1. Introduction
... 1.1 Quantum cosmology and quantum gravity
... 1.2 A brief history of quantum cosmology
2. Hamiltonian Formulation of General Relativity
... 2.1 The $3 + 1$ decomposition
... 2.2 The action
3. Quantisation
... 3.1 Superspace
... 3.2 Canonical quantisation
... 3.3 Path integral quantisation
... 3.4 Minisuperspace
... 3.5 The WKB approximation
... 3.6 Probability measures
... 3.7 Minisuperspace for the Friedmann universe with massive scalar field
4. Boundary Conditions
... 4.1 The no-boundary proposal
... 4.2 The tunneling proposal
5. The Predictions of Quantum Cosmology
... 5.1 The period of inflation
... 5.2 The origin of density perturbations
... 5.3 The arrow of time
6. Conclusion

1. Introduction

These lectures present an introduction to quantum cosmology for an audience consisting for a large part of astronomers, and also a number of particle physicists. As such, the material covered will invariably overlap with that of similar introductory reviews[1,2], although I hope to emphasise, where possible, those aspects of quantum cosmology which are of most interest to astronomers.

This is not an easy task – quantum cosmology is not often discussed at Summer Schools such as this where there is a large emphasis on astrophysical phenomenology, for the very good reason that the ideas involved are at present rather tentative, and quantitative predictions are thus much more difficult to arrive at. Nevertheless we should remember that not too long ago the whole enterprise of cosmology was viewed as the realm of wild speculation. Vladimir Lukash remarked that when he started out in research the advice he was given was that "cosmology was all right for someone like Zeldovich to potter around with after he had already established himself by producing a body of serious work, but it was inappropriate for a young physicist embarking on a career". Happily the situation is quite different today – I am sure the earlier lectures in this School will have convinced you that modern cosmology is a hardcore quantitative science, and that with the new technology and techniques now being developed we can expect to accurately measure all the important cosmological parameters within the next decade and thus enter into a "Golden Era" for cosmological research.

Quantum cosmology, however, still enjoys the sort of status that all of cosmology had until not so very long ago: essentially it is a dangerous field to work in if you hope to get a permanent job. I hope to convince you nevertheless that quantum cosmology represents a vitally important frontier of research, and that although it is by nature somewhat speculative, such speculations are vital if we are to understand the entire history of the universe.

On the face of it the very words "quantum" and "cosmology" do appear to some physicists to be inherently incompatible. We usually think of cosmology in terms of the very large scale structure of the universe, and of quantum phenomena in terms of the very small. However, if the hot big bang is the correct description of the universe – which we can safely assume given the overwhelming evidence described in the earlier lectures – then the universe did start out incredibly small, and there must have been an epoch when quantum mechanics applied to the universe as a whole.

There are people who would take issue with this. In the standard Copenhagen interpretation of quantum mechanics one always has a classical world in which the quantum one is embedded. We have observers who make measurements – the observers themselves are well described by classical physics. If the whole universe is to be treated as a quantum system one does not have such a luxury, and some would argue that our conventional ideas about quantum physics cease to make sense. Yet if quantum mechanics is a universal theory then it is inevitable that some form of "quantum cosmology" was important at the earliest of conceivable times, namely the Planck time, $t_{\rm Planck} = (G\hbar)^{1/2}/c^{5/2} = 5.4 \times 10^{-44}$ sec, (equivalent to 10^{19} GeV as an energy, or 1.6×10^{-35} m as a length). At such scales, where the Compton wavelength of a particle is roughly equal to its gravitational (Schwarzschild) radius, irreducible quantum fluctuations render the classical concept of spacetime meaningless. Whether or not our current efforts at constructing a theory of quantum cosmology are physically valid is therefore really a question of whether our current understanding of quantum physics is adequate for considering the description of processes at the very beginning of the universe, or whether quantum mechanics itself has to be revised at some level. Such a question can really only be answered by extensive work on the problem.

Setting aside the question of the fundamentals of quantum mechanics, let us briefly review the problems which are left unanswered in the standard hot big bang scenario. These are:

1. The *value of* $\eta = N_{\text{baryons}}/N_{\text{photons}}$: the exact value of this parameter is unexplained in the hot big bang but is crucial in determining abundances of light elements through primordial nucleosynthesis.

2. The *horizon problem*: the isotropy of the cosmic microwave background radiation indicates that all regions of the sky must have been in thermal contact at some time in the past. However, in the standard Friedmann-Robertson-Walker (FRW) models regions separated by more than a couple of degrees have non-intersecting particle horizons – i.e., they cannot have been in causal (and hence thermal) contact.

3. The *flatness problem*: given the range of possible values of the ratio of the density of the universe to the critical density at the present epoch, Ω_0, then the FRW models predict that $\Omega(t)$ must have been incredibly close to the spatially flat case $\Omega(t) \simeq 1$ at early times; e.g., assuming $\Omega_0 \gtrsim 0.1$ we find $|\Omega - 1| \lesssim 10^{-26}$ at the lepton era, and $|\Omega - 1| \lesssim 10^{-53}$ at the GUT era.

4. The *unwanted relic problem*: models of the early universe which involve phase transitions often produce copious amounts of topological defects, such as monopoles produced at the GUT scale. If one puts the numbers in one finds that the density of such relics is so great that they would exceed the critical density by such a margin in the standard FRW models that the universe should have ended long ago!

5. The *origin of density perturbations* is unexplained.

6. The *arrow of time* is a physical mystery. On the one hand the laws of physics are CPT-invariant, and on the other there is a *thermodynamic arrow of time*, as prescribed by the Second Law of Thermodynamics, and it appears to match the *cosmological arrow of time*, as prescribed by the expansion of the universe.

7. The *initial conditions* of the universe must be put in by hand, rather than being physically prescribed.

The first four problems on the list are ones that can be explained without appealing to quantum cosmology. The value of η (problem 1) is predicted by models of baryogenesis, which typically take place at the GUT scale. Problems 2–4 are solved by the inflationary universe scenario: an early phase of exponential expansion of the universe drastically changes the past light cone, thereby removing the horizon problem, while also driving Ω close to unity, and diluting unwanted relics to such very low densities that they are close to unobservable.

Problems 5–7, on the other hand, are of a nature which is beyond the scope of the inflationary universe scenario to satisfactorily explain. Inflation provides a mechanism whereby initial small "quantum" perturbations are inflated to all length scales to form a scale-free Harrison-Zeldovich spectrum, but it does not address the question as to exactly how these perturbations arise. Furthermore, a typical model of the very early universe might possess both inflationary and non-inflationary solutions, so that the precise initial conditions of the universe can be crucial for determining whether the universe undergoes a period of inflation *sufficiently long* to be consistent with observation. The length of the period of inflation is precisely the sort of quantitative result that we might hope quantum cosmology should provide. Questions such as the origin of the arrow of time might appear to be of a more philosophical nature – however, quantum cosmology should provide a calculational framework in which such questions can begin to be addressed.

1.1. Quantum cosmology and quantum gravity

Quantum cosmology is perhaps most properly viewed as one attempt among many to grapple with the question of finding a quantum theory of gravity. As a field theory general relativity is not perturbatively renormalisable, and attempts to reconcile general relativity with quantum physics have not yet succeeded despite the attentions of at least one generation of physicists. It is perhaps not surprising that the problem is such a difficult one since general relativity is a theory about the large scale structure of spacetime, and to quantise it we have to quantise spacetime itself rather than simply quantising fields that live in spacetime.

Many ideas have been considered in the quest for a fundamental quantum theory of gravity – whether or not these ideas have brought us closer to that goal is difficult to say without the benefit of hindsight. However, such ideas have certainly profoundly increased our knowledge about the nature of possible physical theories. Some important areas of research have included:

1. *Supergravity*[3–5]. Using supersymmetry, a symmetry between fermions and bosons, one can enlarge the gravitational degrees of freedom to include one or more spin–$\frac{3}{2}$ gravitinos, ψ_μ, in addition to the spin–2 graviton, $g_{\mu\nu}$. Such a symmetry can cure some but not all of the divergences of perturbative quantum general relativity*. In particular, while pure Einstein gravity is perturbatively non-renormalisable at two loops[7], or at one loop if interacting with matter[8–10], in the case of supergravity renormalisability fails only at the 3-loop level[11]▶.

2. *Superstring theory*[13]. Progress can be made if in addition to using supersymmetry, one constructs a theory in which the fundamental objects have an extension rather than being point-like: a theory of strings rather than particles. Much interest in string theory was generated in the mid 1980s with the discovery that certain string theories appear to be finite *at each order* of perturbation theory. In some

* For details of the application of perturbative techniques to quantum cosmology, both with and without supersymmetry, see [6].

▶ The status of the result concerning the 3-loop divergence of supergravity is not quite as rigorous as the other examples mentioned, as the complete 3-loop calculation has not been done. However, it is known that a 3-loop counterterm exists for all extended supergravities and there is no reason to expect the coefficient of the counterterm to be zero, making 3-loop finiteness extremely unlikely. For a review of the ultraviolet properties of supersymmetric field theories see [12].

sense stringy physics "smears out" the problems associated with pointlike interactions. The entire sum of all terms in the perturbation expansion diverges in the case of the bosonic string[14], however, and it is believed that similar results should apply to the superstring[⊛]. Furthermore, despite what many see as the mathematical beauty of string theory, there has unfortunately as yet been no definitive success in deriving concrete phenomenological predictions.

3. *Non-perturbative canonical quantum gravity*[16,17]. The fact that general relativity is not perturbatively renormalisable might simply be a failure of flat spacetime quantum field theory techniques to deal with such an inherently non-linear theory, rather than reflecting an inherent incompatibility between general relativity and quantum physics. Given the divergence of the string perturbation series mentioned above, a non-perturbative formulation of string theory would also be desirable. As a starting point a systematic investigation of a non-perturbative canonical formalism based on general relativity could provide deep insights into quantum gravity. Such a programme has been investigated by Ashtekar and coworkers, mainly from the mid 1980s onwards. One principal difference from the canonical formalism which I will describe in section 2 is that instead of taking the metric to be the fundamental object to describe the quantum dynamics, one bases such a dynamics on a *connection*, in this case an $SL(2,\mathbb{C})$ spin connection[18,19]. If one integrates this connection around a closed loop one arrives at "loop variables", which might be considered to be analogous to the magnetic flux, Φ, obtained by integrating the electromagnetic gauge potential, A_μ, around a closed loop. In the "loop representation" quantum states are represented by functionals of such loops on a 3-manifold[20,21], rather than by functionals of classical fields. Although a number of technical difficulties remain, considerable progress has been made with the Ashtekar formulation, and the reformulation of quantum cosmology in the Ashtekar framework[22–25] could be an area for some interesting future work.

4. *Alternative models of spacetime*. All the above approaches assume that the basic quantum variables, be they a metric or a connection, are defined on differentiable manifolds. Given that it is highly possible that "something strange" happens at the Planck length it is plausible that one might have to abandon this assumption in order to effectively describe the quantum dynamics of gravity. A number of ideas have been considered on these lines. One framework which has been widely used both for numerical relativity and studies of quantum gravity is that of *Regge calculus*[26△] in which one replaces smooth manifolds by spaces consisting of piecewise linear simplicial blocks. Naturally, other possibilities for discretised spacetime structure also exist – *causal sets*[28] being another example which has not yet been so widely explored. A further possibility is *topological quantisation* whereby one replaces a manifold by a set and quantises all topologies on that set[29,30].

⊛ The question of finiteness of superstring perturbation theory is a difficult technical question, which has still to be resolved – see, e.g., [15].

△ For a brief review and extensive bibliography see [27].

Many of the above alternatives fall into the category of being attempts to construct a fundamental theory of quantum gravity. The current quantum cosmology programme is not quite as ambitious. One begins by making the assumption that whatever the exact nature of the fundamental theory of quantum gravity is, in its semiclassical limit it should agree with the semiclassical limit of a canonical quantum formalism based on general relativity alone. Thus we study the semiclassical properties of quantum gravity based solely on Einstein's theory, or some suitable modification of it.

Clearly, any predictions made from such a foundation must be treated with caution. In particular, a new fundamental theory of quantum gravity might introduce radically new physical processes at an energy scale relevant for cosmology. String theory, for example, introduces its own fundamental scale which expressed as a temperature (the *Hagedorn temperature*) is given by:

$$T_{\text{Hagedorn}} = \frac{\sqrt{\hbar c}}{4\pi k_B \sqrt{\alpha'}}, \tag{1.1}$$

the constant α' being the Regge slope parameter, which is inversely proportional to the string tension. It is not known exactly what the value of α' is; however, if T_{Hagedorn} is comparable to or significantly lower than the Planck scale, then it is clear that fundamentally "stringy" processes will be very important in the very early universe, if string theory is indeed the "ultimate" theory.

Although new fundamental physics could drastically change the predictions of quantum cosmology, I believe nevertheless that studying quantum cosmology based even just on Einstein's theory is an important activity. General relativity is a remarkably successful theory; in seeking to replace it by something better it is important that we study processes at the limit of its applicability, thereby challenging our understanding. General relativity is limited by the Planck scale – the physical arenas in which this scale is approached include: (i) very small black holes; (ii) the very early universe. Since it seems unlikely that we will ever be able to create energies of order 10^{19} GeV in the laboratory, the consequences of quantum gravity for the physics of the very early universe will remain the one way of indirectly "testing" it, at least for the foreseeable future.

It is thus important that we consider quantum gravity in a cosmological context. Even if our current attempts do not fully reach the mark, in that we do not yet have a fully-fledged quantum theory of gravity, they nonetheless constitute a vital part of the process of trying to find such a theory.

1.2. A brief history of quantum cosmology

The quantum cosmology programme which I will describe in these lectures has gone through three main identifiable phases to date:

1. *Defining the problem.* The canonical formalism, including the definition of the wavefunction of the universe, Ψ, its configuration space – superspace – and its evolution according to the Wheeler-DeWitt equation, was set up in the late 1960s[31–35].

2. *Boundary conditions.* Quantum cosmology research went into something of a lull during the 1970s but was revived in the mid 1980s when the question of putting appropriate boundary conditions on the wavefunction of the universe was treated seriously. The idea is that such boundary conditions should describe the "creation of the universe from nothing"[36,37], where *nothing* means the absence of space and time. A number of proposals for such boundary conditions emerged – two major contenders being the *"no-boundary"* proposal of Hartle and Hawking[38,39] and the *"tunneling"* proposal advocated by Vilenkin[40–42].

3. *Quantum decoherence.* The mechanism of the transition from quantum physics to classical physics ("quantum decoherence") becomes vitally important when quantum physics is applied to the universe as a whole. The issues involved have begun to occupy many researchers in the early 1990s[▼].

Two other important areas of quantum cosmology (or related) research have been: (i) quantum wormholes and "baby universes"; (ii) supersymmetric quantum cosmology.

Quantum wormholes[⋇] were extremely fashionable in the particle physics community in the years 1988–1990. Such states arise when one considers topology change in the path integral formulation of quantum gravity: *quantum wormholes* are instanton solutions which play an important role in the Euclidean path integral. One deals directly with a "third quantised" formalism, (i.e., quantum field theory over superspace), which includes operators that create and destroy universes – so-called *baby universes.* Much of the excitement in the late 1980s was associated with the idea that such processes could fix the fundamental constants of nature – in particular, driving the cosmological constant to zero.

Supersymmetric quantum cosmology has emerged recently as one of the most active areas of current research[⫛]. In considering the quantum creation of the universe we are of course dealing with the very earliest epochs of the universe's existence, at which time it is believed that supersymmetry would not yet be broken. The inclusion of supersymmetry could therefore be vital from the point of view of physical consistency.

Since the focus of this School is on cosmology, my intention in these lectures is to cover topics 1 and 2 above, and then to proceed to discuss the predictions of quantum cosmology. The third topic of quantum decoherence raises questions which have not been resolved even in ordinary quantum mechanics, since the question of decoherence really amounts to understanding what happens during the "collapse of the wavefunction". Although this is a fascinating issue it has more to do with

▼ See [43] for a review and [44] for a collection of recent papers on the subject.

⋇ See [45] for a review.

⫛ See [46] and references therein.

the fundamentals of quantum mechanics than directly with cosmology. Likewise, I will only briefly touch upon quantum wormholes and supersymmetric quantum cosmology, as these areas are still in their infancy, and one is still at the point of trying to resolve basic questions concerning quantum gravity. I hope the reader will not be disappointed by this – however, given the vast scope of quantum gravity and quantum cosmology one must necessarily be rather selective.

2. Hamiltonian Formulation of General Relativity

2.1. The $3+1$ decomposition

In order to discuss quantum cosmology a fair amount of technical machinery is required. In the canonical formulation we begin by making a $3+1$-split of the 4-dimensional spacetime manifold, $\mathcal{M}$, which will describe the universe, foliating it into spatial hypersurfaces, Σ_t, labeled by a global time function, t. Thus we take the 4-dimensional metric to be given by♦

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\boldsymbol{\omega}^0 \otimes \boldsymbol{\omega}^0 + h_{ij}\boldsymbol{\omega}^i \otimes \boldsymbol{\omega}^j, \tag{2.1}$$

where

$$\begin{aligned} \boldsymbol{\omega}^0 &= \mathcal{N}\mathrm{d}t \\ \boldsymbol{\omega}^i &= \mathrm{d}x^i + \mathcal{N}^i\mathrm{d}t. \end{aligned} \tag{2.2}$$

Such a decomposition is possible in general if the manifold $\mathcal{M}$ is *globally hyperbolic* [47]. The quantity $\mathcal{N}(t, x^k)$ is called the *lapse function* – it measures the difference between the coordinate time, t, and proper time, τ, on curves normal to the hypersurfaces Σ_t, the normal being $n_\alpha = (-\mathcal{N}, 0, 0, 0)$ in the above coordinates. The quantity $\mathcal{N}^i(t, x^k)$ is called the *shift vector* – it measures the difference between a spatial point, p, and the point one would reach if instead of following p from one hypersurface to the next one followed a curve tangent to the normal $\boldsymbol{n}$. That is to say, the spatial coordinates are "comoving" if $\mathcal{N}^i = 0$. Finally, $h_{ij}(t, x^k)$ is the *intrinsic metric* (or *first fundamental form*) induced on the spatial hypersurfaces by the full 4-dimensional metric, $g_{\mu\nu}$. In components we have

$$(g_{\mu\nu}) = \begin{bmatrix} -\mathcal{N}^2 + \mathcal{N}^k\mathcal{N}_k & \mathcal{N}_j \\ \mathcal{N}_i & h_{ij} \end{bmatrix}, \tag{2.3}$$

with inverse

$$(g^{\mu\nu}) = \begin{bmatrix} \frac{-1}{\mathcal{N}^2} & \frac{\mathcal{N}^j}{\mathcal{N}^2} \\ \frac{\mathcal{N}^i}{\mathcal{N}^2} & h^{ij} - \frac{\mathcal{N}^i\mathcal{N}^j}{\mathcal{N}^2} \end{bmatrix}, \tag{2.4}$$

where h^{ij} is the inverse to h_{ij}, and the intrinsic metric is used to lower and raise spatial indices: $\mathcal{N}^k\mathcal{N}_k \equiv h^{jk}\mathcal{N}_j\mathcal{N}_k = h_{jk}\mathcal{N}^j\mathcal{N}^k$ etc.

♦ We use a $(-+++)$ Lorentzian metric signature, and natural units in which $c = \hbar = 1$.

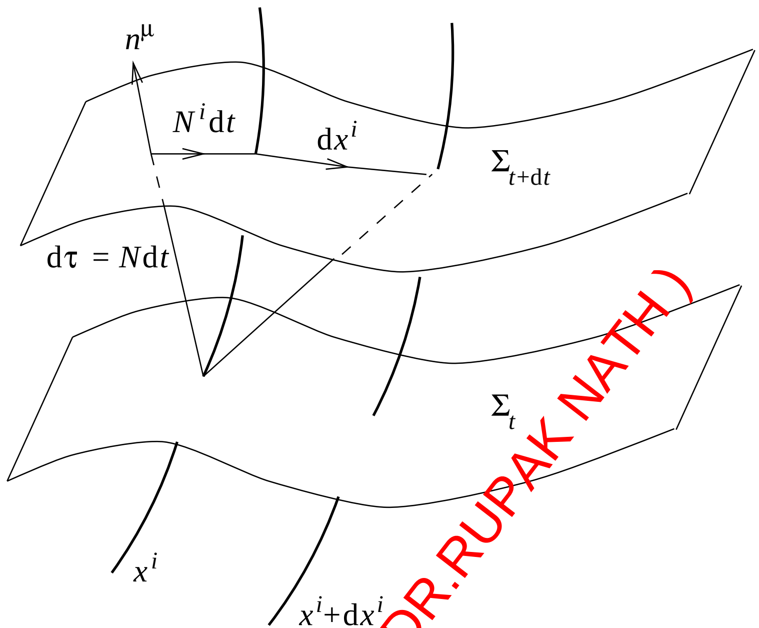

Fig. 1: The $3+1$ decomposition of the manifold, with lapse function, N, and shift vector, N^i.

One can construct an *intrinsic curvature* tensor ${}^3R^i{}_{jkl}(h)$ from the intrinsic metric alone – this of course describes the curvature intrinsic to the hypersurfaces Σ_t. One can also define an *extrinsic curvature*, (or *second fundamental form*), which describes how the spatial hypersurfaces Σ_t curve with respect to the 4-dimensional spacetime manifold within which they are embedded. This is given by

$$\begin{aligned} K_{ij} &\equiv -\, n_{i;j} = -\Gamma^0{}_{ij} n_0 \\ &= \frac{1}{2\mathcal{N}}\left(\mathcal{N}_{i|j} + \mathcal{N}_{j|i} - \frac{\partial h_{ij}}{\partial t}\right), \end{aligned} \tag{2.5}$$

where a semicolon denotes covariant differentiation with respect to the 4-metric, $g_{\mu\nu}$, and a vertical bar denotes covariant differentiation with respect to the 3-metric, h_{ij}: $\mathcal{N}_{i|j} \equiv \mathcal{N}_{i,j} - \Gamma^k{}_{ij}\mathcal{N}_k$ etc.

For a given foliation of $\mathcal{M}$ by spatial hypersurfaces, Σ_t, it is always possible to choose *Gaussian normal coordinates*, in which

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + h_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{2.6}$$

These are comoving coordinates ($\mathcal{N}^i = 0$) with the additional property that t is the proper time parameter ($\mathcal{N} = 1$). This is the standard "gauge choice" that is made in classical cosmology, and in such coordinates $K_{ij} = -\dot{h}_{ij}$, where dot denotes differentiation with respect to t. In making the $3 + 1$ decomposition, however, we are only free to make a specific choice of coordinates such as (2.6) *after* variation of the action if we want to be sure to obtain Einstein's equations, and thus we must retain the lapse and shift function for the time being.

2.2. The action

A relevant action for use in quantum cosmology is that of Einstein gravity plus a possible cosmological term, Λ, and matter, given by

$$S = \frac{1}{4\kappa^2}\left[\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\left({}^4\!R - 2\Lambda\right) + 2\int_{\partial\mathcal{M}} \mathrm{d}^3x\sqrt{h}\,K\right] + S_{\text{matter}}, \tag{2.7}$$

where $\kappa^2 = 4\pi G = 4\pi m_{\text{Planck}}^{-2}$, $K \equiv K^i{}_i$ is the trace of the extrinsic curvature, and for many simple models the matter is specified by a single scalar field, Φ, with potential, $\mathcal{V}(\Phi)$,

$$S_{\text{matter}} = \int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\left(-\frac{1}{2}g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi - \mathcal{V}(\Phi)\right). \tag{2.8}$$

The boundary term[48] in (2.7) does not of course contribute to the classical field equations, and this term is usually omitted in a first course on general relativity. However, in quantum physics we are often interested in phenomena which occur when the classical field equations do not apply, that is "off-shell", and thus it is vitally important to retain the surface term. The simple matter action (2.8) given here should simply be seen as being representative of the type of matter action one might consider. Although the example given by (2.8) is sufficient for studying many inflationary universe models, many other alternatives might also be of interest, such as extra matter from a supergravity multiplet, or matter corresponding to the low-energy limit of string theory. In the latter case, if one works in the "string conformal frame" it is also necessary to alter the gravitational part of the action, as one characteristic of string theory is the presence of the scalar dilaton, Φ, which couples universally to matter (at least perturbatively). In that case (2.7), (2.8) would be replaced by

$$\begin{aligned} S = \frac{1}{4\kappa^2}\Bigg[\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\mathrm{e}^{-2\Phi}\left({}^4\!R + 4g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi - 8\mathcal{V}(\Phi) + \ldots\right) \\ + 2\int_{\partial\mathcal{M}} \mathrm{d}^3x\sqrt{h}\mathrm{e}^{-2\Phi}K\Bigg], \end{aligned} \tag{2.9}$$

where the ellipsis denotes any additional matter degrees of freedom, and we have allowed for the possibility of the generation of a dilaton potential, $\mathcal{V}(\Phi)$, via some non-perturbative symmetry breaking mechanism. However, for simplicity the specific examples we will deal with here will be confined to models of the type (2.7), (2.8).

We now wish to express (2.7), (2.8) in terms of the variables of the $3+1$ split. One can show that

$$ {}^4R = {}^3R - 2R_{nn} + K^2 - K^{ij}K_{ij}, \tag{2.10} $$

where ${}^3R$ is the Ricci scalar of the intrinsic 3-geometry, and

$$ R_{nn} \equiv R_{\alpha\beta}n^\alpha n^\beta = -K^{ij}K_{ij} + K^2 + \left(n^\alpha K + a^\alpha\right)_{;\alpha}, \tag{2.11} $$

with $a^\alpha \equiv n^\beta n^\alpha{}_{;\beta}$. Combining (2.7), (2.10) and (2.11), and noting that the boundary integral involving a^α vanishes identically since $n^\alpha a_\alpha = 0$, we obtain

$$ S \equiv \int \mathrm{d}t\, L = \frac{1}{4\kappa^2}\int \mathrm{d}t\, \mathrm{d}^3x \mathcal{N}\sqrt{h}\left(K_{ij}K^{ij} - K^2 + {}^3R - 2\Lambda\right) + S_{\text{matter}}. \tag{2.12} $$

As in the Hamiltonian formulation of field theory we define canonical momenta in the standard fashion

$$ \pi^{ij} \equiv \frac{\delta L}{\delta \dot{h}_{ij}} = -\frac{\sqrt{h}}{4\kappa^2}\left(K^{ij} - h^{ij}K\right), \tag{2.13} $$

$$ \pi_{\Phi} \equiv \frac{\delta L}{\delta \dot{\Phi}} = \frac{\sqrt{h}}{\mathcal{N}}\left(\dot{\Phi} - \mathcal{N}^i\Phi_{,i}\right), \tag{2.14} $$

$$ \pi^0 \equiv \frac{\delta L}{\delta \dot{\mathcal{N}}} = 0, \tag{2.15} $$

$$ \pi^i \equiv \frac{\delta L}{\delta \dot{\mathcal{N}}_i} = 0. \tag{2.16} $$

The fact that the momenta conjugate to $\mathcal{N}$ and $\mathcal{N}_i$ vanish means that we are dealing with *primary constraints* in Dirac's terminology[49,50].

If we use $\mathcal{N}$, $\mathcal{N}_i$, h_{ij}, Φ and their conjugate momenta as the basic variables we obtain$^\diamondsuit$ a Hamiltonian

$$ \begin{aligned} H &\equiv \int \mathrm{d}^3x \left(\pi^0\dot{\mathcal{N}} + \pi^i\dot{\mathcal{N}}_i + \pi^{ij}\dot{h}_{ij} + \pi_\Phi\dot{\Phi}\right) - L \\ &= \int \mathrm{d}^3x \left(\pi^0\dot{\mathcal{N}} + \pi^i\dot{\mathcal{N}}_i + \mathcal{N}\mathcal{H} + \mathcal{N}_i\mathcal{H}^i\right), \end{aligned} \tag{2.17} $$

where

$$ \mathcal{H} = \frac{\sqrt{h}}{2\kappa^2}\left(G^{\hat{0}}{}_{\hat{0}} - 2\kappa^2 T^{\hat{0}}{}_{\hat{0}}\right) $$

$^\diamondsuit$ Details of all missing steps in this section will be provided upon request in a plain brown envelope.

$$= 4\kappa^2 \mathcal{G}_{ijkl}\pi^{ij}\pi^{kl} - \frac{\sqrt{h}}{4\kappa^2}\left({}^3R - 2\Lambda\right) + \tfrac{1}{2}\sqrt{h}\left[\frac{{\pi_{\Phi}}^2}{h} + h^{ij}\Phi,_i \Phi,_j + 2\mathcal{V}\right], \quad (2.18)$$

$$\mathcal{H}^i = \frac{\sqrt{h}}{2\kappa^2}\left(G^{\hat{0}\hat{\imath}} - 2\kappa^2 T^{\hat{0}\hat{\imath}}\right)$$
$$= -2{\pi^{ij}}_{|j} + h^{ij}\Phi,_j \pi_{\Phi}, \quad (2.19)$$

and

$$\mathcal{G}_{ijkl} = \tfrac{1}{2}h^{-1/2}\left(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}\right) \quad (2.20)$$

is the *DeWitt metric*[33]. The hats in (2.18) and (2.19) denote orthonormal frame components of the Einstein and energy-momentum tensors. In terms of these variables the action (2.12) then becomes

$$S = \int \mathrm{d}t\, \mathrm{d}^3x \left(\pi^0 \dot{\mathcal{N}} + \pi^i \dot{\mathcal{N}}_i - \mathcal{N}\mathcal{H} - \mathcal{N}_i\mathcal{H}^i\right). \quad (2.21)$$

If we vary (2.21) with respect to π^{ij} and π_{Φ} we obtain their defining relations (2.13) and (2.14). The lapse and shift functions now act as Lagrange multipliers; variation of (2.21) with respect to the lapse function, $\mathcal{N}$, yields the *Hamiltonian constraint*

$$\mathcal{H} = 0, \quad (2.22)$$

while variation of (2.21) with respect to the shift vector, $\mathcal{N}_i$, yields the *momentum constraint*

$$\mathcal{H}^i = 0. \quad (2.23)$$

From (2.18) and (2.19) it is clear that these constraints are simply the (00) and $(0i)$ parts of Einstein's equations. In Dirac's terminology these are *secondary* or *dynamical constraints.*

The 3+1 decomposition of our spacetime looks to be very counterintuitive to the usual ideas of general relativity. This is so by choice. Einstein described the l.h.s. of his field equations which encodes the 4-dimensional geometry as a "hall of marble", which does not encourage people to tamper with it. However, to quantise spacetime we must do just that: we must deconstruct spacetime and replace it by something else. Thus far we have not done that – we have simply split our 4-dimensional manifold into a sequence of spatial hypersurfaces, Σ_t. Time is the natural variable to base this split upon since it plays a special role in quantum mechanics – it is a parameter rather than an operator.

Classically the evolution of one spatial hypersurface to the next is completely well-defined, (provided that the manifold, $\mathcal{M}$, is globally hyperbolic), and given initial data, h_{ij}, Φ, on an initial hypersurface, Σ, we can use the Cauchy development to stitch the hypersurfaces Σ_t together to recover the 4-dimensional manifold, $\mathcal{M}$. To quantise the theory, however, we want to perform a path integral over *all* geometries, not just the classically allowed ones. Thus we must consider sequences of geometries at the quantum level which *cannot* be stitched together in a regular Cauchy development to form a 4-manifold which solves Einstein's equations. (See Fig. 2.) We must therefore abandon Einstein's "hall of marble" – spacetime is no longer a fundamental object.

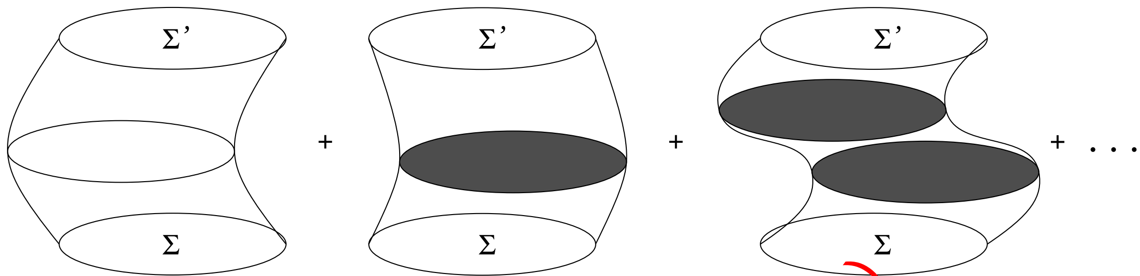

Fig. 2: Quantum geometrodynamics: in addition to the classical Cauchy development from Σ to Σ' (left), the path integral includes a sum over all 4-manifolds which interpolate between the initial and final configurations. The weighting by e^{iS} means that the greater the number of classically forbidden 3-geometries (shaded slices) is in the interpolating 4-manifold, the smaller its contribution is to the path integral.

As was mentioned in the Introduction the "deconstruction" of spacetime that we adopt here is probably the most conservative choice we could make. Even though we abandon the notion of spacetime in discussing the quantum dynamics of gravity, our fundamental objects are still defined on regular 3-manifolds, Σ. A more radical departure would be to replace these spatial hypersurfaces by some more general set by using, for example, the ideas of Regge calculus, causal sets or topological quantisation.

3. Quantisation

3.1. Superspace

As a prelude to the canonical quantisation of gravity let us first introduce the relevant configuration space on which the quantum dynamics will be defined.

Consider the space of all Riemannian 3-metrics and matter configurations on the spatial hypersurfaces, Σ,

$$\text{Riem}\,(\Sigma) \equiv \{h_{ij}(x), \Phi(x) \mid x \in \Sigma\}\,. \tag{3.1}$$

This is an infinite-dimensional space, on account of the label $x = \{x^i\}$, which specifies the point on the hypersurface, but there are a finite number of degrees of freedom at each point, $x \in \Sigma$. In fact, we are really interested in geometry here and configurations which can be related to each other by a diffeomorphism, i.e., a coordinate transformation, should be considered to be equivalent since their intrinsic geometry is the same. Thus we factor out by diffeomorphisms on the spatial hypersurfaces and identify *superspace*⋆ as

$$\frac{\mathrm{Riem}\,(\Sigma)}{\mathrm{Diff}_0(\Sigma)}\,,$$

where the subscript zero denotes the fact that we consider only diffeomorphisms which are connected to the identity. This infinite-dimensional space will be the basic configuration space of quantum cosmology.

The DeWitt metric (2.20) then provides a metric on superspace which we can write as

$$\mathcal{G}_{AB}(x) \equiv \mathcal{G}_{(ij)(kl)}(x), \tag{3.2}$$

where the indices A, B run over the independent components of the intrinsic metric h_{ij}:

$$A, B \ \in \{h_{11}, h_{12}, h_{13}, h_{22}, h_{23}, h_{33}\}.$$

The DeWitt metric has a $(-+++++)$ signature at each point $x \in \Sigma$, regardless of the signature of the spacetime metric, $g_{\mu\nu}$. To incorporate all the degrees of freedom, we also need to extend the range of the indices A, B to include the matter fields, by appropriately defining $\mathcal{G}_{\Phi\Phi}(x)$ (and other components if more than one matter field is present), thereby obtaining the full *supermetric*.

An inverse DeWitt metric, $\mathcal{G}^{AB}(x) = \mathcal{G}^{(ij)(kl)}(x)$, can be defined by the requirement

$$\mathcal{G}^{(ij)(kl)}\mathcal{G}_{(kl)(mn)} = \tfrac{1}{2}\left(\delta^i{}_m\delta^j{}_n + \delta^i{}_n\delta^j{}_m\right), \tag{3.3}$$

which gives

$$\mathcal{G}^{(ij)(kl)} = \tfrac{1}{2}\sqrt{h}\left(h^{ik}h^{jl} + h^{il}h^{jk} - 2h^{ij}h^{kl}\right). \tag{3.4}$$

3.2. Canonical quantisation

We will perform canonical quantisation by taking the wavefunction of the universe, $\Psi[h_{ij}, \Phi]$, to be a functional on superspace. Unlike the case of ordinary quantum mechanics, the wavefunction, Ψ, does not depend explicitly on time here. This is related to the fact that general relativity is an "already parametrised" theory – the original Einstein-Hilbert action is time-reparametrisation invariant. Time is contained implicitly in the dynamical variables, h_{ij} and Φ.

According to Dirac's quantisation procedure[49] we make the following replacements for the canonical momenta

$$\pi^{ij} \to -i\frac{\delta}{\delta h_{ij}}, \quad \pi_{\Phi} \to -i\frac{\delta}{\delta\Phi}, \quad \pi^0 \to -i\frac{\delta}{\delta\mathcal{N}}, \quad \pi^i \to -i\frac{\delta}{\delta\mathcal{N}_i}\,. \tag{3.5}$$

⋆ The use of the terminology "superspace" for the configuration space of quantum cosmology predates the discovery of supersymmetry, and of course is completely different from the combined manifold of commuting and anticommuting coordinates which is called "superspace" in supersymmetric theories.

and then demand that the wavefunction is annihilated by the operator version of the constraints. For the primary constraints we have

$$\begin{aligned} \hat{\pi}\Psi &= -i\frac{\delta\Psi}{\delta\mathcal{N}} = 0, \\ \hat{\pi}^i\Psi &= -i\frac{\delta\Psi}{\delta\mathcal{N}_i} = 0, \end{aligned} \tag{3.6}$$

which implies that Ψ is independent of $\mathcal{N}$ and $\mathcal{N}^i$. The momentum constraint yields

$$\hat{\mathcal{H}}^i\Psi = 0 \qquad \Rightarrow \qquad i\left[\frac{\delta\Psi}{\delta h_{ij}}\right]_{|j} = 2\kappa^2\hat{T}^{\hat{0}\hat{\imath}}\Psi. \tag{3.7}$$

Using (3.7) one can show[51] that Ψ is the same for configurations $\{h_{ij}(x), \Phi(x)\}$ which are related by a coordinate transformation in the spatial hypersurface, Σ, which accords with the rationale for factoring out by diffeomorphisms in our definition of superspace. Finally, the Hamiltonian constraint yields

$$\hat{\mathcal{H}}\Psi = \left[-4\kappa^2\mathcal{G}_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} + \frac{\sqrt{h}}{4\kappa^2}\left(-{}^3R + 2\Lambda + 4\kappa^2\hat{T}^{\hat{0}\hat{0}}\right)\right]\Psi = 0, \tag{3.8}$$

where for our scalar field example

$$\hat{T}^{\hat{0}\hat{0}} = \frac{-1}{2h}\frac{\delta^2}{\delta\Phi^2} + \tfrac{1}{2}h^{ij}\Phi_{,i}\,\Phi_{,j} + \mathcal{V}(\Phi). \tag{3.9}$$

Eq. (3.8) is known as the *Wheeler-DeWitt equation*[32,33]. In fact, it is not a single equation but is actually one equation at each point, $x \in \Sigma$. It is a second-order hyperbolic functional differential equation on superspace. On account of factor-ordering ambiguities it is not completely well-defined, although there exist "natural" choices[33,52] of ordering for which the derivative pieces become a Laplacian in the supermetric.

3.3. Path integral quantisation

An alternative to canonical quantisation which perhaps better accommodates an intuitive understanding (c.f. Fig. 2) of the quantisation procedure is the path integral approach. Path integral techniques in quantum gravity were pioneered in the late 1970s[53,54]. The starting point for this is Feynman's idea that the amplitude to go from one state with an intrinsic metric, h_{ij}, and matter configuration, Φ, on an initial hypersurface, Σ, to another state with metric, h'_{ij} and matter configuration Φ' on an final hypersurface, Σ', is given by a functional integral of e^{iS} over all 4-geometries, $g_{\mu\nu}$, and matter configurations, ϕ, which interpolate between the initial and final configurations

$$\langle h'_{ij}, \Phi', \Sigma' | h_{ij}, \Phi, \Sigma\rangle = \sum_{\mathcal{M}} \int \mathfrak{D}\boldsymbol{g}\,\mathfrak{D}\phi\; e^{iS[g_{\mu\nu},\phi]}. \tag{3.10}$$

In ordinary quantum field theory in flat spacetime the path integral suffers from the difficulty that since the action $S[g_{\mu\nu}, \phi]$ is finite the integral oscillates and therefore fails to converge. Furthermore, the solution which extremises the action is given by solving a hyperbolic equation between initial and final boundary surfaces – which is not a mathematically well-posed problem, and may have either no solutions or an infinite number of solutions. To deal with this problem one performs a "Wick rotation" to "imaginary time" $t \to -i\tau$ and considers a path integral formulated in terms of the Euclidean action, $I = -iS$. The action is then positive-definite, so that the path integral is exponentially damped and should converge. Also the problem of finding the extremum becomes that of solving an elliptic equation with boundary conditions, and this is well-posed.

One may attempt to apply a similar approach to quantum gravity, replacing S in (3.10) by the Euclidean action⊞ $I[g_{\mu\nu}, \phi] = -iS[g_{\mu\nu}, \phi]$, and taking the sum in (3.10) to be over all metrics with signature $(++++)$, which induce the appropriate 3-metrics h_{ij} and h'_{ij} on the past and future hypersurfaces. This approach to quantum gravity has had some important successes – most notably, it provides: (i) an elegant way of deriving the thermodynamic properties of black holes[55–57]; and (ii) a natural means for discussing the effects of gravitational instantons[58–60]. Gravitational instantons have been found to provide the dominant contribution to the path integral in processes such as pair creation of charged black holes in a magnetic field[61–63], and therefore provide a means of calculating the rates of such processes semiclassically.

The problems associated with the Euclidean approach to quantum gravity are considerable, however. Firstly, unlike ordinary field theories• such as Yang-Mills theory the gravitational action is not positive-definite[64], and thus the path integral does not converge if one restricts the sum to real Euclidean-signature metrics. To make the path integral converge it is necessary to include complex metrics in the sum[64,65]. However, there is no unique contour to integrate along in superspace[66–68] and the result one obtains for the path integral may depend crucially on the contour that is chosen[66]. Furthermore, the measure in (3.10) is ill-defined. It is really only in the last ten years that mathematicians have succeeded in making path integration in ordinary quantum field theory rigorously defined. Clearly, we may have to wait some time before the same can be said of path integrals in quantum gravity.

The physicists' approach is to set aside the issues involved in making the formalism rigorous and to see what can be learned nevertheless. We thus take the

⊞ Strictly speaking one should call this the Riemannian action, since "Euclidean" spaces are those which are *flat*, whereas *curved* manifolds with $(++++)$ signature are known as *Riemannian spaces*. However, the terminology "Euclidean" action which has been taken over from flat space quantum field theory seems to have stuck, despite the fact that we are of course no longer dealing with $\mathbb{R}^4$.

• The action for fermi fields in ordinary quantum field theory is not positive-definite, but that is not a problem since one can treat them as anticommuting quantities so that the path integral over them converges.

wavefunction Ψ of the universe on a hypersurface, Σ, with intrinsic 3-metric, h_{ij}, and matter configuration, Φ, to be defined[38,39] by the functional integral

$$\Psi\left[h_{ij}, \Phi, \Sigma\right] = \sum_{\mathcal{M}} \int \mathfrak{D}\boldsymbol{g}\,\mathfrak{D}\phi\, \mathrm{e}^{-I[g_{\mu\nu},\phi]}. \tag{3.11}$$

where the sum is over a class of 4-metrics, $g_{\mu\nu}$, and matter configurations, ϕ, which take values h_{ij} and Φ on the boundary Σ. Alternative definitions of the wavefunction have been proposed. In particular, Linde[69] has argued that one should Wick rotate with the opposite sign, i.e., $t \to +i\tau$ instead of $t \to -i\tau$ as above, leading to a factor e^{+I} instead of e^{-I} in (3.11). Alternatively, one could stick with a Lorentzian path integral[70], with e^{iS} instead of e^{-I} in (3.11). In any case, in order to achieve convergence of the path integral[⊖] it is necessary to include complex manifolds in the sum, which somewhat obscures the distinctions between these alternative proposed definitions of Ψ. The real distinction between the alternative definitions arises when it comes to imposing boundary conditions, thereby restricting the 4-manifolds included in the sum in (3.11). For example, one could view the Euclidean sector as being the appropriate sector of the quantum theory in which an "initial" boundary condition on Ψ should be imposed, which would make (3.11) the natural starting point, as is the case for the no-boundary proposal[38,39]. Alternative boundary conditions would favour the Lorentzian path integral[70].

The path integral definition of the wavefunction (3.11) is consistent with the earlier definition based on canonical quantisation to the extent that wavefunctions defined according to (3.11) can be shown to satisfy the Wheeler-DeWitt equation (3.8) provided that the action, the measure and the class of paths summed over are invariant under diffeomorphisms[71].

In the canonical quantisation formalism any particular solution to the Wheeler-DeWitt equation will depend upon the specification of boundary conditions on the wavefunction. In the path integral formulation the particular solution of the Wheeler-DeWitt equation that one obtains will similarly depend on the contour of integration chosen in superspace, and the class of 4-metrics one sums over in (3.11). Unfortunately it is not known how the choice of contour and class of paths prescribes the boundary conditions on the wavefunction in the general case, although it can be answered for specific models. The question of boundary conditions is naturally of prime importance for cosmology, and we shall return to this question in §4.

3.4. Minisuperspace

In practice to work with the infinite dimensions of the full superspace is not possible, at least with the techniques that have been developed to date. One useful

---

⊖ Linde's suggested modification [69] to (3.11) yields a convergent path integral for the scale factor, which is all that one needs in the simplest minisuperspace models, but does not lead to convergence if one includes matter or inhomogeneous modes of the metric.

approximation therefore is to truncate the infinite degrees of freedom to a finite number, thereby obtaining some particular *minisuperspace* model. An easy way to achieve this is by considering homogeneous metrics, since as was observed earlier for each point $x \in \Sigma$ there are a finite number of degrees of freedom in superspace.

The results we shall obtain by this approach will be somewhat satisfying in that they do appear to have some predictive power. However, the truncation to minisuperspace has not as yet been made part of a rigorous approximation scheme to full superspace quantum cosmology. As they are currently formulated minisuperspace models should therefore be viewed as toy models, which we nonetheless hope will capture some of the essence of quantum cosmology. Since we are simultaneously setting most of the field modes and their conjugate momenta to zero, which violates the uncertainty principle, this approach might seem rather *ad hoc*. However, in classical cosmology homogeneity and isotropy are important simplifying assumptions which do have a sound observational basis. Therefore it is not entirely unreasonable to hope that a consistent truncation to particular minisuperspace models with particular symmetries might be found in future♣.

Let us thus consider *homogeneous* cosmologies for simplicity. Instead of having a separate Wheeler-DeWitt equation for each point of the spatial hypersurface, Σ, we then simply have a single Wheeler-DeWitt equation for all of Σ. The standard FRW metrics, with

$$h_{ij}\mathrm{d}x^i\mathrm{d}x^j = a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2\right)\right], \qquad k=-1,0,1, \tag{3.12}$$

are of course one special example. In that case our model with a single scalar field simply has two minisuperspace coordinates, $\{a, \Phi\}$, the cosmic scale factor and the scalar field. Many more general homogeneous (but anisotropic) models can also be considered. Indeed all such models can be classified◁ and apart from the FRW models the other cases of interest are: (i) the *Kantowski-Sachs models*[77,78], which have a 3-metric

$$h_{ij}\mathrm{d}x^i\mathrm{d}x^j = a^2(t)\mathrm{d}r^2 + b^2(t)\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2\right), \tag{3.13}$$

and thus three minisuperspace coordinates, $\{a, b, \Phi\}$; and (ii) the Bianchi models.

The *Bianchi models* are the most general homogeneous cosmologies with a 3-dimensional group of isometries. These groups are in a one–to–one correspondence with 3-dimensional Lie algebras, which were classified long ago by Bianchi[79]. There are nine distinct 3-dimensional Lie algebras, and consequently nine types of Bianchi cosmology. The 3-metric of each of these models can be written in the form

$$h_{ij}\mathrm{d}x^i\mathrm{d}x^j = h_{ij}(t)\boldsymbol{\omega}^i \otimes \boldsymbol{\omega}^j, \tag{3.14}$$

♣ For some discussions of the validity of the minisuperspace approximation see [72–75].

◁ See [76] for a review.

where $\boldsymbol{\omega}^i$ are the invariant 1-forms associated associated with the isometry group. The simplest example is Bianchi I, which corresponds to the Lie Group $\mathbb{R}^3$. In that case we can choose $\boldsymbol{\omega}^1 = \mathrm{d}x$, $\boldsymbol{\omega}^2 = \mathrm{d}y$, and $\boldsymbol{\omega}^3 = \mathrm{d}z$, so that

$$h_{ij}\mathrm{d}x^i\mathrm{d}x^j = a^2(t)\mathrm{d}x^2 + b^2(t)\mathrm{d}y^2 + c^2(t)\mathrm{d}z^2, \tag{3.15}$$

and the minisuperspace coordinates are $\{a, b, c, \Phi\}$. If we take the spatial directions to be compact such a universe will have toroidal topology. In the special case that $a(t) = b(t) = c(t)$ we retrieve the spatially flat $(k = 0)$ FRW universe.

The most complicated, and possibly the most interesting, Bianchi model is Bianchi IX, associated to the Lie group $SO(3, \mathbb{R})$. In this case the invariant 1-forms may be written as

$$\begin{aligned} \boldsymbol{\omega}^1 &= -\sin\psi\,\mathrm{d}\theta + \sin\theta\cos\psi\,\mathrm{d}\varphi, \\ \boldsymbol{\omega}^2 &= \quad\cos\psi\,\mathrm{d}\theta + \sin\theta\sin\psi\,\mathrm{d}\varphi, \\ \boldsymbol{\omega}^3 &= \quad\cos\theta\,\mathrm{d}\varphi + \mathrm{d}\psi, \end{aligned} \tag{3.16}$$

in terms of the Euler angles, (ψ, θ, φ), on the 3-sphere, S^3. The spatial sections of the geometry resulting from (3.14), (3.16) have the topology of S^3, but are only spherically symmetric in the special case that $h_{11}(t) = h_{12}(t) = \ldots h_{33}(t)$, which corresponds to the $k = +1$ FRW universe. Geometrically the spatial hypersurfaces can thus be thought of as squashed, twisted 3-spheres [see Fig. 3]. Bianchi IX has played an important role in classical cosmological studies of the initial singularity – it is the basis of the so-called "mixmaster universe"[80,81]. As a classical dynamical system Bianchi IX is extremely interesting because it appears to be chaotic, but only just on the verge of being so. Over the years there has been much debate as to whether Bianchi IX is or is not chaotic, and this seems to have been recently resolved by an explicit demonstration that it is not integrable■[82]. The corresponding minisuperspace model will have six independent coordinates in addition to the scalar field coordinate.

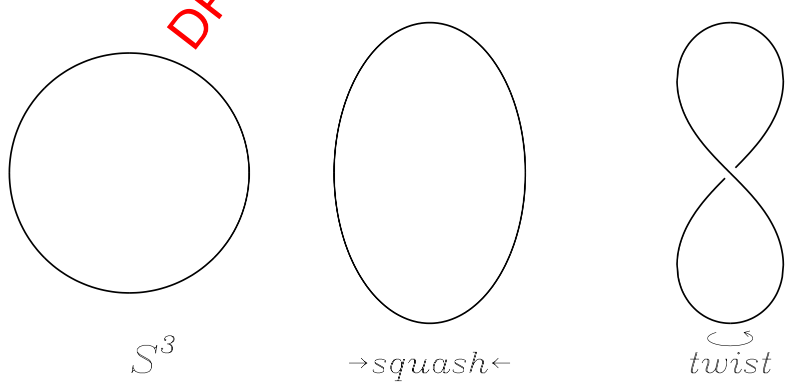

Fig. 3: Schematic geometry of spatial hypersurfaces in the Bianchi IX universe.

■ Technically speaking, what has been proved is the failure of integrability in the Painlevé sense. While this does not guarantee the existence of chaotic regimes, it does make their existence extremely probable.

Let us now consider the minisuperspace corresponding to an arbitrary homogeneous cosmology. We will assume that the minisuperspace is of dimension n, and will denote the minisuperspace coordinates by $\{q^{A}\}$. Since $\mathcal{N}^i = 0$ by assumption, it follows from the definitions (2.5) and (3.4) that

$$\mathcal{G}^{ijkl}\dot{h}_{ij}\dot{h}_{kl} = 4\sqrt{h}\mathcal{N}^2\left(K_{ij}K^{ij} - K^2\right), \tag{3.17}$$

and consequently the Lorentzian action (2.12) now takes the form

$$S = \int \mathrm{d}t\left[\frac{1}{2\mathcal{N}}\mathcal{G}_{AB}(q)\dot{q}^{A}\dot{q}^{B} - \mathcal{N}\mathcal{U}(q)\right], \tag{3.18}$$

where

$$\mathcal{G}_{AB}\mathrm{d}q^{A}\mathrm{d}q^{B} = \int \mathrm{d}^3x\left[\frac{1}{8\kappa^2}\mathcal{G}^{ijkl}\delta h_{ij}\delta h_{kl} + \sqrt{h}\delta\Phi\delta\Phi\right], \tag{3.19}$$

is the *minisupermetric*, which is now of finite dimension, n, and

$$\mathcal{U} = \int \mathrm{d}^3x\sqrt{h}\left[\frac{1}{4\kappa^2}\left(-{}^3R + 2\Lambda\right) + \mathcal{V}(\Phi)\right]. \tag{3.20}$$

The action (3.18) is simply equivalent to that for a "point particle" moving in a "potential" $\mathcal{U}$. Variation of (3.18) with respect to q^{A} thus yields a geodesic equation with force term,

$$\frac{1}{\mathcal{N}}\frac{d}{dt}\left(\frac{\dot{q}^{A}}{\mathcal{N}}\right) + \frac{1}{\mathcal{N}^2}\Gamma^{A}{}_{BC}\dot{q}^{B}\dot{q}^{C} = -\mathcal{G}^{AB}\frac{\partial\mathcal{U}}{\partial q^{B}}, \tag{3.21}$$

where $\Gamma^{A}{}_{BC}$ are the Christoffel symbols determined from the minisupermetric, while variation of (3.18) with respect to $\mathcal{N}$ yields the Hamiltonian constraint

$$\frac{1}{2\mathcal{N}^2}\mathcal{G}_{AB}(q)\dot{q}^{A}\dot{q}^{B} + \mathcal{U}(q) = 0. \tag{3.22}$$

The general solution to (3.21), (3.22) will have $2n - 1$ independent parameters, one of which is always trivial in the sense that it corresponds to a choice of origin of the time parameter. In studying any particular minisuperspace model we must take care to check that what we have done above is consistent, as it does not always follow that substituting a particular ansatz into an action before varying it will yield the same result as substituting the same ansatz into the field equations obtained from variation of the original action. Eqs. (3.21) and (3.22) should correspond respectively to the (ij) and (00) components of the original Einstein equations, while the $(0i)$ equation is trivially satisfied in the present case.

Quantisation is greatly simplified because now that our configuration space is finite-dimensional we are effectively dealing with the quantum *mechanics* of a constrained system. The canonical momenta and Hamiltonian are respectively given by

$$\pi_A = \frac{\partial L}{\partial \dot{q}^A} = \frac{\mathcal{G}_{AB}\dot{q}^B}{\mathcal{N}}\,, \tag{3.23}$$

$$H = \pi_A \dot{q}^A - L = \mathcal{N}\left[\tfrac{1}{2}\mathcal{G}^{AB}\pi_A\pi_B + \mathcal{U}(q)\right] \equiv \mathcal{N}\mathsf{H}. \tag{3.24}$$

The π_A are related to the canonical momenta (2.13)–(2.16) defined earlier by integration over the 3-volume of the hypersurfaces of homogeneity in (3.19). In terms of the new variables the action (3.18) and Hamiltonian constraint (3.22) are respectively

$$S = \int \mathrm{d}t \left[\pi_A \dot{q}^A - \mathcal{N}\mathsf{H}\right], \tag{3.25}$$

$$\tfrac{1}{2}\mathcal{G}^{AB}\pi_A\pi_B + \mathcal{U}(q) = 0. \tag{3.26}$$

Under canonical quantisation (3.26) yields the Wheeler-DeWitt equation

$$\hat{\mathsf{H}}\Psi = \left[-\tfrac{1}{2}\nabla^2 + \mathcal{U}(q)\right]\Psi = 0, \tag{3.27}$$

where

$$\nabla \equiv \frac{1}{\sqrt{-\mathcal{G}}}\partial_A\left[\sqrt{-\mathcal{G}}\,\mathcal{G}^{AB}\partial_B\right] \tag{3.28}$$

is the Laplacian operator of the minisupermetric. In arriving at (3.27) we have made an explicit "natural choice" of factor ordering[33,52] in order to accommodate the factor-ordering ambiguity. This choice is favoured by independent minisuperspace calculations of the prefactor, using zeta function regularisation and a scale invariant measure, which can then be related to factor ordering dependent terms through a semiclassical expansion of the Wheeler-DeWitt equation[83].

An alternative "natural choice"[35,84–86] of factor ordering would yield a *conformally invariant* Wheeler-DeWitt equation,

$$\hat{\mathsf{H}}\Psi = \left[-\tfrac{1}{2}\nabla^2 + \frac{n-2}{8(n-1)}\mathcal{R} + \mathcal{U}(q)\right]\Psi = 0, \tag{3.29}$$

where $\mathcal{R}$ is the scalar curvature obtained from the minisupermetric.

3.5. The WKB approximation

In view of the difficulties associated with solving the Wheeler-DeWitt equation in general, the best we can realistically hope for in many minisuperspace models is to look for appropriate approximate solutions in the semiclassical limit[‡], in which

$$\Psi \simeq \sum_n \Psi_n \equiv \sum_n \mathcal{A}_n \mathrm{e}^{-I_n}, \tag{3.30}$$

‡ The semiclassical limit of the full superspace Wheeler-DeWitt equation has been treated more formally by a number of authors. See, e.g., [87], [88] and references therein.

where the sum is over saddle points of the path integral, the $\mathcal{A}_n$ being appropriate (possibly complex) prefactors. In general we might expect to find regions in which the wavefunction is exponential, $\Psi \simeq \mathrm{e}^{-I}$, and regions in which it is oscillatory, $\Psi \simeq \mathrm{e}^{iS}$. The latter could be viewed as the wavefunction of a universe in the classical "Lorentzian" or "oscillatory" region, while the former would correspond to a universe in a classically inaccessible "Euclidean" or "tunneling" region. As has already been mentioned, the sum in (3.30) will in general contain a number of saddle points with an action, I_n, which is neither purely real nor purely imaginary.

Our own universe is of course Lorentzian at late times, and therefore the only minisuperspace models which can be of direct physical relevance are those for which the Wheeler-DeWitt equation does possess approximate solutions of the oscillatory type. Approximate solutions of this type can be obtained by performing a WKB expansion, for which purpose it is necessary to restore $\hbar$ in the minisuperspace Wheeler-DeWitt equation (3.27). If we assume that each component Ψ_n satisfies (3.27) separately, then

$$
\begin{aligned}
0 = \hat{\mathrm{H}}\Psi_n &= \left[-\tfrac{1}{2}\hbar^2\nabla^2 + \mathcal{U}\right]\mathcal{A}_n \mathrm{e}^{-I_n/\hbar} \\
&= \mathrm{e}^{-I_n/\hbar}\left\{\left[-\tfrac{1}{2}\left(\boldsymbol{\nabla} I_n\right)^2 + \mathcal{U}\right]\mathcal{A}_n + \hbar\left[\boldsymbol{\nabla} I_n \cdot \boldsymbol{\nabla}\mathcal{A}_n + \tfrac{1}{2}\mathcal{A}_n\nabla^2 I_n\right] + \mathrm{O}(\hbar^2)\right\},
\end{aligned}
\tag{3.31}
$$

where the dot implies contraction with the minisupermetric $\mathcal{G}^{AB}$. The $\mathrm{O}(\hbar^0)$ and $\mathrm{O}(\hbar)$ terms give two equations for I_n and $\mathcal{A}_n$. If we decompose I_n into real and imaginary parts according to $I_n = \mathsf{R}_n - i\mathsf{S}_n$ then the real and imaginary parts of the $\mathrm{O}(\hbar^0)$ term in (3.31) give

$$-\tfrac{1}{2}\left(\boldsymbol{\nabla}\mathsf{R}_n\right)^2 + \tfrac{1}{2}\left(\boldsymbol{\nabla}\mathsf{S}_n\right)^2 + \mathcal{U} = 0, \tag{3.32}$$

$$\boldsymbol{\nabla}\mathsf{R}_n \cdot \boldsymbol{\nabla}\mathsf{S}_n = 0. \tag{3.33}$$

Provided that the imaginary part of the action varies much more rapidly than the real part, i.e., $(\boldsymbol{\nabla}\mathsf{R}_n)^2 \ll (\boldsymbol{\nabla}\mathsf{S}_n)^2$, then (3.32) is the Lorentzian Hamilton-Jacobi equation for S_n:

$$\tfrac{1}{2}\mathcal{G}^{AB}\frac{\partial\mathsf{S}_n}{\partial q^A}\frac{\partial\mathsf{S}_n}{\partial q^B} + \mathcal{U}(q) = 0. \tag{3.34}$$

Comparison of (3.34) with (3.27) suggests a strong correlation between coordinates and momenta, and invites the identification

$$\pi_A = \frac{\partial\mathsf{S}_n}{\partial q^A}. \tag{3.35}$$

If we differentiate (3.34) w.r.t. q^C we obtain

$$\tfrac{1}{2}\mathcal{G}^{AB}{}_{,C}\frac{\partial\mathsf{S}_n}{\partial q^A}\frac{\partial\mathsf{S}_n}{\partial q^B} + \mathcal{G}^{AB}\frac{\partial\mathsf{S}_n}{\partial q^A}\frac{\partial^2\mathsf{S}_n}{\partial q^B\partial q^C} + \frac{\partial\mathcal{U}}{\partial q^C} = 0. \tag{3.36}$$

If we define a minisuperspace vector field

$$\frac{\mathrm{d}}{\mathrm{d}s} \equiv \mathcal{G}^{AB} \frac{\partial S_n}{\partial q^A} \frac{\partial}{\partial q^B}, \tag{3.37}$$

then combining (3.35), (3.36) and (3.37) we obtain

$$\frac{\mathrm{d}\pi_C}{\mathrm{d}s} + \tfrac{1}{2}\mathcal{G}^{AB}{}_{,C}\, \pi_A \pi_B + \frac{\partial \mathcal{U}}{\partial q^C} = 0, \tag{3.38}$$

which after raising indices is the same geodesic equation (3.21) obtained earlier provided we identify the parameter, s, with the proper time on the geodesics.

We can now solve the equation given by the $O(\hbar)$ term of (3.31). Since $|\nabla R_n| \ll |\nabla S_n|$ it follows that the terms involving R_n can be neglected, and thus

$$\nabla S_n \cdot \nabla \mathcal{A}_n \equiv \mathcal{G}^{AB} \frac{\partial S_n}{\partial q^A} \frac{\partial \mathcal{A}_n}{\partial q^B} \equiv \frac{\mathrm{d}\mathcal{A}_n}{\mathrm{d}s} = -\tfrac{1}{2}\mathcal{A}_n \nabla^2 S_n, \tag{3.39}$$

which may be readily integrated. We thus obtain a first order WKB wavefunction

$$\Psi_n = \mathcal{C}_n \exp\left(iS_n - \tfrac{1}{2} \int \mathrm{d}s\, \nabla^2 S_n \right), \tag{3.40}$$

where $\mathcal{C}_n$ is an arbitrary (complex) constant to be appropriately normalised, and we have reverted to natural units in which $\hbar = 1$.

The wavefunction (3.40) could be considered to be the analogue of the wavefunction for coherent states in ordinary quantum mechanics,

$$\psi_n(x,t) = c_n e^{ip_n x} \exp\left(\frac{-(x - \bar{x}_n(t))^2}{\sigma^2} \right), \tag{3.41}$$

which describes a wave packet which is "peaked" about a classical particle trajectory, $\bar{x}_n(t)$, and which thus roughly "predicts" classical behaviour. This becomes problematic, however, if we consider a superposition, $\psi = \sum_n \psi_n$, of such states since interference between different wave packets will in general destroy the classical behaviour. In order to interpret the total wavefunction as saying that the particle follows a roughly classical trajectory, $\bar{x}_n(t)$, with probability $|c_n|^2$, it is necessary that a *decoherence* mechanism should exist which renders this quantum mechanical interference negligible[89].

The issue of quantum decoherence is clearly also of great importance in quantum cosmology, since in order to interpret Ψ in (3.30) in a similar fashion a similar mechanism must exist. It has been recently argued, furthermore, that decoherence is a necessary feature of the WKB interpretation of quantum cosmology, since without decoherence the existence of chaotic cosmological solutions would lead to a breakdown of the WKB approximation[90]. This is analogous to similar problems with the commutativity of the limits $t \to \infty$ and $\hbar \to 0$ in ordinary quantum mechanics when applied to chaotic systems.

The issues involved in decoherence pose complex conceptual questions for the fundamentals of quantum mechanics itself, quite apart from the problems specific to quantum cosmology[⊠]. Here we will merely assume that such a mechanism exists,

⊠ For further details see, e.g., [91]–[94] and references therein.

and we will take the view that Ψ can be considered to "predict" a classical spacetime if there exist WKB-type solutions (3.40), which yield a strong correlation between π^A and q^A according to (3.35). The sense in which the minisuperspace positions and momenta are "strongly correlated" can be made more precise through the use of quantum distribution functions, such as the Wigner function[91,95]. By use of the Wigner function one may show[95] that wavefunctions of the oscillatory type, $\Psi \sim \mathrm{e}^{i\mathrm{S}}$, predict a strong correlation between coordinates and momenta, whereas wavefunctions of the type $\Psi \sim \mathrm{e}^{-\mathrm{I}}$, which are also typical minisuperspace solutions, do not. Such exponential wavefunctions can thus be considered as describing universes in a purely quantum "tunneling" regime, before the quantum to classical transition. We will interpret wavefunctions, $\Psi \sim \mathrm{e}^{i\mathrm{S}}$, as corresponding to classical spacetime, or rather a set of classical spacetimes as S is a first integral of the equations of motion.

3.6. Probability measures

Given a solution, Ψ, to the Wheeler-DeWitt equation it is necessary to construct a probability measure in order to make predictions. One central question in quantum cosmology is how one should construct such a measure.

The minisuperspace Wheeler-DeWitt equation (3.27) is a second-order equation very much like the Klein-Gordon equation in ordinary field theory, and it readily follows from (3.27) that the current[33]

$$\mathcal{J} = -\tfrac{1}{2}i\left(\overline{\Psi}\boldsymbol{\nabla}\Psi - \Psi\boldsymbol{\nabla}\overline{\Psi}\right) \tag{3.42}$$

is conserved: $\boldsymbol{\nabla}\cdot\mathcal{J} = 0$. The similarity to the Klein-Gordon current extends to the fact that the natural inner product[33] constructed from $\mathcal{J}$ is not positive-definite and so gives rise to negative probabilities. In quantum field theory this is not a problem since one can split the wavefunction up into positive and negative frequency components which correspond to particles and anti-particles. However, as has already been mentioned there is no well-defined notion of positive frequencies in superspace on account of its lack of symmetries[96]. A further problem is that many natural wavefunctions would have zero norm with this definition. For example, the no-boundary wavefunction is real and gives $\mathcal{J} = 0$.

The similarity of Ψ to the Klein-Gordon field has suggested to many people that one should turn Ψ into an operator, $\hat{\Psi}$, thereby introducing quantum field theory on superspace, or "third quantisation". One then arrives at operators which create and annihilate universes. However, as we do not perform measurements over a statistical ensemble of universes it is not clear how we can arrive at sensible probabilities using such a formulation.

The difficulties with the Klein-Gordon current of course led Dirac to introduce the Dirac equation, and it is worth mentioning that a similar resolution of the problem is available in supersymmetric quantum cosmology. In particular, one can go to a theory which includes fermionic variables by considering quantum cosmology based on supergravity◀ rather than the purely bosonic Einstein theory. The constraints of supergravity, which may be viewed as the Dirac square root of the constraints of general relativity[98,99], are reducible to first-order equations. Furthermore, this also translates into simplifications in homogeneous minisuperspace models – the appropriate constraint equation which determines the quantum evolution of the wavefunction can be considered to be the Dirac square root of the Wheeler-DeWitt equation[100–103]. As a result it is possible to construct[100] a Dirac-type probability density which is conserved by the equation $\mathcal{Q}\Psi = 0$, where $\mathcal{Q}$ is the supercharge.

Another alternative to the question of the probability measure is to use $|\Psi|^2$ directly as a probability measure[52,104], by defining the probability of the universe being in a region, $\mathcal{A}$, of superspace by

$$\mathcal{P}(\mathcal{A}) \propto \int_{\mathcal{A}} |\Psi|^2 \, {}^*1 \qquad (3.43)$$

where ^{*1} is the volume-element on superspace, $*$ being the Hodge dual in the supermetric. This definition of a necessarily positive-definite probability density works very well for homogeneous minisuperspaces, for which the volume form ^{*1} is independent of $x \in \Sigma$. This is perhaps not surprising since as was observed in §3.4 the assumption of homogeneity reduces the problem to one of quantum mechanics, and $|\Psi|^2$ is of course the probability density in conventional quantum mechanics.

Problems with the definition (3.43) do arise since even in some simple examples the wavefunction is not normalisable, but instead $\langle\Psi|\Psi\rangle = \infty$. One further problem is that whereas in ordinary quantum mechanics $|\Psi|^2$ describes a probability density in configuration space – i.e., the space of particle positions – in quantum cosmology the configuration space is (mini)superspace and time is implicitly contained in the (mini)superspace coordinates. These coordinates cannot therefore be thought of as the mere analogues of particle positions. As a result the recovery of the conservation of probability and the standard interpretation of the quantum mechanics for small subsystems is not necessarily straightforward in the approach based on (3.43), and may involve understanding some subtle questions about the role of time in quantum gravity.

Ultimately a formulation such as (3.43) which is based on absolute probabilities may not be required since it is impossible to measure statistical ensembles of universes and thus all we can really test are *conditional probabilities* rather than absolute probabilities. For example, a relevant testable probability might be the probability, $\mathcal{P}(\mathcal{A}|\mathcal{B})$, of finding Ψ in a region $\mathcal{A}$ of superspace given that Ψ started in another region $\mathcal{B}$ of superspace. Page[2,105] has explored the construction of conditional probabilities in quantum cosmology without the use of absolute probabilities.

◀ In fact, a naïve first-order Hamiltonian formulation for the minisuperspaces of homogeneous cosmologies was found early on [97], but until the development of supergravity there was no natural interpretation of the Dirac-type constraint equation obtained.

Clearly the issues surrounding the choice of probability measure involve some deep conceptual problems which may perhaps get to the heart of the broader conceptual basis of quantum gravity. Such issues have been discussed by a number of authors in the context of quantum cosmology[2,43,104–107] and I will not address them here in any detail.

For the purposes of examining how we might hope to make predictions from the proposed boundary conditions of §4.1,4.2 we shall merely consider quantum cosmology in the WKB limit. In this limit it follows from (3.40) that each of the components Ψ_n in (3.30) has a conserved Klein-Gordon-type current (3.42) given by

$$\mathcal{J}_n \simeq |\mathcal{A}_n|^2 \nabla \mathsf{S}_n, \tag{3.44}$$

which flows very nearly along the direction of the classical trajectories. The current conservation law $\nabla_{\!A}\, \mathcal{J}_n{}^A = 0$ implies that

$$\mathrm{d}\mathcal{P} = \mathcal{J}_n{}^A \mathrm{d}\Sigma_{A}, \tag{3.45}$$

is a conserved probability measure on the set of trajectories with tangent $\nabla \mathsf{S}_n$, where $\mathrm{d}\Sigma_A$ is the element of a hypersurface, Σ, in minisuperspace which cuts across the flow and intersects each curve in the congruence once and only once. One finds that for a pencil of trajectories near the classical trajectory the probability density (3.45) is positive-definite. Vilenkin has argued[104] that positive-definiteness of the probability measure is really only required in the semiclassical limit, as this is the only limit in which we obtain a universe accessible to observation where the conventional laws of physics apply. Therefore given that the current (3.42) works in the WKB limit, this is all that is needed if we are content that "probability" and "unitarity" are only approximate concepts in quantum gravity. One can also show[2,52] that the manifestly positive definition (3.43) yields essentially the same result as (3.45) in the WKB limit.

3.7. Minisuperspace for the Friedmann universe with massive scalar field

Let us now apply our results to the particularly simple case of a homogeneous, isotropic universe with a single scalar field, with a potential which allows for inflationary behaviour. A quadratic potential is possibly the simplest example with this property, and thus has been much studied in quantum cosmology.

For convenience we will introduce a numerical normalisation factor $\sigma^2 = \kappa^2/(3\mathsf{V})$ into the metric, where V is the 3-volume of the unit hypersurface – e.g., $\mathsf{V} = 2\pi^2$ for the 3-sphere, $k = +1$. We are considering *closed universes only*, which requires making topological identifications for the $k = 0, -1$ cases, so that V remains finite. In place of (2.1), (2.2) and (3.12) we then have a metric

$$\mathrm{d}s^2 = \sigma^2 \left\{ -\mathcal{N}^2 \mathrm{d}t^2 + a^2(t) \left[\frac{\mathrm{d}r^2}{1 - kr^2} + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\varphi^2 \right) \right] \right\}. \tag{3.46}$$

We also take the scalar field action to be normalised by

$$S_{\text{matter}} = \frac{3}{\kappa^2} \int_{\mathcal{M}} \mathrm{d}^4x \sqrt{-g} \left(-\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - \frac{V(\phi)}{2\sigma^2} \right). \tag{3.47}$$

Thus $\mathcal{N} \to \sigma\mathcal{N}$, $a \to \sigma a$, $\Phi = \sqrt{3}\phi/\kappa$ and $\mathcal{V} = 3V/(2\kappa^2\sigma^2)$ relative to our earlier definitions, and a, ϕ and V are now dimensionless. As for the general homogeneous minisuperspace discussed in the previous section, we use a gauge in which $\mathcal{N}_i = 0$. From (3.46) it follows that ${}^3R = \frac{6k}{\sigma^2 a^2}$ and $K_{ij} = -\frac{\dot{a}}{\sigma \mathcal{N} a} h_{ij}$, and thus the action takes the form (3.18) with a minisupermetric

$$\mathcal{G}_{AB} \mathrm{d}q^A \mathrm{d}q^B = -a \mathrm{d}a^2 + a^3 \mathrm{d}\phi^2, \tag{3.48}$$

and potential

$$\mathcal{U} = \tfrac{1}{2} \left(a^3 V(\phi) - ka \right). \tag{3.49}$$

Alternatively, it is sometimes useful to express (3.48) in terms of a conformal gauge

$$\begin{aligned} \mathcal{G}_{AB} \mathrm{d}q^A \mathrm{d}q^B &= \mathrm{e}^{3\alpha} \left(-\mathrm{d}\alpha^2 + \mathrm{d}\phi^2 \right), & (3.50) \\ &= -(4uv)^{-1/4} \mathrm{d}u \mathrm{d}v, & (3.51) \end{aligned}$$

where $\alpha = \ln a$, or alternatively in null coordinates

$$\begin{aligned} u &= \tfrac{1}{2} \mathrm{e}^{2(\alpha - \phi)} = \tfrac{1}{2} a^2 \mathrm{e}^{-2\phi}, \\ v &= \tfrac{1}{2} \mathrm{e}^{2(\alpha + \phi)} = \tfrac{1}{2} a^2 \mathrm{e}^{2\phi}. \end{aligned} \tag{3.52}$$

The canonical momenta are given by

$$\pi^0 = 0, \qquad \pi_a = \frac{-a\dot{a}}{\mathcal{N}}, \qquad \pi_\phi = \frac{a^3 \dot{\phi}}{\mathcal{N}}, \tag{3.53}$$

and the classical equations of motion yield

$$\mathsf{H} = \tfrac{1}{2} \left(-\frac{a\dot{a}^2}{\mathcal{N}^2} + \frac{a^3 \dot{\phi}^2}{\mathcal{N}^2} - ka + a^3 V \right) = 0, \tag{3.54}$$

$$\frac{1}{\mathcal{N}} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\dot{a}}{\mathcal{N}} \right) + \frac{2a\dot{\phi}^2}{\mathcal{N}^2} - aV = 0, \tag{3.55}$$

$$\frac{1}{\mathcal{N}} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\dot{\phi}}{\mathcal{N}} \right) + 3 \frac{\dot{a}\dot{\phi}}{a\mathcal{N}^2} + \tfrac{1}{2} \frac{\mathrm{d}V}{\mathrm{d}\phi} = 0, \tag{3.56}$$

which are equivalent to the geodesic equation (3.21) with metric (3.48) and potential (3.49). The lapse function is not of physical relevance classically since we can choose an alternate proper time parameter, $\mathrm{d}\tau = \mathcal{N}\mathrm{d}t$, or equivalently choose a gauge $\mathcal{N} = 1$ in (3.54)–(3.56) so that t is the proper time. One of these equations depends on the other two by virtue of the Bianchi identity, as is always the case in general relativity. We thus effectively have two independent differential equations in two unknowns, one first order and one second order, or equivalently an autonomous system of three first order differential equations. The solution therefore depends on three free parameters, but as mentioned above one of these amounts to a choice of the origin of time which is of no physical importance. Thus there is a two-parameter family of physically distinct solutions.

The classical solutions cannot be written in a simple closed form except for certain special values of k and the potential $V(\phi)$, which unfortunately does not even include the quadratic potential▷ $V(\phi) = m^2\phi^2$. The qualitative property of the solutions may nonetheless be determined by studying the 3-dimensional phase space. Instead of choosing a particular potential, however, let us suppose that we are in a region of the phase space for which V can be approximated by a constant. Such conditions are more or less met in the "slow-rolling approximation"[109] of inflationary cosmology, in which $|V'/V| \ll 6$ and $|V''/V| \ll 9$. In this approximation the dynamics is described by setting $\dot{\phi} \simeq 0$ in (3.54) and (3.55), and setting $\ddot{\phi} \simeq 0$ in (3.56). In approximating the nearly flat region of the potential V by a cosmological constant we ignore the slow change of the scalar field determined from (3.56). We thus obtain a simplified model which possesses classical inflationary solutions, provided the constant V is chosen to be *positive*. This will serve as a useful test model for quantum cosmology.

In the case that V is constant, eq. (3.56) integrates to give $\dot{\phi} = Ca^{-3}$, where C is an arbitrary constant, and it therefore follows that the Friedmann equation (3.54) can be written in terms of an elliptic integral in $a^2(\eta)$,

$$\eta = \frac{1}{2}\int_0^{a^2} \frac{\mathrm{d}z}{\sqrt{Vz^3 - kz^2 + C^2}}\,, \tag{3.57}$$

where η is the conformal time parameter defined by $\mathrm{d}\eta = a^{-1}\mathrm{d}t$. It is thus possible to express the general solution in terms of elliptic functions. Since the properties of such functions are not very transparent perhaps, we can alternatively plot the 2-dimensional phase space – e.g., in terms of the variables $\dot{\phi}$ and $\dot{\alpha}$, as in Fig. 4 in order to understand the qualitative features of the solutions. There are four critical points, at $(\dot{\phi}, \dot{\alpha}) = \left\{\left(0, \pm 1/\sqrt{V}\right), \left(\pm 1/\sqrt{2V}, 0\right)\right\}$, the first two being nodes, $A_\pm$, which are endpoints for all values of the spatial curvature, k, and the latter two saddle points, $B_\pm$, for the $k = +1$ solutions.

The general solution for the spatially flat case ($k = 0$), which corresponds to the bold separatrices shown in Fig. 4, is given by[110]

$$\begin{aligned} a &= C_a \left|\sinh \tfrac{3}{2}\sqrt{V}t\right|^{1/3} \left|\cosh \tfrac{3}{2}\sqrt{V}t\right|^{1/3}, \\ \phi &= \tfrac{1}{3}\ln\left|\tanh \tfrac{3}{2}\sqrt{V}t\right| + C_\phi, \end{aligned} \tag{3.58}$$

▷ The Einstein equations for the FRW universe with a massive scalar field can be solved approximately [108] in various limits, however.

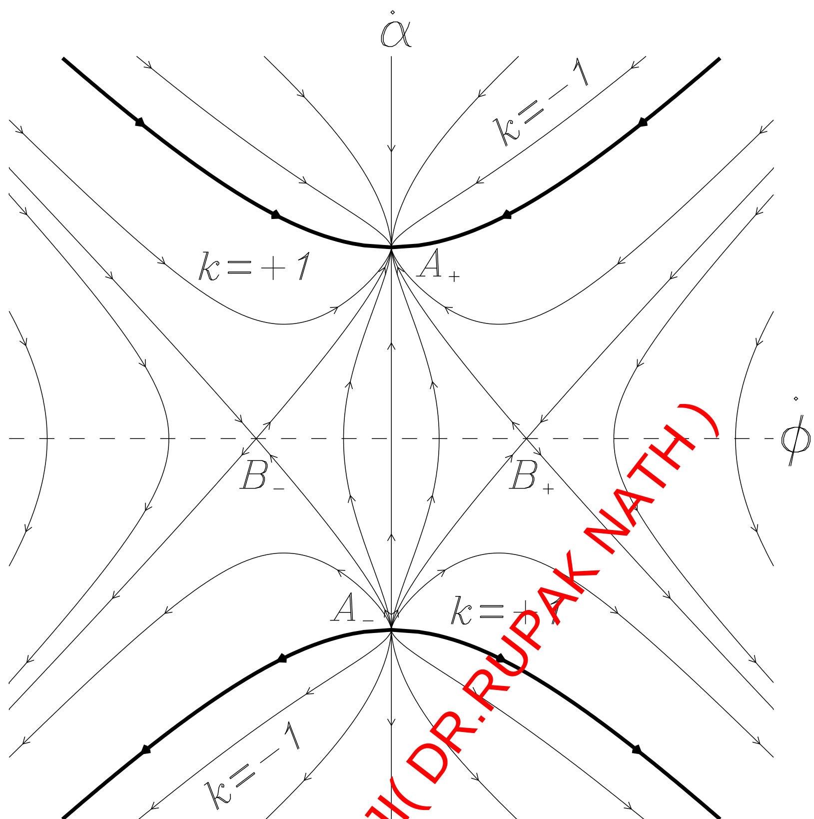

Fig. 4: The 2-dimensional phase plot of $\dot{\alpha} = \dot{a}/a$ versus $\dot{\phi}$ for the simplified model with constant potential, V.

in closed form, where C_a and C_ϕ are arbitrary constants. At late times the solutions (3.58) have an exponential scale factor, $a \to C_a \exp(\sqrt{V}\,t)$ as $t \to \infty$, and constant scalar field $\phi \to C_\phi$. Furthermore, one can see from Fig. 4 that all the $k = -1$ solutions in the upper half-plane, and a number of the $k = +1$ solutions are also attracted to the point A_+ with a similar inflationary behaviour at late times. (The corresponding point A_- on the $k = 0$ separatrix corresponds to the time-reversed solution, with an inflationary phase as $t \to -\infty$.) The simplified universe corresponding to Fig. 4 of course is far from being the complete picture, as the model does not allow for any exit from inflation. However, Fig. 4 illustrates the typical situation that a given model will possess regimes with inflationary behaviour and regimes with non-inflationary behaviour. In Fig. 4 the $k = +1$ solutions to the right and left of the separatrices that pass through $B_\pm$ fall into the latter category, for example. The situation becomes even more involved when one considers the full 3-dimensional phase space for some particular potential $V(\phi)$.

The case of the $k = +1$ solutions in Fig. 4 illustrates the general feature that classical dynamics are highly dependent on initial conditions. In order to obtain a sufficiently long inflationary epoch to overcome the problems mentioned in the Introduction, (of order 65 e-folds growth in the scale factor), the initial values of ϕ

and $\dot\phi$ must be restricted to a particular region of the phase space. In particular, $\dot\phi$ must be small initially. Classically, there is no *a priori* reason for one choice of initial conditions over any other choice, unless further ingredients are added. The degree to which inflationary initial conditions are preferred relative to other initial conditions – i.e., how probable is inflation? – is precisely the sort of question that we might therefore hope quantum cosmology could answer.

It is possible to attempt to solve this question without resorting to quantum cosmology. To do this one must construct a measure on the set of all universes[111,112], and then compare the number of inflationary solutions with a sufficiently long exponential phase to the number of other solutions. Preliminary results[112] seemed to indicate that almost all models with a massive scalar field undergo a period of inflation. However, a more careful analysis[108] revealed that the answer is ambiguous, as both the set of inflationary solutions and the set of non-inflationary solutions have infinite measure.

Alternatively, if as we expect the universe began in some sort of tunneling process or similar transition from a quantum regime, then we could expect the "initial" classical parameters to be determined, at least in a probabilistic fashion, from more fundamental quantum processes. The question of the most probable state of the universe is then pushed back a level and becomes: "what is a typical wavefunction for the universe?"

In the context of the present minisuperspace model, therefore, we can proceed by quantising the Wheeler-DeWitt equation (3.54), to obtain

$$\begin{aligned}\hat{\mathrm{H}}\Psi = \left[-\tfrac{1}{2}\nabla^2 + \mathcal{U}\right]\Psi &= -\tfrac{1}{2}\left[\frac{1}{a^3}\left(a\frac{\partial}{\partial a}\,a\frac{\partial}{\partial a} - \frac{\partial^2}{\partial\phi^2}\right) - ka + a^3V(\phi)\right]\Psi \\ &= \tfrac{1}{2}\mathrm{e}^{-3\alpha}\left[\frac{\partial^2}{\partial\alpha^2} - \frac{\partial^2}{\partial\phi^2} - k\mathrm{e}^{4\alpha} + \mathrm{e}^{6\alpha}V(\phi)\right]\Psi \\ &= (4uv)^{1/4}\left[2\frac{\partial^2}{\partial u\partial v} - \frac{k}{2} + (uv)^{1/2}V\right]\Psi \\ &= 0\end{aligned} \tag{3.59}$$

in terms of the various sets of coordinates given earlier. In general boundary conditions will have to specified in order to solve (3.59). However, we can consider the approximate form of the WKB solutions without considering boundary conditions for the time being.

We will confine ourselves to regions in which the potential can be approximated by a cosmological constant, as in the analysis of Fig. 4, so that we can drop the term involving derivatives with respect to ϕ in (3.59), thereby obtaining a simple 1-dimensional problem which is amenable to a standard WKB analysis. The first order WKB wavefunction (3.40) which solves (3.59) in this approximation is

$$\Psi(a,\phi) \simeq \begin{cases} \dfrac{\mathcal{B}(\phi)}{a\left(a^2V(\phi)-k\right)^{1/4}} \exp\left[\dfrac{\pm i}{3V(\phi)}\left(a^2V(\phi)-k\right)^{3/2}\right], & a^2V > k, \quad (3.60) \\ \dfrac{\mathcal{C}(\phi)}{a\left(k-a^2V(\phi)\right)^{1/4}} \exp\left[\dfrac{\pm 1}{3V(\phi)}\left(k-a^2V(\phi)\right)^{3/2}\right], & a^2V < k. \quad (3.61) \end{cases}$$

If V is positive, as was assumed above, then oscillatory type solutions will thus exist for large values of the scale factor, while the exponential type solutions will exist for small values of the scale factor if $k = +1$.

The oscillatory solutions are of the form $\Psi \sim e^{i\mathsf{S}}$ (neglecting the prefactor), where S satisfies the Hamilton-Jacobi equation (3.34). Comparing this to the Hamiltonian constraint (3.26) we find a strong correlation (3.35) between momenta and coordinates. For large scale factors, $a^2V \gg |k|$, so that $\mathsf{S} \simeq \pm\frac{1}{3}a^3\sqrt{V}$. In this limit (3.23), (3.35) and (3.50) thus yield

$$\begin{aligned} \pi_\alpha &= \frac{\partial \mathsf{S}}{\partial \alpha} \quad \Rightarrow \quad \dot{\alpha} \simeq \pm\sqrt{V}, \\ \pi_\phi &= \frac{\partial \mathsf{S}}{\partial \phi} \quad \Rightarrow \quad \dot{\phi} \simeq 0, \end{aligned} \tag{3.62}$$

which correspond in fact to the inflationary points, $A_\pm$, of Fig. 4. The oscillatory wavefunction thus "picks out" classical inflationary universes.

Since the minisupermetric (3.50) is conformal to 2-dimensional Minkowski space in the coordinates (α, ϕ), it is convenient to represent it by a Carter-Penrose conformal diagram (see Fig. 5). In each case we plot $(p-q)$ horizontally and $(p+q)$ vertically, where $\tan p = \alpha + \phi$, and $\tan q = \alpha - \phi$. The boundary consists of points corresponding to past timelike infinity, $\mathfrak{i}^- = \{(a,\phi)\,|\,a=0, \phi \text{ finite}\}$, future timelike infinity, $\mathfrak{i}^+ = \{(a,\phi)\,|\,a=\infty, \phi \text{ finite}\}$, left and right spacelike infinity, $\mathfrak{i}^0_{L,R} = \{(a,\phi)\,|\,a = \text{ finite}, \phi = \pm\infty\}$; and past and future null boundaries, $\mathcal{I}^-_{L,R} = \{(a,\phi)\,|\,a=0, \phi=\pm\infty\}$ and $\mathcal{I}^+_{L,R} = \{(a,\phi)\,|\,a=0, \phi=\pm\infty\}$. In each case the subscript L (left) is associated with $\phi \to -\infty$, and the subscript R (right) with $\phi \to +\infty$. The approximate region for which oscillatory WKB solutions exist is shown in Fig. 5(a,b) for the approximate minisuperspace with a cosmological constant, in Fig. 5(c,d) for $V(\phi) = m^2\phi^2$, and in Fig. 5(e,f) for potentials, $V(\phi)$, typically found in higher-derivative gravity theories and in string theory with supersymmetry breaking.

Naturally it is of interest to know whether the inflationary WKB wavefunctions are *typical* solutions to the Wheeler-DeWitt equation. To determine a typical wavefunction for the universe, we need to make a choice of boundary conditions for Ψ when solving (3.59).

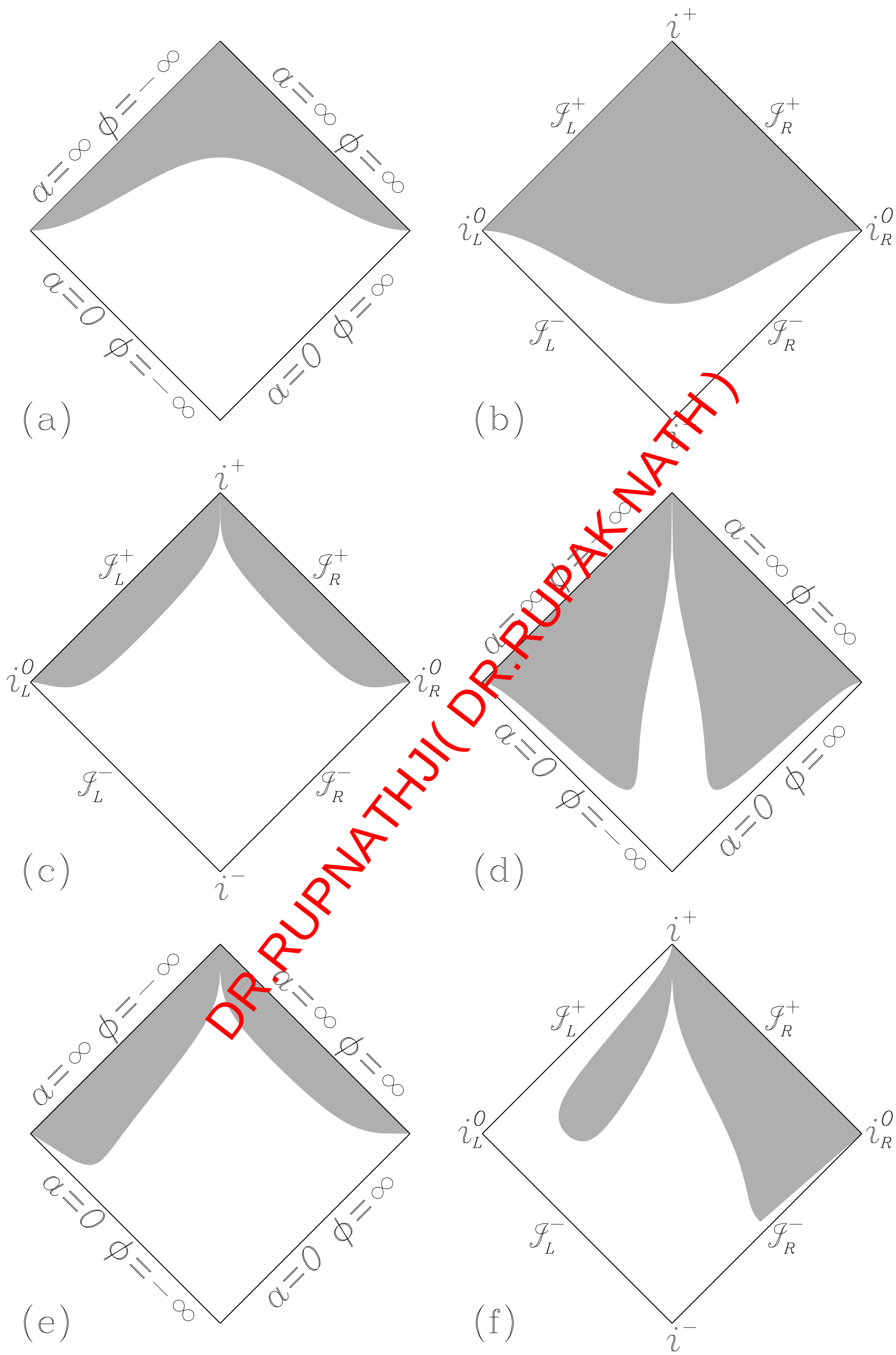

Fig. 5: Conformal diagrams of the 2-dimensional minisuperspace. The region where oscillatory WKB solutions exist, as given by the rough criterion $a^2V > 1$, is shaded for various potentials: (a) $V = 0.25$ (const); (b) $V = 4$ (const); (c) $V = 0.25\phi^2$; (d) $V = 25\phi^2$; (e) $V = \left[1 - \mathrm{e}^{-\phi/f}\right]^2$ with $f = 1.5$; (f) $V = 4\sinh^2\phi\exp\left[-f\mathrm{e}^{-2\phi}\right]$ with $f = 0.1$.

4. Boundary Conditions

The specification of boundary conditions for the Wheeler-DeWitt equation may seem a disappointment, as it might appear that we are just replacing an arbitrary initial choice of parameters which describe the classical evolution of the universe by an arbitrary initial choice of parameters which describe its quantum evolution. However, if quantum mechanics is a universal theory then it must have applied at the earliest epochs of the existence of the universe, in which case it is natural that the quantum dynamics precedes the classical dynamics. This justifies a quantum boundary condition for the universe as being more fundamental than a classical one. In any case, the only alternative to choosing quantum boundary conditions would be that mathematical consistency might be enough to guarantee a unique solution to the Wheeler-DeWitt equation, as DeWitt originally hoped[33]. If the experience gained from the study of minisuperspace models translates to superspace, then this would not appear to be the case, however.

The question naturally arises as to whether there should be some natural boundary condition, which once and for all determines the quantum evolution of the universe at early times, or alternatively whether the nature of the quantum dynamics might be somewhat indifferent to such choices. Deep conceptual problems are involved in trying to make headway with this question. Unlike other situations in quantum physics, where boundary conditions are readily specified by the symmetry of particular problems, such as spherical symmetry in the case of the hydrogen atom, the origin of the universe poses a situation in which all intuition must be abandoned and we can at best proceed on aesthetic grounds alone.

Having made a choice of boundary condition, we can of course solve the Wheeler-DeWitt equation and study the physical consequences for the evolution of the universe. However, without some additional principle we should by rights study many different boundary conditions before we can begin to have any confidence about the predictions made. To arrive at a principle which would circumvent this problem is an immense challenge: it would more or less amount to an additional law of physics which must be appended to the others which describe the quantum evolution of the universe. The situation might be considered to be the same as trying to describe the phase transition from gas to liquid if all the physical phenomena that we knew about related to the gaseous phase only. The universe appears to have undergone a phase transition when it was formed, but the only experience we have available involves the "after" state of the universe alone.

Progress can of course only be made by attempting to define natural boundary conditions for the wavefunction of the universe, and examining the consequences. This became an important activity in the 1980s. I will only discuss the two most studied boundary condition proposals, the "no-boundary proposal" and the "tunneling proposal". However, other proposals have been put forward, including the "all possible boundaries proposal" of Suen and Young[113] and the "symmetric initial condition" of Conradi and Zeh[114,115].

4.1. The no-boundary proposal

The proposal of Hartle and Hawking[38,39] is that one should restrict the sum in the definition of the wavefunction of the universe (3.11) to include only compact Euclidean 4-manifolds, $\mathcal{M}$, for which the spatial hypersurface Σ on which Ψ is defined forms the only boundary, and only matter configurations which are regular on these geometries. The universe then has no singular boundary to the past, as is the case for the standard FRW cosmology. The sum (3.11) thus includes manifolds such as those shown in Fig. 6, but not those shown in Fig. 2. As Hawking[39] puts it: *the boundary conditions of the universe are that it has no boundary.*

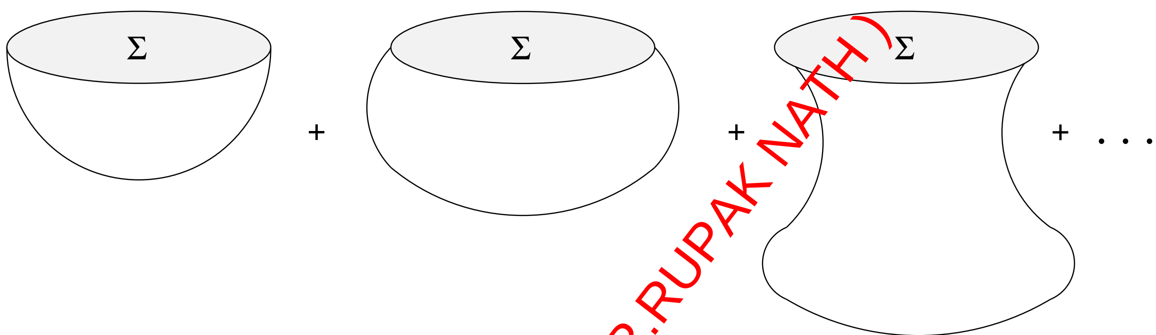

Fig. 6: Geometries allowed by the Hartle-Hawking no-boundary proposal.

Intuitively, what Hartle and Hawking had in mind in formulating this proposal was to get rid of the initial singularity by "smoothing the geometry of the universe off in imaginary time". For example, whereas a surface with $\sqrt{h} = 0$ would be singular in a Lorentzian signature metric, this is not necessarily the case if the metric is of Euclidean signature, as can be seen from the example of S^4 shown in Fig. 7. Ideally, the no-boundary proposal should tell us what initial conditions to set when we take manifolds with $\sqrt{h} \to 0$ or any similar limit consistent with the proposal. If the limit is taken at an initial time $\tau = 0$, the no-boundary proposal would lead to conditions on $h_{ij}(x, 0)$, $\phi(x, 0)$ and their derivatives. In practice, quantum cosmology is rarely studied beyond the semiclassical approximation, in which $\Psi \simeq \mathcal{A} e^{-I_{\rm cl}}$, where $I_{\rm cl}$ is the classical (possibly complex) action evaluated along the solution to the Euclidean field equations. In the semiclassical approximation one therefore works only with boundary conditions on the metric and matter fields which correspond to the no-boundary proposal at the classical level. In particular, we demand: (i) that the 4-geometry is closed; and (ii) that the saddle points of the functional integral correspond to regular solutions of the classical field equations which match the prescribed initial data on Σ.

One question which is not explained by the no-boundary proposal is the choice of a contour of integration for the path integral. As was mentioned earlier, the path integral over real Euclidean metrics does not converge, and thus it is necessary to include complex metrics to make the path integral converge. Such metrics will generally include ones which are neither truly Euclidean nor truly Lorentzian, and thus

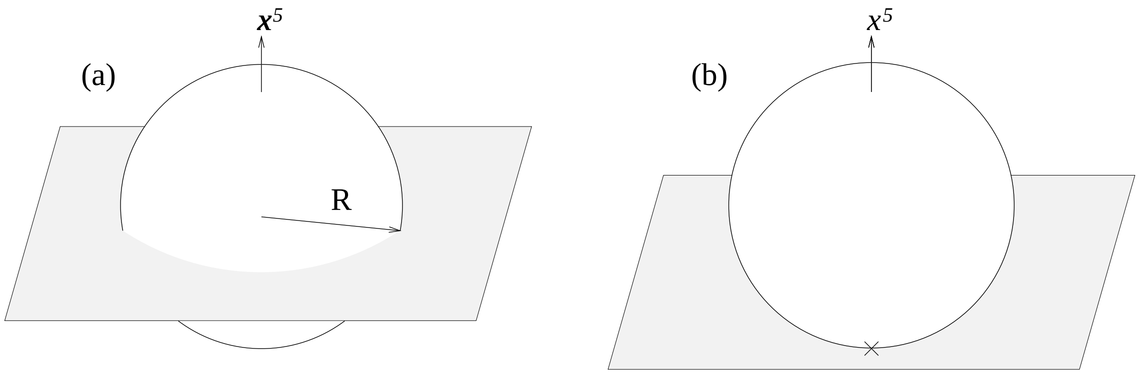

Fig. 7: Slicing a 4-sphere of radius R embedded in flat 5-dimensional space: (a) a surface $|x^5| = a < R$ intersects the 4-sphere in a 3-sphere of non-zero radius; (b) when $|x^5| = R$ the 3-sphere shrinks to zero radius but there is no singularity of the 4-geometry.

the naïve picture of a compact Euclidean geometry sewn onto a Lorentzian one, which is suggested by "smoothing the geometry of the universe off in imaginary time", is not completely accurate. In general, one might expect a truly complex metric to interpolate the Euclidean and Lorentzian ones, and in general the initial geometry might be only *approximately Euclidean* and the final geometry only *approximately Lorentzian*[116]. Unfortunately the criteria for achieving convergence of the path integral do not single out a unique contour of integration, and the no-boundary proposal does not appear to offer any further clues as to how the contour should be chosen[117].

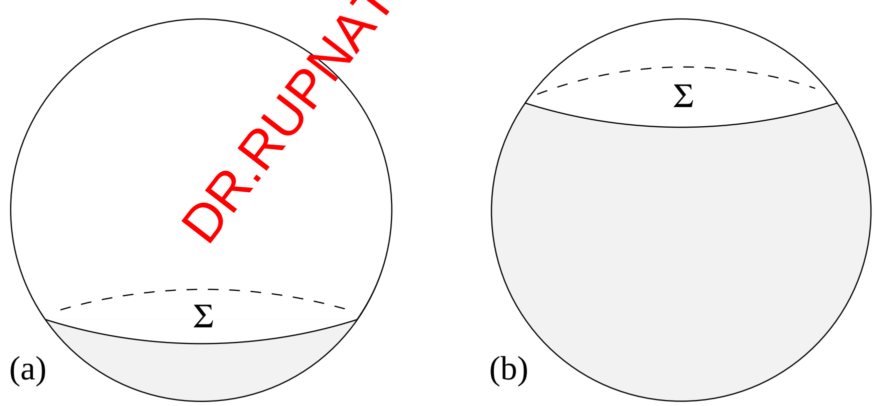

Fig. 8: Euclidean solutions which correspond to matching a given 3-sphere hypersurface, Σ, to a 4-sphere which is: (a) less than half filled; (b) more than half filled.

Non-uniqueness of the contour of integration is already a problem in the simplest conceivable non-trivial minisuperspace model, namely the $k = +1$ FRW universe with a cosmological term and no other matter – the "de Sitter minisuperspace" model. At the semiclassical level one can calculate $e^{-I_{cl}}$ by the steepest descents method. Hartle and Hawking discussed this in their original "no boundary" paper[38], and argued heuristically that one particular saddle point would yield the dominant contribution to the path integral, namely the saddle point corresponding to the classical Euclidean solution which matches the S^3 hypersurface Σ to a less than half filled 4-sphere, rather than the solution which matches S^3 to a more than half filled 4-sphere (see Fig. 8). However, their argument did not stand up to a more rigorous analysis. Halliwell and Louko[66] found a means of evaluating the minisuperspace path integral exactly, and thereby explicitly determined convergent contours. They showed that this simple model possesses inequivalent contours for which the path integral converges. These pass through different saddle points and lead to different semiclassical wavefunctions, Ψ. There are thus many different no-boundary wavefunctions, each corresponding to a different choice of contour. The problem persists in more complicated models[67,68]. Since different no-boundary wavefunctions could lead to different physical predictions, the ambiguity associated with the choice of contour would appear to be the most significant problem with the no-boundary proposal which still needs to be resolved.

Let us return to the minisuperspace example of §3.7 and examine the implications of imposing the Hartle-Hawking boundary condition. For simplicity we will specialise to $k = +1$ models. The other values of k have also been discussed in the literature.[2,105,118]

The minisuperspace no-boundary wavefunction for the $k = +1$ models is given by

$$\Psi_{\rm HH}[a,\phi] = \int^{(a,\phi)} \mathfrak{D}\mathfrak{a}\, \mathfrak{D}\upphi\, \mathfrak{D}\mathcal{N}\, \mathrm{e}^{-I[\mathfrak{a},\upphi,\mathcal{N}]}, \tag{4.1}$$

where▲

$$I = \tfrac{1}{2} \int_0^{\tau_f} \mathrm{d}\tau \mathcal{N} \left[\frac{-\mathfrak{a}}{\mathcal{N}^2} \left(\frac{\mathrm{d}\mathfrak{a}}{\mathrm{d}\tau} \right)^2 + \frac{\mathfrak{a}^3}{\mathcal{N}^2} \left(\frac{\mathrm{d}\upphi}{\mathrm{d}\tau} \right)^2 - \mathfrak{a} + \mathfrak{a}^3 V \right]. \tag{4.2}$$

The integral is taken over a class of paths which match the values

$$\mathfrak{a}(\tau_f) = a, \qquad \upphi(\tau_f) = \phi, \tag{4.3}$$

on the final surface, and the origin of the Euclidean time coordinate, τ, has been chosen to be zero.

There are two approaches we can take to solving the Wheeler-DeWitt equation (3.59): either (i) attempt to interpret the Hartle-Hawking boundary condition directly in terms of boundary conditions of Ψ on minisuperspace; or (ii) take a saddle-point approximation to the path integral (4.1). For consistency these two approaches should agree.

Let us first consider (3.59) directly. Hawking and Page[52,119] have argued that one can approximate the Hartle-Hawking boundary condition in minisuperspace by saying that in an appropriate measure one should have $\Psi = 1$ when $a \to 0$ with ϕ

▲ I shall use a different font to distinguish the minisuperspace coordinates in the functional integral, $(\mathfrak{a}, \upphi)$, from their boundary values, (a, ϕ), in the 3-geometry on the hypersurface Σ.

regular, and $\Psi = 1$ also along the past null boundaries, in order to provide sufficient Cauchy data to solve (3.59) everywhere in the (α, ϕ) plane. It is possible to solve (3.59) exactly for a massless scalar field[86,120], (i.e., $V = 0$), but exact solutions are not known for the massive scalar field. Approximate solutions can be found in various regimes[52,119,121], however.

Firstly, since Ψ must be regular as $a \to 0$, we see that Ψ must be independent of ϕ in this limit, $\frac{\partial \Psi}{\partial \phi} \simeq 0$, in order to overcome the divergence of the a^{-3} factor (using coordinates (a, ϕ)) in (3.59). Furthermore, for $a^2 V \ll 1$ the approximation $V \simeq 0$ is a good one[⋈], and for $k = \pm 1$ (3.59) becomes a Bessel equation in the variable $\frac{1}{2}a^2$. For $k = +1$, which is the case of interest to us here, we therefore obtain the solution[52,119]

$$\Psi \simeq \mathrm{I}_0(\tfrac{1}{2}a^2), \tag{4.4}$$

where $\mathrm{I}_0(z) \equiv \sum_{n=0}^{\infty} \frac{1}{(n!)^2} \left(\frac{z}{2}\right)^{2n}$ is the zero order modified Bessel function, and the normalisation has been fixed to satisfy the boundary condition $\Psi \to 1$ as $a \to 0$. For large a (with $a^2 V \ll 1$),

$$\Psi \sim \frac{1}{\sqrt{\pi}\, a} \exp\left(\tfrac{1}{2}a^2\right) \left[1 + \mathrm{O}(a^{-2})\right], \tag{4.5}$$

i.e., the wavefunction is of exponential type. It therefore agrees with the WKB approximation (3.61) in the limit $a^2 V \ll 1$ for large a provided we take the $(-)$ solution of (3.61) with normalisation

$$\mathcal{C}(\phi) = \frac{1}{\sqrt{\pi}} \exp\left(\frac{+1}{3V(\phi)}\right). \tag{4.6}$$

Let us now consider the limit $V(\phi) \gg 1$ but avoid regimes in which the ϕ dependence of (3.59) is significant by assuming that V is approximately constant, as in the approximation of Fig. 4, so that the ϕ derivative can still be neglected. The term involving the spatial curvature k in (3.59) is now negligible compared to the last term and can also be neglected, yielding the solution[52,119]

$$\Psi \simeq \mathsf{c}(\phi)\, \mathrm{J}_0(\tfrac{1}{3}a^3\sqrt{V}), \tag{4.7}$$

where $\mathrm{J}_0(z) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} \left(\frac{z}{2}\right)^{2n}$ is the zero order ordinary Bessel function. Since the spatial curvature term in (3.59) is always dominant for $a \to 0$, the approximation which led to (4.7) no longer applies in that limit, so the factor c cannot be normalised by the boundary condition $\Psi(0, \phi) = 1$. For large a, however,

$$\Psi \sim \frac{\mathsf{c}}{\sqrt{2\pi \mathsf{S}}} \cos\left(\mathsf{S} - \frac{\pi}{4}\right), \tag{4.8}$$

[⋈] We assume that $V(\phi)$ grows less strongly than $\mathrm{e}^{6|\phi|}$ as $|\phi| \to \infty$, i.e., $|V'/V| < 6$, so that this approximation remains valid for arbitrarily large $|\phi|$.

where $\mathsf{S} = \frac{1}{3}a^3\sqrt{V}$, which is a superposition of the two oscillatory WKB modes (3.60) for large S. Using the WKB connection formula to match the $(-)$ solution of (3.61) to the oscillatory region, we find agreement with the asymptotic limit (4.8), provided

$$\mathsf{c} = \sqrt{\frac{2\pi}{3}}\,\mathcal{C} = \sqrt{\tfrac{2}{3}}\exp\left(\frac{1}{3V(\phi)}\right). \tag{4.9}$$

The ϕ-dependent corrections to these wavefunctions, which result from perturbations in directions in which $\dot{\phi} \not\simeq 0$ have been discussed by Page[2,119,122].

Let us now consider the alternative method of determining Ψ by making a saddle-point approximation to the path integral. In the semiclassical approximation the wavefunction takes the form

$$\Psi \sim \exp\left(-I_{\rm cl}(a,\phi)\right), \tag{4.10}$$

where $I_{\rm cl}$ denotes the Euclidean action (4.2) evaluated at a classical (possibly complex) solution to the Euclidean field equations

$$\frac{1}{\mathcal{N}^2}\left(\frac{\mathrm{d}\mathsf{a}}{\mathrm{d}\tau}\right)^2 - \frac{\mathsf{a}^2}{\mathcal{N}^2}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2 - 1 + \mathsf{a}^2 V(\phi) = 0, \tag{4.11}$$

$$\frac{1}{\mathcal{N}\mathsf{a}}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{1}{\mathcal{N}}\frac{\mathrm{d}\mathsf{a}}{\mathrm{d}\tau}\right) + \frac{2}{\mathcal{N}^2}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2 + V(\phi) = 0, \tag{4.12}$$

$$\frac{1}{\mathcal{N}}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{1}{\mathcal{N}}\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right) + \frac{3}{\mathcal{N}\mathsf{a}}\frac{\mathrm{d}\mathsf{a}}{\mathrm{d}\tau}\frac{\mathrm{d}\phi}{\mathrm{d}\tau} - \tfrac{1}{2}\frac{\mathrm{d}V}{\mathrm{d}\phi} = 0, \tag{4.13}$$

which follow from (3.54)–(3.56) by replacing $t \to -i\tau$. We will henceforth restrict ourselves to the gauge in which $\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}\tau} = 0$, i.e., $\mathcal{N} = \text{const}$, in which case the functional integral over $\mathcal{N}$ in (4.1) must be replaced by an ordinary integral.

The Hartle-Hawking boundary condition demands that

$$\mathsf{a}(0) = 0, \qquad \left.\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right|_0 = 0. \tag{4.14}$$

To see this consider the Euclidean 4-metric which is given by

$$\mathrm{d}s^2 = \mathcal{N}^2\mathrm{d}\tau^2 + \mathsf{a}^2(\tau)\mathrm{d}{\Omega_3}^2. \tag{4.15}$$

The Hartle-Hawking boundary condition requires that we close this 4-geometry in a regular fashion as $\tau \to 0$. This is achieved if $\mathsf{a} \sim \pm\mathcal{N}\tau$ as $\tau \to 0$, since (4.15) is then the same as the metric of the 4-sphere in spherical polar coordinates in this limit. This suggests that we demand $\mathsf{a}(0) = 0$ and $\frac{1}{\mathcal{N}}\frac{\mathrm{d}\mathsf{a}}{\mathrm{d}\tau}\big|_0 = \pm 1$. However, the second condition is guaranteed by the constraint equation (4.11) if the first condition is imposed. The second condition of (4.14) is obtained by noting that (4.13) will give a regular solution for ϕ in the limit $\tau \to 0$ only if $\left.\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right|_0 = 0$, since the middle term of (4.13) diverges otherwise.

If as before we make the simplifying assumption that V can be approximated by a cosmological constant, and $\frac{\mathrm{d}\phi}{\mathrm{d}\tau} \simeq 0$, then it follows that there exist solutions satisfying the boundary conditions (4.3)and (4.14), which may be written

$$\mathfrak{a}(\tau) \simeq \frac{a \sin(\mathcal{N}\sqrt{V}\,\tau)}{\sin(\mathcal{N}\sqrt{V}\,\tau_f)}, \tag{4.16}$$

where

$$\sin^2(\mathcal{N}\sqrt{V}\,\tau_f) = a^2 V, \tag{4.17}$$

which follows from solving the constraint (4.11). If $a^2V < 1$ then (4.17) will give an infinite number of real solutions for the constant $\mathcal{N}$, which may be conveniently parametrised

$$\mathcal{N} = \mathcal{N}_{n\pm} \equiv \frac{1}{\sqrt{V}\,\tau_f}\left[(n+\tfrac{1}{2})\pi \pm \cos^{-1}(a\sqrt{V})\right], \quad n \in \mathbb{Z}, \tag{4.18}$$

with $\cos^{-1}(a\sqrt{V})$ taken in the principal range $(0, \frac{\pi}{2})$. Substitution of (4.18) in (4.16) then gives

$$\mathfrak{a}(\tau) \simeq (-1)^n V^{-1/2} \sin(\mathcal{N}\sqrt{V}\,\tau). \tag{4.19}$$

In addition to the real solutions (4.18) of (4.17), there will also be complex solutions if $a^2V > 1$.

Using the solution (4.19) it is straightforward to evaluate the classical action (4.2). For $n = 0$, for example, we find

$$I_\pm = \frac{-1}{3V(\phi)}\left[1 \pm \left(1 - a^2 V(\phi)\right)^{3/2}\right]. \tag{4.20}$$

If we substitute (4.19) into (4.15) we obtain the metric of the 4-sphere, and thus the classical solutions correspond to matching a given 3-sphere to 4-sphere(s). Furthermore, for $n = 0$ we find $\frac{\mathrm{d}\mathfrak{a}}{\mathrm{d}\tau}\big|_{\tau_f} = \mp\mathcal{N}\sqrt{1 - a^2V}$, so that the $(-)$ solution of (4.18) and (4.20) has $\frac{\mathrm{d}\mathfrak{a}}{\mathrm{d}\tau}\big|_{\tau_f} > 0$, which corresponds to matching the 3-geometry to a less than half filled 4-sphere (Fig. 8(a)), while the $(+)$ solution similarly corresponds to the more than half filled 4-sphere case (Fig. 8(b)). Values of $n > 0$ would appear to give cases in which the 4-geometry pinches off to zero a number of times and then "bounces back" resulting in linear chains of contiguous 4-spheres[123,124]. However, this interpretation of the saddle points as corresponding to universes with bounce solutions does not appear to remain valid[116] once one considers complex solutions to the field equations (4.11)–(4.13).

As was discussed earlier the no-boundary condition does not prescribe a unique contour of integration, and thus it is not completely clear which of the above saddle points should be included in the semiclassical wavefunction (3.30). Halliwell and Hartle have shown[117] that the points with $n < 0$, which have a negative lapse function, lead to difficulties with the recovery of quantum field theory in curved spacetime from quantum cosmology, and thus one might hope that these points should be avoided in the contour of integration. However, no clear grounds present themselves for omitting the other saddle points. For the purpose of making predictions from the no-boundary proposal we shall therefore make a choice by assuming that the contour is such that the solution corresponding to the less than half filled 4-sphere, with action I_-, provides the dominant contribution. (This is the choice that Hartle and Hawking originally made[38].) Neglecting the prefactor, we therefore obtain a no-boundary wavefunction

$$\Psi_{\rm HH}(a,\phi) \propto \exp\left(\frac{1}{3V(\phi)}\right) \exp\left[\frac{-1}{3V(\phi)}\left(1-a^2V(\phi)\right)^{3/2}\right], \tag{4.21}$$

in the region $a^2V < 1$. Using the WKB matching procedure one can show that the corresponding solution in the region $a^2V > 1$ is

$$\Psi_{\rm HH}(a,\phi) \propto \exp\left(\frac{1}{3V(\phi)}\right) \cos\left[\frac{1}{3V(\phi)}\left(a^2V(\phi)-1\right)^{3/2}-\frac{\pi}{4}\right], \tag{4.22}$$

which is the superposition of the two WKB components of (3.60):

$$\begin{aligned} &\Psi_{\rm HH} = \Psi_- + \Psi_+, \\ &\Psi_\pm \propto \exp\left(\frac{1}{3V(\phi)}\right) \exp\left\{\pm i\left[\frac{1}{3V(\phi)}\left(a^2V(\phi)-1\right)^{3/2}-\frac{\pi}{4}\right]\right\}. \end{aligned} \tag{4.23}$$

One may observe that (4.21) and (4.22) agree with the solutions (4.4) and (4.7) found earlier by direct examination of the Wheeler-DeWitt equation in the appropriate limits (4.5) and (4.8). Thus although the saddle point corresponding to the less than half filled 4-sphere does not appear to be picked out in any special way by the path integral, it is favoured by the "approximate" boundary condition[52,119] that $\Psi \to 1$ as $a \to 0$.

One must add the caveat that the approximate boundary condition should be amended if one is to consider genuinely complex solutions of the Euclidean field equations (4.11)–(4.13). This issue has been considered by Lyons[116]. In general one must analytically continue the boundary condition (4.14) demanded by the no-boundary proposal. Although the simple picture of matching a real Euclidean solution to a real Lorentzian solution at the junction is no longer maintained, one nonetheless finds solutions which are initially approximately Euclidean and at late times are approximately Lorentzian, with classical inflationary behaviour[116].

4.2. The tunneling proposal

An alternative approach advocated by Vilenkin is that the boundary condition for the wavefunction, Ψ, should be such as to embody the notion that the universe "tunnels into existence from nothing" without making such specific restrictions on the "initial" geometry as the Hartle-Hawking proposal does. Conceivably there are many possible ways in which such a notion could be translated mathematically into a definition of the wavefunction, Ψ, and indeed many alternative formulations of the tunneling proposal have been put forward[◊]. Some early versions were phrased in a similar fashion to the no-boundary proposal: in particular, Vilenkin[70] proposed defining the wavefunction by a functional integral over Lorentzian metrics which interpolate between a given matter configuration, Φ, and 3-geometry, h_{ij}, and a vanishing 3-geometry, $\varnothing$, lying to its past

$$\Psi\left[h_{ij}, \Phi, \Sigma\right] = \sum_{\mathcal{M}} \int_{\varnothing}^{(h,\Phi)} \mathfrak{D}\boldsymbol{g}\, \mathfrak{D}\phi\, \mathrm{e}^{iS[g_{\mu\nu},\phi]}. \tag{4.24}$$

Vilenkin has also given an alternative formulation of the tunneling proposal in terms of a boundary condition on superspace[40,41] rather than a restriction on manifolds included in the path integral. In order to formulate boundary conditions on superspace it is necessary to consider its boundary, which can be thought of as consisting of 3-metric and matter configurations for which the 3-curvature is infinite, or $|\Phi| \to \infty$ etc. As we have already seen from the example of S^4 (Fig. 7), not all singular 3-geometries will correspond to singular 4-geometries, as it is possible to obtain a singular 3-geometry by a degenerate slicing of the 4-geometry. Therefore we should distinguish points on the boundary of superspace which correspond to genuine singularities of the 4-geometry from those that correspond to degenerate slicings[★]. We call the former the *singular boundary* of superspace, and the latter the *non-singular boundary*.

The tunneling proposal of Vilenkin[41] is that *the wavefunction, Ψ, should be everywhere bounded, and at singular boundaries of superspace Ψ includes only outgoing modes, i.e., those that carry a flux out of superspace.* Thus ingoing modes can only enter at the nonsingular boundary. This definition is somewhat vague as there is no obvious rigorous definition of positive and negative frequency modes in superspace due to the fact that it possesses no Killing vectors[96], and thus there is no clear notion which modes are "ingoing" and which are "outgoing". Furthermore, the structure of superspace is not completely understood, and it has not been rigorously shown that its boundary can be split into singular and non-singular parts.

The tunneling proposal has in fact been formulated with the minisuperspace WKB approximation in mind, in which case the notion of the boundary of minisuperspace and the notions of ingoing and outgoing modes are more clearly defined. Since each oscillatory WKB mode $\Psi \sim \mathrm{e}^{iS_n}$ has a current (3.44), we can classify the modes as ingoing or outgoing according to the direction of $\boldsymbol{\nabla} S_n$ on the surface in question. Heuristically, the idea underlying the Vilenkin boundary condition is that the ensemble of universes described by Ψ should not include any universes

◊ Linde's proposal [69] embodies a similar philosophy. However, it gives a different wavefunction to Vilenkin's "tunneling" wavefunction in simple minisuperspace models [125].

★ This distinction can be made more precise using Morse functions [42].

contracting down from infinite size[40], but only those that correspond to "tunneling from nothing". As we shall see, the "outgoing flux" condition[41] accords with this notion at least in the case of simple minisuperspace models.

The outgoing-flux version of the tunneling proposal agrees with the path integral formulation (4.24) in the case of the simplest de Sitter minisuperspace[42,66], but the two versions would not appear to be equivalent in general[68]. Whereas the no-boundary wavefunction fixes the initial data but leaves the contour of path integration ambiguous, the tunneling proposal fixes the contour of integration but leaves some ambiguity in the specification of the initial data, even when the outgoing-flux condition is imposed[68,126].

Consider the minisuperspace model of §3.7. In the diagrams of Fig. 5 all surfaces $\mathcal{J}^{\pm}_{L,R}$ and the points $\mathtt{i}^0_{L,R}$ and $\mathtt{i}^+$ will be part of the singular boundary, whereas the point $\mathtt{i}^-$, which corresponds to $a \to 0$ ($\alpha \to -\infty$) with ϕ finite, is the only point of the nonsingular boundary. Of course, the condition that $a \to 0$ is not enough to guarantee regularity of the 4-geometry (4.15), and in general one might not expect boundary points to cleanly fall into the category of the "singular boundary" or the "nonsingular boundary". However, as we have already observed in the last section, in the present minisuperspace model the additional requirement for regularity that $\frac{1}{\mathcal{N}}\frac{\mathrm{d}a}{\mathrm{d}\tau}\big|_0 = \pm 1$ is guaranteed by the constraint equation (4.11) if $a(0) = 0$ is imposed.

The oscillatory WKB region, as shaded in the conformal diagrams of Fig. 5 and Fig. 9, is always bounded by $\mathtt{i}^0_R$, $\mathcal{J}^+_R$ and $\mathtt{i}^+$, and in all cases except that of Fig. 5(f)[▽] the oscillatory region is bounded by $\mathtt{i}^0_L$ and $\mathcal{J}^-_L$ also. The wavefunction in this region is given by a superposition of terms $\mathrm{e}^{i\mathsf{S}_n}$, where S_n is a solution to the Hamilton-Jacobi equation (3.34), which in terms of the variables (α, ϕ) may be written

$$-\left(\frac{\partial \mathsf{S}}{\partial \alpha}\right)^2 + \left(\frac{\partial \mathsf{S}}{\partial \phi}\right)^2 + \mathsf{U}(\alpha, \phi) = 0, \tag{4.25}$$

where $\mathsf{U}(\alpha, \phi) \equiv 2\mathrm{e}^{3\alpha}\mathcal{U}(\alpha, \phi) = \mathrm{e}^{4\alpha}\left(\mathrm{e}^{2\alpha}V(\phi) - 1\right)$, and for convenience in what follows we have suppressed the index n. The characteristics of (4.25) satisfy

$$\frac{\mathrm{d}\alpha}{2\mathsf{S}_{,\alpha}} = \frac{\mathrm{d}\phi}{-2\mathsf{S}_{,\phi}} = \frac{\mathrm{d}\mathsf{S}}{2\mathsf{U}} = \frac{\mathrm{d}(\mathsf{S}_{,\alpha})}{\mathsf{U}_{,\alpha}} = \frac{\mathrm{d}(\mathsf{S}_{,\phi})}{\mathsf{U}_{,\phi}}\,. \tag{4.26}$$

Thus each $\mathsf{S}(\alpha, \phi)$ describes a congruence of classical paths with

$$\frac{\mathrm{d}\phi}{\mathrm{d}\alpha} = -\frac{\mathsf{S}_{,\phi}}{\mathsf{S}_{,\alpha}} = \frac{\mp \mathsf{S}_{,\phi}}{\sqrt{\mathsf{S}_{,\phi}{}^2 + \mathsf{U}}}. \tag{4.27}$$

[▽] Unlike the other cases the potential of Fig. 5(f), which corresponds to a potential in which supersymmetry is broken through gaugino condensation in string theory, vanishes as $\phi \to -\infty$. This limit is the "weak coupling limit" of string theory.

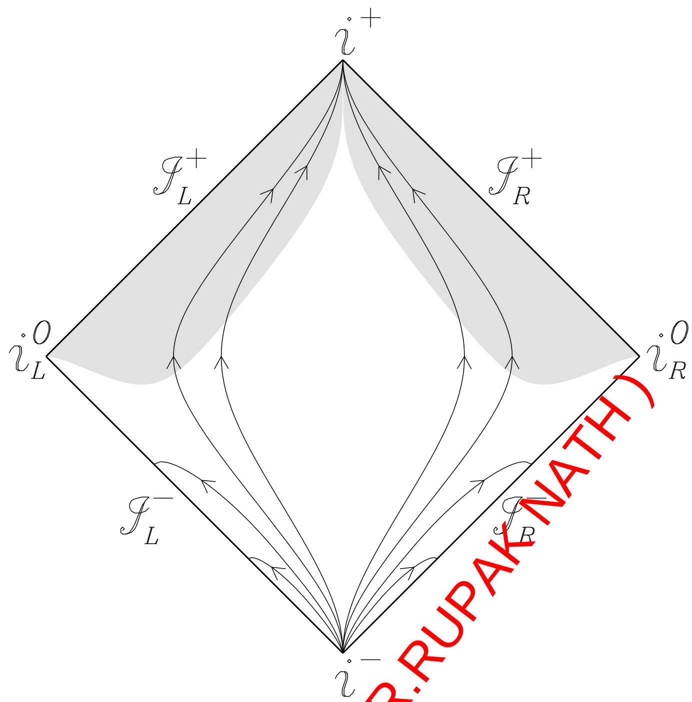

Fig. 9: Some schematic probability flows consistent with the tunneling proposal. An indicative oscillatory WKB region is shown (shaded area) for the potential $V = \phi^2$.

Since $\mathsf{U} > 0$ in the oscillatory region (assuming $V(\phi) > 0$), it follows that these integral curves satisfy $\left|\frac{\mathrm{d}\phi}{\mathrm{d}\alpha}\right| < 1$, i.e., they are "timelike" in the minisuperspace coordinates and have an endpoint at $\mathfrak{i}^+$. Since $\pi_\alpha \equiv -\frac{\mathrm{e}^{3\alpha}\dot{\alpha}}{\mathcal{N}} = \mathsf{S}_{,\alpha}$ it follows that the WKB modes correspond to expanding universes ($\dot{\alpha} > 0$) with $\pi_\alpha = \mathsf{S}_{,\alpha} < 0$, or contracting universes with $\pi_\alpha = \mathsf{S}_{,\alpha} > 0$, if we assume[⊘] $\mathcal{N} > 0$. Since all paths originate at finite values of α the latter solutions will extend from $\mathfrak{i}^+$ into the interior of the minisuperspace, i.e., they are "ingoing modes" and are to be excluded on account of the tunneling boundary condition. This boundary condition only allows the expanding solutions which are "outgoing" at $\mathfrak{i}^+$. In the approximation in which (4.5) applies it follows that the tunneling proposal demands that only the modes with $\Psi \sim \exp\left[\frac{-i}{3V(\phi)}\left(e^{2\alpha}V(\phi) - 1\right)^{3/2}\right]$ are admitted in the oscillatory region. The tunneling wavefunction is thus complex in the oscillatory region, in contrast to the

⊘ This choice is a matter of convention and the opposite choice would reverse the roles of the "ingoing" and "outgoing" modes. Equivalently, the coordinate t is an arbitrary label from the point of view of general relativity, and it is a matter of convention whether we choose t to increase or decrease towards the "future"; the "future" being defined by the expansion of the universe (the cosmological arrow) or the increase of entropy (the thermodynamic arrow). In terms of Fig. 4 it amounts to an arbitrary choice between A_+ and A_- as representing the "late-time behaviour" of the inflationary solutions.

no-boundary wavefunction (4.22) which is real.

To extend the tunneling wavefunction into the "tunneling region" (the unshaded regions of Fig. 5 and Fig. 9), we can use the WKB matching procedure. Of course, there will also be additional solutions that remain entirely in the tunneling region and never cross into the oscillatory region. In the case of the null boundaries, $\mathcal{I}^-_{L,R}$, which border this region we observe that as $\alpha \to -\infty$ (3.59) becomes the wave equation in the (α, ϕ) coordinates, and thus the solutions are asymptotically null which leads to a notion of "ingoing" and "outgoing" modes, so that the tunneling condition can be imposed. Some possible probability flows consistent with the tunneling proposal are shown in the conformal diagram of Fig. 9.

5. The Predictions of Quantum Cosmology

Given the many problems and uncertainties in the quantum cosmology programme that have been discussed above, one could easily form the opinion that it is premature to talk about the predictions that quantum cosmology makes. Nevertheless, it is important to investigate the types of predictions we might expect quantum cosmology to make about the universe, as well as realistically evaluating the limitations of the these predictions. I will concentrate on three key areas in this section. Other topics, most notably the variability of the constants of nature, have been discussed recently by Vilenkin[125,127].

5.1. The period of inflation

The construction of a suitable probability measure should allow us to answer questions such as whether inflation is a feature of a typical universe. As we saw in §3.6, the construction of a general probability measure is problematic. Thus we shall limit our discussion to the oscillatory WKB limit in which case the Klein-Gordon-type current (3.42) is adequate.

The issue of the duration of the period of inflation has been a point of some debate between proponents of the no-boundary wavefunction and proponents of the tunneling wavefunction. Consider the canonical minisuperspace model of §3.7. We have seen that in the WKB limit the Hartle-Hawking boundary condition gives rise to a wavefunction (4.22) in the oscillatory region, which is strongly peaked about the set of classical solutions (3.62) that correspond to an inflating universe: $a(t) \propto \mathrm{e}^{\sqrt{V}\,t}$, $\phi(t) \simeq \phi_0 = \text{const}$.

Vilenkin[40–42,70] has also studied the minisuperspace model of §3.7, but with a different choice of factor ordering in the Wheeler-DeWitt equation (3.59). He concludes that similarly to (4.22) the tunneling boundary condition leads to a WKB wavefunction

$$\Psi_{\mathrm{V}}(a,\phi) \propto \exp\left(\frac{-1}{3V(\phi)}\right) \exp\left[\frac{-i}{3V(\phi)}\left(a^2V(\phi)-1\right)^{3/2} + \frac{i\pi}{4}\right] \tag{5.1}$$

in the oscillatory region, which is also peaked about the classical trajectories of the inflationary universes (3.62).

Both the no-boundary and tunneling wavefunctions thus predict inflation, at least in the context of this simple minisuperspace model. The important question is *how much* inflation do the models predict? This will be largely decided by the value, ϕ_0, of the scalar field with which the universe "nucleates" in the semiclassical regime.

In order to study the probability flux arising from (3.42) in the WKB limit it is necessary to focus on one particular WKB component. The no-boundary wavefunction (4.22) is of course real, and the resulting current (3.42) is identically zero. However, $\Psi_{\rm HH}$ is the superposition (4.23) of two WKB components, which will correspond to contracting and expanding universes. As discussed in §3.5 it is assumed that a decoherence mechanism exists so that the interference between the two components is negligible[89], and we can therefore assume that the universe is peaked about one or other WKB component for the purposes of determining the probability measure. The difference between the no-boundary and tunneling wavefunctions may therefore not seem great once decoherence to a classical universe is assumed. In particular, if we take the "outgoing" WKB component in the no-boundary case then the only significant difference between $\Psi_{\rm HH}$ and $\Psi_{\rm T}$ is the ϕ-dependent part of the prefactor, $\exp\left(\frac{\pm 1}{3V(\phi)}\right)$, which is obtained from boundary conditions set in the tunneling region. This difference will have ramifications for the probability flux, however.

One question which we might hope to answer in our minisuperspace model would be: given that a Lorentzian universe nucleates, what is the probability that it inflates by a sufficient amount (~ 65 e-folds) to solve the problems of the standard cosmology mentioned in §1? The answer to this would involve integrating the probability flux (3.45) on the surface separating the tunneling and oscillatory regions, which is roughly given by $a^2 V(\phi) = 1$. However, our discussion here is limited by the fact that our WKB approximation applies to trajectories with $\dot{\phi} \simeq 0$. For such trajectories the probability current, $\mathcal{J}$, points chiefly in the direction of the a-coordinate in minisuperspace. We can therefore attempt to answer the question approximately by evaluating the probability current (3.44) on a surface Σ in minisuperspace with $a = \text{const}$. For sufficiently large values of a such surfaces will lie almost entirely within the oscillatory regime (see Fig. 10). On these surfaces we obtain a probability flux

$$\mathrm{d}\mathcal{P} = \mathcal{J} \cdot \mathrm{d}\Sigma \propto \begin{cases} \exp\left(\frac{+2}{3V(\phi)}\right) & \text{no-boundary wavefunction, } \Psi_{\rm HH}, \\ \exp\left(\frac{-2}{3V(\phi)}\right) & \text{tunneling wavefunction, } \Psi_{\rm V}. \end{cases} \tag{5.2}$$

Although the integral of (5.2) may diverge for particular potentials $V(\phi)$ of interest, this should not be viewed as a problem since, as was emphasised in §3.6, questions in quantum cosmology can only refer to conditional rather than absolute probabilities.

In the present case let us assume that inflation occurs for large values of the scalar field, as is the case for potentials of the "chaotic" type, $V = \lambda\phi^{2p}$. There will then be a minimum value of the scalar field, $\phi_{\rm suff}$, for which sufficient inflation is obtained. For $V = \lambda\phi^{2p}$ a universe with $\phi = \phi_0$ initially will undergo

$$N_e \simeq \frac{3}{2p}\left[{\phi_0}^2 - \tfrac{2}{9}p(2p-1)\right] \tag{5.3}$$

e-folds of inflation⊡, so that in the case of the quadratic potential ($p = 1$) we find $\phi_{\rm suff} \gtrsim 6.6$, for example. The relevant conditional probability for sufficient inflation on an $a =$ const surface is then given by

$$\mathcal{P}\left(\phi_0 > \phi_{\rm suff} \mid \phi_1 < \phi_0 < \phi_2\right) = \frac{\displaystyle\int_{\phi_{\rm suff}}^{\phi_2} {\rm d}\phi_0 \exp\left(\frac{\pm 2}{3V(\phi_0)}\right)}{\displaystyle\int_{\phi_1}^{\phi_2} {\rm d}\phi_0 \exp\left(\frac{\pm 2}{3V(\phi_0)}\right)}, \tag{5.4}$$

where the (+) case refers to the no-boundary wavefunction and the (−) case to the tunneling wavefunction, and the values ϕ_1 and ϕ_2 are respectively lower and upper cutoffs on the allowed values of ϕ, which must be determined by physical criteria. A minimum cut-off might be expected for a variety of physical reasons, such as avoiding classical universes which rapidly recollapse.

In their original investigations of the question of the duration of the period of inflation for the quadratic scalar potential, Hawking and Page[52] took $\phi_2 = \infty$, in which case both integrals in (5.4) are dominated by large values of ϕ, leading to a probability $\mathcal{P} \simeq 1$, and thus a "prediction" of inflation. However, this result has been criticised by Vilenkin[41] since for $k = +1$ we have $\mathcal{V} = \frac{9}{16} m_{\rm Planck}^4 V$ so that values of $m\phi \gtrsim 4/3$ are in excess of the Planck scale, and the semiclassical approximation will no longer apply. Vilenkin suggested that an upper cutoff, ϕ_2, should be introduced at the Planck scale. If this is the case then provided the lower cutoff ϕ_1 is sufficiently close to zero, we would find that the integral in the denominator of (5.4) becomes very large in the case of the no-boundary wavefunction (+ sign), leading to $\mathcal{P} \ll 1$, whereas this would not be the case for the tunneling wavefunction (− sign), and thus the latter would predict more inflation.

Introducing a cutoff at the Planck scale might be deemed a rather arbitrary procedure, since without any knowledge of Planck scale physics it is impossible to be sure whether the "real" answer is better approximated by the introduction of a cutoff or not. However, it has been argued on the basis of investigations at the 1-loop level that the wavefunction is damped at large values of ϕ by quantum corrections[128,129]. This renders the wavefunction normalisable and would justify a cutoff, ϕ_2, near the Planck scale.

⊡ In the slow-rolling approximation [109] it follows from (3.54) and (3.56) (with $\mathcal{N} = 1$) that the number of e-folds is $N_e = 6\int_{\phi_e}^{\phi_0} {\rm d}\phi\, V/V'$, where ϕ_0 is the initial value and ϕ_e the final value of the scalar field at the end of the inflationary epoch. For $V = \lambda\phi^{2p}$ we have $N_e = \frac{3}{2p}\left({\phi_0}^2 - {\phi_e}^2\right)$. The value of ϕ_e can be estimated from the limit set by $|V'/V| \ll 6$ and $|V''/V| \ll 9$.

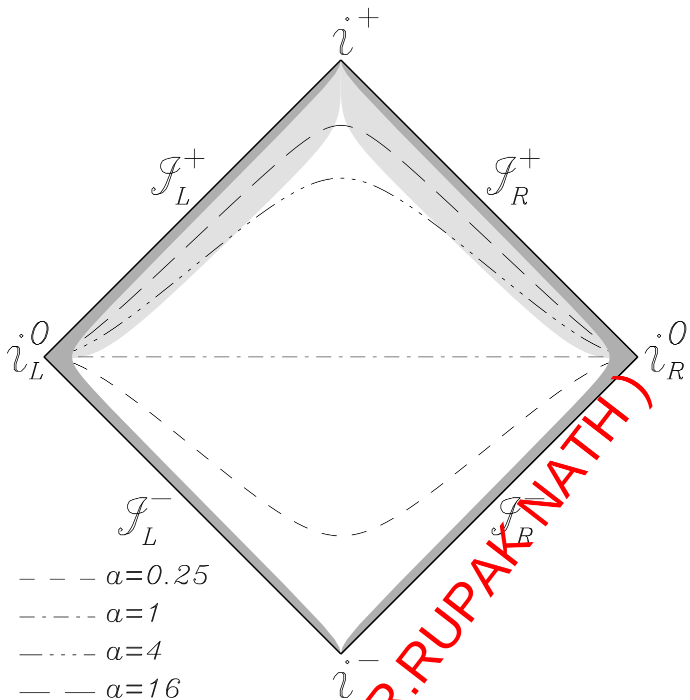

Fig. 10: Conformal diagram for $V = 0.04\phi^2$. The oscillatory region, given roughly by $a^2V > 1$, is lightly shaded. Lines a = const are superimposed. For very large values of ϕ these lie almost entirely in the oscillatory region. The region of ϕ-values excluded by a Planck scale cutoff is darkly shaded.

Given that the no-boundary wavefunction apparently yields a wavefunction peaked around the lower cutoff ϕ_1, it is important to determine what a reasonable value of this cutoff should be. This issue has been considered by Grishchuk and Rozhansky[130], and also more recently by Lukas[131], who have conducted numerical investigations to analyse the behaviour of the caustic[132] in the (a, ϕ) plane which separates the Euclidean and Lorentzian solutions. This is illustrated in Fig. 11, where classical solutions to the Euclidean field equations (3.54)–(3.56) (with $\mathcal{N} = 1$) are shown. The solutions beginning at $a = 0$ initially follow lines with $\dot{\phi} \simeq 0$, but the approximation eventually breaks down when the trajectories curl back and recollapse with $a \to 0$ and $|\phi| \propto -\ln a \to \infty$. They thus represent a flow from $\imath^-$ to $\mathscr{I}^-$ in the conformal diagrams. The solutions cross each other on the caustic, which for large values of ϕ corresponds to $a^2V(\phi) = 1$, in accordance with our expectation from the WKB approximation.

To the right of the Euclidean solutions in Fig. 11 we would find Lorentzian solutions with $\dot{\phi} \simeq 0$ sufficiently far away from the caustic[132]. These solutions are not depicted here. The "nucleation of a universe" would thus correspond to a solution of the Wheeler-DeWitt equation which was initially peaked about a classical Euclidean trajectory with $\dot{\phi} \simeq 0$, and which then crossed over the caustic to be peaked about a corresponding trajectory with $\dot{\phi} \simeq 0$ in the Lorentzian region.

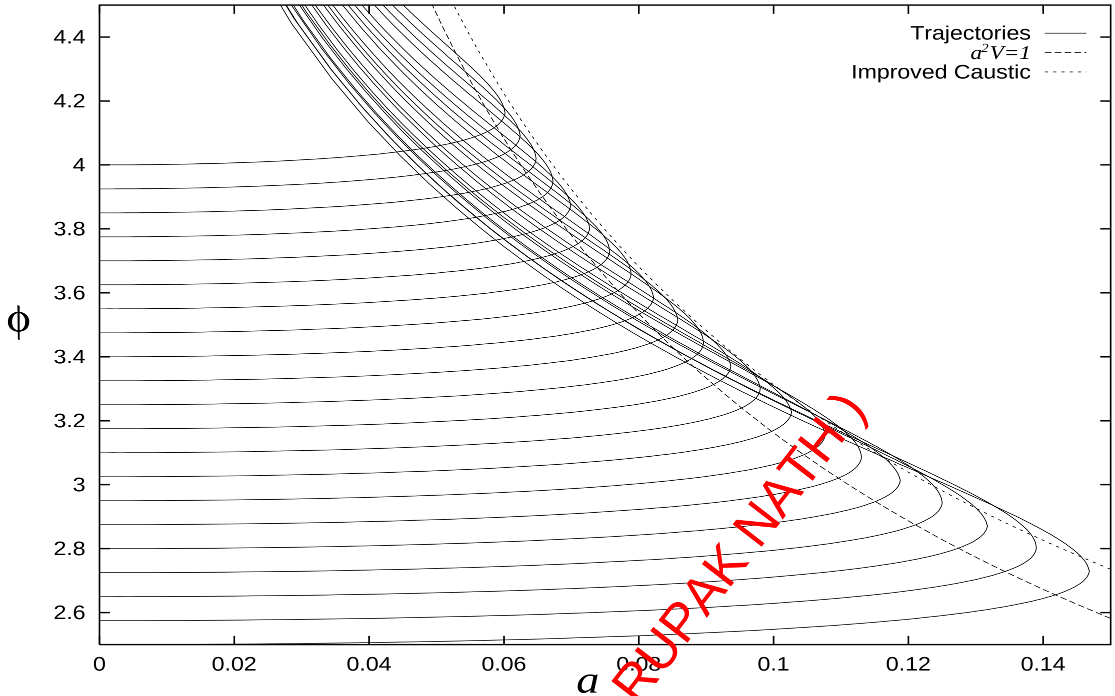

Fig. 11: Classical Euclidean trajectories for the potential $V = \phi^4$. The approximate caustic $a^2V = 1$ is indicated by a dashed line, and the improved caustic by a dotted line. (From [131].)

While crossing the caustic, the wavefunction would be peaked about a complex solution which was neither truly Euclidean nor truly Lorentzian.

It is evident from Fig. 11 that for small values of ϕ the caustic begins to deviate from the curve $a^2V = 1$. In fact, one finds[130,131] that for sufficiently small values of ϕ Lorentzian universes never nucleate at all. Accordingly, a portion of the shaded region in the conformal diagrams should be excised about the $\phi = 0$ axis in the vicinity of i^+. The cutoff, ϕ_*, at which this occurs has been estimated by Lukas[131] as being

$$\phi_* \simeq p + \left\{2p\left[1 - (1 + \mathrm{e}^{-1})^{-1/2}\right]\right\}^{1/2}, \tag{5.5}$$

in the case of the chaotic potentials, $V = \lambda\phi^{2p}$. Provided that the interpretation of (5.4) remains valid♠, we should therefore set $\phi_1 = \phi_*$ as the lower bound in the integrals. In the case of the quadratic potential, $V = m^2\phi^2$, we then have $\phi_1 \simeq 1.5$, and using the values of $\phi_{\rm suff}$ and ϕ_2 found earlier it is straightforward to

♠ As a cautionary note, one should observe that since the approximation of the caustic by the curve $a^2V = 1$ breaks down for small ϕ, the use of the surfaces $a =$ const which was assumed in (5.4) may not be appropriate for small ϕ, and ideally one should determine the probability flux across the caustic itself. Of course, in a more careful analysis one would consider complex solutions to the field equations in line with [116], rather than real Euclidean and real Lorentzian solutions with a junction condition. One would hope that this would not alter the conclusions of the analysis much.

check that for typical values of $m \ll 1$, (5.4) gives $\mathcal{P} \ll 1$ for $\Psi_{\rm HH}$ and $\mathcal{P} \sim 1$ for $\Psi_{\rm V}$ in accordance with the earlier discussion.

The above analysis would appear to indicate that the no-boundary wavefunction effectively "predicts" a value of $\phi \simeq \phi_*$, which in the case of the quadratic potential is unfortunately less than $\phi_{\rm suff}$. However, the calculation is model-dependent, and if one could find a potential for which $\phi_* \gtrsim \phi_{\rm suff}$ then both $\Psi_{\rm HH}$ and $\Psi_{\rm V}$ would yield $\mathcal{P} \sim 1$. If we compare (5.3) and (5.5) (for $N_e \simeq 65$) we see that in the case of the chaotic potentials, $V = \lambda\phi^{2p}$, this requirement is equivalent to $p \gtrsim 62$. However, it requires an enormously small value of λ to keep such a V below the Planck scale and this is not promising. Lukas[131] has also estimated the value of ϕ_* for some other potentials but did not find any candidates with $\phi_* \gtrsim \phi_{\rm suff}$. However, given the model-dependence of the calculations it cannot be ruled out that some other potential, or the coupled effects of two or more scalar fields, might give $\phi_* \gtrsim \phi_{\rm suff}$ and thereby make inflation a "prediction" of the no-boundary wavefunction.

5.2. The origin of density perturbations

It is expected that the anisotropies in the cosmic microwave background radiation, which were first definitively observed in 1992, have their origin in quantum fluctuations in the very early universe. Such fluctuations can be described using the formalism of quantum field theory in curved spacetime. Since the era of quantum cosmology is in a sense prior to that in which quantum field theory in curved spacetime is applicable, one would hope to trace the origin of the primordial perturbations back to quantum cosmology. Indeed, this can be done[133–135] and I will very briefly outline the main results.

With homogeneous minisuperspaces as a starting point, one can add small inhomogeneous perturbations to the metric and matter fields:

$$h_{ij}(x,t) = a^2(t)\left(\Omega_{ij} + \varepsilon_{ij}\right), \tag{5.6}$$
$$\Phi(x,t) = \Phi_0(t) + \delta\Phi(x,t), \tag{5.7}$$
$$\mathcal{N}(x,t) = \mathcal{N}_0(t) + \delta\mathcal{N}(x,t), \tag{5.8}$$
$$\mathcal{N}_i(x,t) = 0 + \delta\mathcal{N}_i(x,t). \tag{5.9}$$

Here we have restricted attention to the $k = +1$ FRW minisuperspace model, the subscript zero denotes the unperturbed quantities, and Ω_{ij} is the unperturbed standard round metric on the 3-sphere. The perturbations can be expanded in terms of spherical harmonics on the 3-sphere[133]. If one substitutes the ansatz (5.6)–(5.9) into the classical action (2.12) and expands to quadratic order one obtains an action which can be split into the original minisuperspace action, S_0, and an additional action, S_2, quadratic in the perturbations,

$$S = S_0[q^A, \mathcal{N}_0] + S_2[q^A, \mathcal{N}_0, \varepsilon_{ij}, \delta\Phi, \delta\mathcal{N}, \mathcal{N}_i], \tag{5.10}$$

and a corresponding Hamiltonian

$$H = \mathcal{N}_0 \left(\mathsf{H}_0 + \int \mathrm{d}^3x\, \mathcal{H}_2 \right) + \int \mathrm{d}^3x\, \delta\mathcal{N}\, \mathcal{H}_1 + \int \mathrm{d}^3x\, \delta\mathcal{N}_i\, \mathcal{H}^i\,, \tag{5.11}$$

where H_0 is the unperturbed Hamiltonian (3.54), and the terms $\mathcal{H}_1$ and $\mathcal{H}_2$ are linear and quadratic in the perturbations respectively. There is now a non-trivial momentum constraint at each point $x \in \Sigma$, $\mathcal{H}^i(x) = 0$, while the Hamiltonian constraint splits into a piece linear in the perturbations, $\mathcal{H}_1(x) = 0$, plus a homogeneous piece

$$\mathsf{H}_0 + \mathsf{H}_2 \equiv \mathsf{H}_0 + \int \mathrm{d}^3x\, \mathcal{H}_2 = 0. \tag{5.12}$$

We may quantise (5.12) in the standard fashion to obtain a modified Wheeler-Dewitt equation

$$\hat{\mathsf{H}}\Psi = \left[-\tfrac{1}{2}\nabla^2 + \mathcal{U}(q) + \hat{\mathsf{H}}_2 \right] \Psi = 0, \tag{5.13}$$

in place of (3.27), where the Laplacian is still defined in terms of the original minisuperspace coordinates according to (3.28), and $\hat{\mathsf{H}}_2$ is a second order differential operator which results from quantisation of the homogeneous part of the perturbations. In the case of pure scalar field modes, for example, the perturbations may be decomposed as

$$\delta\Phi(x,t) = \frac{1}{\sigma} \sum_{nml} \mathsf{f}_{nlm}(t) Q^n{}_{lm}(x), \tag{5.14}$$

where the $Q^n{}_{lm}$ satisfy the 3-dimensional Laplace equation on S^3,

$$^{(3)}\Delta Q^n{}_{lm} = -(n^2 - 1) Q^n{}_{lm}, \tag{5.15}$$

and one finds[1,89,133]

$$\hat{\mathsf{H}}_2 = \tfrac{1}{2} \sum_{nml} \left\{ \frac{-1}{a^3} \frac{\partial^2}{\partial \mathsf{f}^2_{nlm}} + \left[(n^2 - 1)a + m^2 a^3 \right] \mathsf{f}^2_{nlm} \right\} \tag{5.16}$$

in the case of the quadratic scalar potential.

It is possible to find solutions to the Wheeler-DeWitt equation (5.13) in which the minisuperspace coordinates, q^A, are treated semiclassically in the WKB approximation, while the perturbations are treated quantum mechanically. One can show[1,89,133,134] that the solutions take the form

$$\Psi = \mathcal{C}(q, \Phi)\, \mathrm{e}^{i\mathsf{S}_0(q,\Phi)}\, \tilde{\psi}, \tag{5.17}$$

where S_0 satisfies the unperturbed Hamilton-Jacobi equation (3.34), the prefactor $\mathcal{C}$ depends only on the unperturbed minisuperspace coordinates, and the functions $\tilde{\psi}$ satisfy the functional Schrödinger equation

$$i \frac{\partial \tilde{\psi}}{\partial t} = \hat{\mathsf{H}}_2 \tilde{\psi}. \tag{5.18}$$

The different modes of the scalar perturbations (5.14) do not interact, for example, and in this case

$$\tilde{\psi} = \prod_{nml} \psi_{nlm}(t, \mathsf{f}_{nlm}), \tag{5.19}$$

where each mode ψ_{nlm} separately satisfies (5.17) with $\hat{\mathsf{H}}_2$ given by (5.16).

The wavefunction (5.17) is thus peaked about classical trajectories, with corrections, $\tilde{\psi}$, which satisfy the functional Schrödinger equation along these trajectories. This is in fact precisely the starting point for the treatment of matter modes in a curved spacetime background using the formalism of quantum field theory in curved spacetime. Thus the above result is important in that it demonstrates that quantum cosmology is consistent with the standard approach to the quantum treatment of cosmological perturbations. As an added bonus the imposition of a boundary condition on Ψ, such as the no-boundary or tunneling condition, will result in the choice of particular solutions of the functional Schrödinger equation, and consequently a particular vacuum state for the matter modes. Both the no-boundary condition and the tunneling condition pick out[41,136–139] a de-Sitter invariant state known as the "Euclidean" or Bunch-Davies vacuum[140,141], which is the state that is often assumed in cosmological calculations of density perturbations. In fact, this state is picked out by many boundary conditions[137,138], and thus could be regarded as a natural quantum state for matter in quantum cosmology.

One may solve the functional Schrödinger equation and study the growth[133] of the modes. As discussed above, the results obtained are the same as those which are found if one begins with quantum matter fields in de-Sitter invariant vacuum states in background inflationary spacetimes[142,143]. In particular, one obtains a scale-free spectrum of density perturbations which act as seeds for the formation of galaxies and other structures in the universe. Such a spectrum accords well with the spectrum deduced from the COBE measurements of cosmic microwave anisotropies[144].

5.3. The arrow of time

The question of the origin of the arrow of time in the face of the CPT-invariance of the laws of physics is one of the deepest unresolved conceptual issues in theoretical physics, and it has been the subject of more than one major conference[44,145]. The question of the nature of time assumes prime importance in quantum gravity and quantum cosmology[♡]. One might argue that if the basic laws of physics are time symmetric, then the arrow of time could have its origin in a boundary condition for the wavefunction of the universe.

There are many different arrows of time observed in the universe. Some of these such as the psychological arrow of time (we remember only the past), and

♡ The conceptual issues surrounding the nature of time in quantum cosmology would entail a series of lectures in themselves. For further discussion see, e.g., [24], [146]–[148] and references therein.

the electromagnetic arrow of time (the choice of retarded as opposed to advanced solutions of Maxwell's equations), could well be argued to be consequences of the thermodynamic arrow of time, which arises from the second law of thermodynamics. However, the expansion of the universe provides an alternate cosmological arrow of time which is not obviously directly related to the increase of entropy. Whether such a relationship does exist is a question which might potentially be resolved by quantum cosmology.

In 1985 Hawking[149] proposed that the thermodynamic arrow and cosmological arrows of time were correlated: that is to say if the universe were spatially closed then entropy would decrease in the contracting phase[⊚]. In arriving at this proposal he was influenced by the fact that the no-boundary wavefunction is CPT-invariant, and also by some early studies of simple minisuperspace models which possessed quasi-periodic solutions[151] which "bounced" instead of recontracting to a singularity as $a \to 0$. However, Hawking soon changed his mind on the issue, which he terms his "greatest mistake"[152]. This change of mind was brought about by a number of factors. Firstly, Page pointed out that CPT-invariance of the no-boundary wavefunction does not imply CPT-invariance for an individual WKB component of the wavefunction, which would correspond to the history of a classical universe[153]. Furthermore, minisuperspace models which were subsequently studied, such as that of the Kantowski-Sachs universe, were found not to admit bounce solutions but always possessed singularities to the future[154,155]. Finally, as was mentioned in §4.1, the bounce solutions do not feature even in simple minisuperspace models once one considers the contribution of complex metrics[116]. In general, the approximate minisuperspace boundary condition that $\Psi \to 1$ as $a \to 0$ must be altered to allow for approximately Euclidean metrics which also contribute a rapidly oscillating component to the wavefunction[116,156].

Hawking, Laflamme and Lyons[156] have recently argued that a thermodynamic arrow of time results from the imposition of the Hartle-Hawking boundary condition. More precisely, they have considered the evolution of primordial fluctuations as outlined in §5.2, but accounting for the "approximately Euclidean" geometries[116] which appear to be required if one considers complex metrics. They find[156] that the gravitational wave perturbations have an amplitude that remains in the linear regime and is roughly time-symmetric about the time of maximum expansion. Such perturbations cannot be said to give rise to an arrow of time. Density perturbations behave differently, however. They start out small but grow large and become non-linear as the universe expands, and moreover this growth continues during the contracting phase of the spatially closed universe. This growth of inhomogeneity therefore provides an arrow of time which could be considered to be an essentially thermodynamic arrow. Since it does not match the cosmological arrow in the contracting phase, the only reason for the coincidence of the two arrows in the present epoch would appear to be an anthropic one. In particular, the conditions that would

⊚ This idea was first suggested by Gold [150], and has resurfaced a number of times in different contexts.

prevail in a contracting phase would appear to preclude the existence of life[152,156]. Thus the fact that we are around to make observations means we must find ourselves in a cosmological epoch in which the two arrows coincide.

It should be added that the debate about whether the cosmological and thermodynamic arrows of time coincide has not yet been closed, however, and it is still maintained by Kiefer and Zeh[157] that the boundary condition on the Wheeler-DeWitt equation must be such that the thermodynamic arrow would reverse in a recontracting universe.

6. Conclusion

I hope to have shown you that although research in quantum cosmology is still rather speculative and open-ended, its framework nonetheless has the potential not only to provide answers to questions surrounding the origin and early evolution of the universe, but also to help us unravel the mysteries of quantum gravity. Quantum cosmology is a field in which a great deal remains to be done, and the results of §5 should be regarded as a hopeful indication of the types of predictions we might hope to make. It is too early to draw definitive conclusions about the relative merits of the various boundary condition proposals. The results of §5.1 ostensibly favour the "tunneling proposal". However, this conclusion is only based on a few simple models, and therefore some caution must be exercised.

As a final note, it is worth mentioning that recent results show that supersymmetry provides a means of restricting possible boundary conditions for the Wheeler-DeWitt equation, or the corresponding Dirac square-root equation. The supergravity constraint equations for various homogeneous minisuperspace models appear to be so restrictive that they only pick out the most symmetric quantum states[102,103,158]. In particular, simple analytic solutions for Ψ in the supersymmetric Bianchi-IX minisuperspace have been found[102] in the empty and filled fermion sectors, which have a natural interpretation as[158] wormhole states[121,159], or as[160] Hartle-Hawking no-boundary states. Some doubts were initially cast on the relevance of these solutions, as the states appear to have no counterpart in 4-dimensional supergravity[161,162]. However, more recent work[46] on supersymmetric minisuperspaces corresponding to Bianchi "class A"[76] models indicates that infinitely many physical states with finite (even) fermion number can be found, and these are direct analogues of physical states in full supergravity. This result shows that such minisuperspace models are likely to be physically very important for quantum cosmology.

Acknowledgement I would like to thank A. Lukas for giving me permission to use Fig. 11, which originally appeared in [131].

References

1. J. J. Halliwell, in *Quantum Cosmology and Baby Universes*, eds. S. Coleman, J. B. Hartle T. Piran and S. Weinberg, (World Scientific, Singapore, 1991), p. 159.
2. D. N. Page, in *Proceedings of the Banff Summer Institute on Gravitation, August 1990*, eds. R. B. Mann and P. S. Wesson, (World Scientific, Singapore, 1991).
3. P. van Nieuwenhuizen, Phys. Rep. **68** (1981) 189.
4. P. West, *Introduction to Supersymmetry and Supergravity*, (World Scientific, Singapore, 1986).
5. A. Salam and E. Sezgin (eds.), *Supergravities in Diverse Dimensions*, (World Scientific, Singapore, 1989).
6. G. Esposito, *Quantum Gravity, Quantum Cosmology and Lorentzian Geometries*, (Springer *Lecture Notes in Physics* **m12**, Berlin, 1992).
7. M. H. Goroff and A. Sagnotti, Phys. Lett. **160B** (1985) 81; Nucl. Phys. **B266** (1986) 709.
8. G. 't Hooft and M. Veltman, Ann. Inst. H. Poincaré **20** (1974) 69.
9. S. Deser and P. van Nieuwenhuizen, Phys. Rev. **D10** (1974) 411.
10. S. Deser, P. van Nieuwenhuizen and H. S. Tsao, Phys. Rev. **D10** (1974) 3337.
11. S. Deser, J. H. Kay and K. S. Stelle, Phys. Rev. Lett. **38** (1977) 527.
12. P. S. Howe and K. S. Stelle, Int. J. Mod. Phys. **A4** (1989) 1871.
13. M. B. Green, J. H. Schwarz and E. Witten, *Superstring theory*, (Cambridge University Press, 1987).
14. D. J. Gross and V. Periwal, Phys. Rev. Lett. **60** (1988) 2105.
15. S. Davis, Preprint DAMTP-R/94/27-rev (1994), hep-th/9503231.
16. A. Ashtekar, *Lectures on Non-perturbative Canonical Gravity*, (World Scientific, Singapore, 1991).
17. C. Rovelli, Class. Quantum Grav. **8** (1991) 1613.
18. A. Sen, J. Math. Phys. **22** (1981) 1781.
19. A. Ashtekar, Phys. Rev. Lett. **57** (1986) 2244; Phys. Rev. **D36** (1987) 1587.
20. C. Rovelli and L. Smolin, Phys. Rev. Lett. **61** (1989) 1155; Nucl. Phys. **B331** (1990) 80.
21. M. P. Blencowe, Nucl. Phys. **B341** (1990) 213.
22. H. Kodama, Prog. Theor. Phys. **80** (1988) 1024; Phys. Rev. **D42** (1990) 2548.
23. C. Soo and L. N. Chang, Int. J. Mod. Phys. **D3** (1994) 529.

24. L. Smolin and C. Soo, Nucl. Phys. **B449** (1995) 289.

25. J. Louko, Phys. Rev. **D51** (1995) 586.

26. T. Regge, Nuovo Cimento **19** (1961) 558.

27. R. M. Williams and P. A. Tuckey, Class. Quantum Grav. **9** (1992) 1409.

28. L. Bombelli, J. Lee, D. Meyer and R. D. Sorkin, Phys. Rev. Lett. **59** (1987) 521.

29. C. J. Isham, Class. Quantum Grav. **6** (1989) 1509.

30. C. J. Isham, Y. Kubyshin and P. Renteln, Class. Quantum Grav. **7** (1990) 1053.

31. J. A. Wheeler, in *Relativity, Groups and Topology*, eds. C. DeWitt and B. S. DeWitt, (Gordon and Breach, New York, 1963), p. 315.

32. J. A. Wheeler, in *Batelles Rencontres*, eds. C. DeWitt and J. A. Wheeler, (Benjamin, New York, 1968), p. 242.

33. B. S. DeWitt, Phys. Rev. **160** (1967) 1113.

34. C. W. Misner, Phys. Rev. **186** (1969) 1319; in *Relativity*, eds. M. Carmeli, Finkler and L. Witten, (Plenum, New York, 1970), p. 55.

35. C. W. Misner, in *Magic Without Magic: John Archibald Wheeler*, ed. J. Klauder, (W.H. Freeman, San Francisco, 1972), p. 441.

36. S. W. Hawking, in *Astrophysical Cosmology*, eds. H. A. Brück, G. V. Coyne and M. S. Longair, (Pontifica Academia Scientarium *Scripta Varia* **48**, Vatican City, 1982) p. 563.

37. A. Vilenkin, Phys. Lett. **117B** (1982) 25; Phys. Rev. **D27** (1983) 2848.

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

39. S. W. Hawking, Nucl. Phys. **B239** (1984) 257.

40. A. Vilenkin, Phys. Rev. **D33** (1986) 3560.

41. A. Vilenkin, Phys. Rev. **D37** (1988) 888.

42. A. Vilenkin, Phys. Rev. **D50** (1994) 2581.

43. J. B. Hartle, in *Quantum Cosmology and Baby Universes*, eds. S. Coleman, J. B. Hartle T. Piran and S. Weinberg, (World Scientific, Singapore, 1991), p. 65.

44. J. J. Halliwell, J. Perez-Mercader, and W. H. Zurek (eds.), *The Physical Origins of Time Asymmetry*, (Cambridge University Press, 1994).

45. A. Strominger, in *Quantum Cosmology and Baby Universes*, eds. S. Coleman, J. B. Hartle, T. Piran and S. Weinberg, (World Scientific, Singapore, 1991), p. 269.

46. A. Csordás and R. Graham, Phys. Rev. Lett. **74** (1995) 4129.

47. R. M. Wald, *General Relativity*, (Chicago University Press, 1984).

48. J. W. York, Phys. Rev. Lett. **28** (1972) 1082.

49. P. A. M. Dirac, Proc. R. Soc. Lond. **A246** (1958) 326; 333.

50. A. Hanson, T. Regge and C. Teitelboim, *Constrained Hamiltonian Systems*, (Accademia Nationale dei Lincei, Rome, 1976).

51. P. W. Higgs, Phys. Rev. Lett. **1** (1958) 373.

52. S. W. Hawking and D. N. Page, Nucl. Phys. **B264** (1986) 185.

53. G. W. Gibbons and S. W. Hawking, Phys. Rev. **D15** (1977) 2752.

54. S. W. Hawking, in *General Relativity: An Einstein Centenary Survey* (Cambridge University Press, 1979), p. 746.

55. J. B. Hartle and S. W. Hawking, Phys. Rev. **D13** (1976) 2188.

56. G. W. Gibbons and S. W. Hawking, Phys. Rev. **D15** (1977) 2738.

57. G. W. Gibbons and M. J. Perry, Proc. R. Soc. Lond. **A358** (1978) 467.

58. T. Eguchi and A. J. Hanson, Phys. Lett. **74B** (1978) 249.

59. G. W. Gibbons and S. W. Hawking, Phys. Lett. **78B** (1978) 430; Commun. Math. Phys. **66** (1979) 291.

60. G. W. Gibbons and C. N. Pope, Commun. Math. Phys. **66** (1979) 267.

61. G. W. Gibbons, in *Fields and Geometry*, ed. A. Jadczyk, (World Scientific, Singapore, 1986).

62. D. Garfinkle and A. Strominger, Phys. Lett. **256B** (1991) 146.

63. H. F. Dowker, J. P. Gauntlett, S. B. Giddings and G. T. Horowitz, Phys. Rev. **D50** (1994) 2662.

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

65. K. Schleich, Phys. Rev. **D36** (1987) 2342.

66. J. J. Halliwell and J. Louko, Phys. Rev. **D39** (1989) 2206.

67. J. J. Halliwell and J. Louko, Phys. Rev. **D40** (1989) 1868.

68. J. J. Halliwell and J. Louko, Phys. Rev. **D42** (1990) 3997.

69. A. D. Linde, Zh. Eksp. Teor. Fiz. **87** (1984) 369 [Sov. Phys. JETP **60** (1984) 211]; Lett. Nuovo Cimento **39** (1984) 401; Rep. Prog. Phys. **47** (1984) 925.

70. A. Vilenkin, Phys. Rev. **D30** (1984) 509; Nucl. Phys. **B252** (1985) 141.

71. J. J. Halliwell and J. B. Hartle, Phys. Rev. **D43** (1991) 1170.

72. K. V. Kuchař and M. P. Ryan, Phys. Rev. **D40** (1989) 3982.

73. S. Sinha and B. L. Hu, Phys. Rev. **D44** (1991) 1028.

74. F. D. Mazzitelli, Phys. Rev. **D46** (1992) 4758.

75. A. Ishikawa and T. Isse, Mod. Phys. Lett. **A8** (1993) 3413.

76. M. A. H. MacCallum, in *General Relativity: An Einstein Centenary Survey* eds. S. W. Hawking and W. Israel, (Cambridge University Press, 1979), p. 533.

77. A. S. Kompaneets and A. S. Chernov, Zh. Eksp. Teor. Fiz. **47** (1964) 1939 [Sov. Phys. JETP **20** (1965) 1303].

78. R. Kantowski and R. K. Sachs, J. Math. Phys. **7** (1966) 443.

79. L. Bianchi, Mem. di Mat. Soc. Ital. Sci. **11** (1897) 267.

80. C. W. Misner, Phys. Rev. Lett. **22** (1969) 1071.

81. V. A. Belinskii, E. M. Lifschitz and I. M. Khalatnikov, Usp. Fiz. Nauk. **102** (1970) 463 [Sov. Phys. Usp. **13** (1970) 745].

82. A. Latifi, M. Musette and R. Conte, Phys. Lett. **194A** (1994) 83.

83. J. Louko, Ann. Phys. (N.Y.) **181** (1988) 318.

84. J. J. Halliwell, Phys. Rev. **D38** (1988) 2468.

85. I. G. Moss, Ann. Inst. H. Poincaré **49** (1988) 341.

86. D. N. Page, J. Math. Phys. **32** (1991) 3427.

87. T. Banks, Nucl. Phys. **B249** (1985) 332.

88. S. P. Kim, Phys. Rev. **D52** (1995) 3382.

89. J. J. Halliwell, Phys. Rev. **D39** (1989) 2912.

90. E. Calzetta and J. J. Gonzalez, Phys. Rev. **D51** (1995) 6821.

91. S. Habib and R. Laflamme, Phys. Rev. **D42** (1990) 4056.

92. E. Calzetta, Phys. Rev. **D43** (1991) 2498.

93. B. L. Hu, J. P. Paz and S. Sinha, in *Directions in General Relativity, Vol. I*, eds. B. L. Hu, M. P. Ryan and C. V. Vishveshwara, (Cambridge University Press, 1993), p. 145.

94. B. L. Hu, in [44], p. 475.

95. J. J. Halliwell, Phys. Rev. **D36** (1987) 3626.

96. K. V. Kuchař, J. Math. Phys. **22** (1981) 2640.

97. M. P. Ryan, *Hamiltonian Cosmology* (Springer *Lecture Notes in Physics* **13**, Berlin, 1972).

98. C. Teitelboim, Phys. Rev. Lett. **38** (1977) 1106.

99. R. Tabensky and C. Teitelboim, Phys. Lett. **69B** (1977) 453.

100. A. Macías, O. Obregón and M. P. Ryan, Class. Quantum Grav. **4** (1987) 1477.

101. P. D. D'Eath and D. I. Hughes, Phys. Lett. **214B** (1988) 498; Nucl. Phys. **B378** (1992) 381.

102. R. Graham, Phys. Rev. Lett. **67** (1991) 1381; Phys. Lett. **277B** (1992) 393.

103. P. D. D'Eath, S. W. Hawking and O. Obregón, Phys. Lett. **300B** (1993) 44.
104. A. Vilenkin, Phys. Rev. **D39** (1989) 1116.
105. D. N. Page, Phys. Rev. **D34** (1986) 2267.
106. J. B. Hartle, Phys. Rev. **D37** (1988) 2818, **D43** (1991) 1434.
107. D. N. Page, Phys. Rev. Lett. **70** (1993) 4034.
108. S. W. Hawking and D. N. Page, Nucl. Phys. **B298** (1988) 789.
109. E. W. Kolb and M. S. Turner, *The Early Universe*, (Addison-Wesley, Reading, Mass., 1990).
110. D. L. Wiltshire, Phys. Rev. **D36** (1987) 1634.
111. M. Henneaux, Lett. Nuovo Cimento **38** (1983) 609.
112. G. W. Gibbons, S. W. Hawking and J. M. Stewart, Nucl. Phys. **B281** (1987) 736.
113. W. M. Suen and K. Young, Phys. Rev. **D39** (1989) 2201.
114. H. D. Conradi and H. D. Zeh, Phys. Lett. **154A** (1991) 321.
115. H. D. Conradi, Phys. Rev. **D46** (1992) 612, Class. Quantum Grav. **12** (1995) 2423.
116. G. W. Lyons, Phys. Rev. **D46** (1992) 1546.
117. J. J. Halliwell and J. B. Hartle, Phys. Rev. **D41** (1990) 1815.
118. G. W. Gibbons and L. P. Grishchuk, Nucl. Phys. **B313** (1989) 736.
119. D. N. Page, in *Quantum Concepts in Space and Time*, eds. R. Penrose and C. J. Isham, (Oxford University Press, 1986) p. 274.
120. D. J. Kaup and A. P. Vitello, Phys. Rev. **D9** (1974) 1648.
121. S. W. Hawking and D. N. Page, Phys. Rev. **D42** (1990) 2655.
122. D. N. Page, Class. Quantum Grav. **7** (1990) 1841.
123. J. J. Halliwell and R. C. Myers, Phys. Rev. **D40** (1989) 4011.
124. I. Klebanov, L. Susskind, and T. Banks, Nucl. Phys. **B317** (1989) 665.
125. A. Vilenkin, in *Proceedings of the International School of Astrophysics 'D. Chalonge', Eriche, 1995*, ed. N. Sanchez, (Kluwer, Dordrecht), to appear.
126. J. Louko and T. Vachaspati, Phys. Lett. **223B** (1989) 21.
127. A. Vilenkin, Phys. Rev. Lett. **74** (1995) 846.
128. A. O. Barvinsky and A. Yu. Kamenshchik, Class. Quantum Grav. **7** (1990) L181.
129. A. O. Barvinsky, Phys. Rep. **230** (1993) 237.
130. L. P. Grishchuk and L. V. Rozhansky, Phys. Lett. **234B** (1990) 9.
131. A. Lukas, Phys. Lett. **347B** (1995) 13.
132. L. P. Grishchuk and L. V. Rozhansky, Phys. Lett. **208B** (1988) 369.

133. J. J. Halliwell and S. W. Hawking, Phys. Rev. **D31** (1985) 1777.
134. S. Wada, Nucl. Phys. **B276** (1986) 729.
135. J. J. Halliwell and P. D'Eath, Phys. Rev. **D35** (1985) 1100.
136. R. Laflamme, Phys. Lett. **198B** (1987) 156.
137. S. Wada, Phys. Rev. Lett. **59** (1987) 2375.
138. T. Vachaspati, Phys. Lett. **217B** (1989) 228.
139. T. Vachaspati and A. Vilenkin, Phys. Rev. **D37** (1988) 898.
140. T. S. Bunch and P. C. W. Davies, Proc. R. Soc. Lond. **A360** (1978) 117.
141. B. Allen, Phys. Rev. **D32** (1985) 3136.
142. S. W. Hawking, Phys. Lett. **115B** (1982) 295.
143. A. H. Guth and S. Y. Pi, Phys. Rev. Lett. **49** (1982) 1110; Phys. Rev. **D32** (1985) 1899.
144. K. M. Gorski, G. Hinshaw, A. J. Banday, C. L. Bennett, E. L. Wright, A. Kogut, G. F. Smoot, and P. Lubin, Astrophys. J. **430** (1994) L89.
145. T. Gold (ed.), *The Nature of Time*, (Cornell University Press, Ithaca, 1967).
146. A. Higuchi and R. M. Wald, Phys. Rev. **D51** (1995) 544.
147. S. C. Beluardi and R. Ferraro, Phys. Rev. **D52** (1995) 1963.
148. D. Marolf, Class. Quantum Grav. **12** (1995) 2469.
149. S. W. Hawking, Phys. Rev. **D32** (1985) 2489.
150. T. Gold, in *La Structure at l'Évolution de l'Univers, 11th Solvay Conference*, (Edition Stoops, Brussels, 1958), p. 81; Am. J. Phys. **30** (1962) 403.
151. S. W. Hawking and Z. C. Wu, Phys. Lett. **151B** (1985) 15.
152. S. W. Hawking, Vist. Astr. **37** (1993) 559; and in [44], p. 346.
153. D. N. Page, Phys. Rev. **D32** (1985) 2496.
154. R. Laflamme and E. P. S. Shellard, Phys. Rev. **D35** (1987) 2315.
155. R. Laflamme, *Time and Quantum Cosmology*, (Ph.D. thesis, University of Cambridge, 1988); in [44], p. 358.
156. S. W. Hawking, R. Laflamme and G. W. Lyons, Phys. Rev. **D47** (1993) 5342.
157. C. Kiefer and H. D. Zeh, Phys. Rev. **D51** (1995) 4145.
158. P. D'Eath, Phys. Rev. **D48** (1993) 713.
159. S. W. Hawking, Phys. Rev. **D37** (1988) 904.
160. R. Graham and H. Luckock, Phys. Rev. **D49** (1994) R4981.
161. B. de Wit, H. Nicolai and H. J. Matschull, Phys. Lett. **318B** (1994) 115.
162. S. M. Carroll, D. Z. Freedman, M. E. Ortiz and D. N. Page, Nucl. Phys. **B423** (1994) 661.