PHYSICS REPORTS (Review Section of Physics Letters) 137, No. 4 (1986) 231–275. North-Holland, Amsterdam

INTERPRETING THE WAVE FUNCTION OF THE UNIVERSE

Frank J. TIPLER

Department of Mathematics and Department of Physics, Tulane University, New Orleans, Louisiana 70118, U.S.A.

Received April 1985

Contents:

Abstract:

The Many-Worlds Interpretation of quantum mechanics is used to determine the meaning of the universal wave function of quantum cosmology. More precisely, the Many-Worlds Interpretation is used to distinguish those quantities in quantum cosmology which are measurable, and hence physically meaningful, from those which are not. A number of rather surprising conclusions are drawn from the analysis. First, it is not possible to measure the expectation value of any universal operator and so such expectation values are physically meaningless. Second, it is physically (although not mathematically) meaningless to talk about eigenstates of the Universal Hamiltonian. Third, before a measurement of the radius of the Universe is made it makes no sense to say the Universe has a radius. Fourth, after the first *two* radius measurements are made, the Universe can be said to have *all* radii consistent with the support of the universal wave function and the wave function of the measuring apparatus; in effect, the Universe splits into an infinite number of branches, in each of which the time evolution is observed to be very close to the classical evolution. Fifth, any universal wave function which does not violate a Quantum Copernican Cosmological Principle will automatically solve the Flatness Problem without having to invoke inflation. In effect, quantum cosmology allows us to have "inflation without inflation". All of these conclusions are illustrated with a closed Friedmann universe quantized in conformal time. My quantization procedure allows only one solution to Schrödinger's equation, and this solution solves the Flatness Problem. I show that the ADM quantization method plus the Hartle–Hawking initial foundary condition gives the same result.

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 137, No. 4 (1986) 231–275.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 31.00, postage included.

INTERPRETING THE WAVE FUNCTION OF THE UNIVERSE

Frank J. TIPLER

Department of Mathematics and Department of Physics, Tulane University, New Orleans, Louisiana 70118, U.S.A.

NORTH-HOLLAND–AMSTERDAM

1. Introduction

In the past few years there has been a vast upsurge of interest in quantum gravity, particularly as applied to cosmology. There are several reasons for this. First and foremost is Hawking's remarkable discovery that particle creation must occur in the strong gravitational field of a black hole. Hawking's result supported the earlier work by Zel'dovich and Parker indicating that quantum effects in the strong gravitational field of the very early universe could dominate the evolution in this regime. Path integral techniques developed by Hawking and Hartle to study particle creation around an evaporating black hole were soon applied to particle creation in the early universe. These techniques were originally used only semi-classically, with the gravitational field being considered as a classical field, but Hawking has spent the last few years attempting to find a path integral formulation of a fully quantized general relativity theory, culminating in his paper with Hartle wherein the authors claimed to have discovered "The Wave Function of the Universe [1]". But it is not enough to discover the wave function of the Universe. It is also necessary to discover what such a function means; that is, it is necessary to analyze the physical significance of the wave function representing the quantum state of the Universe. It is the main purpose of the present paper to accomplish this task of interpreting the Universal Wave Function.

Surprisingly, no one has attempted to do this before in a really systematic way. Most studies in quantum cosmology concentrate on the problem of obtaining the appropriate Schrödinger equation for the wave function. The problem of interpreting the wave function is them attacked by analyzing the motion of a wave packet, or by looking at the time evolution of the expectation value of some observable. It is assumed that the motion of a wave packet or the evolution of an expectation value (of the radius of the Universe, say) would resemble the observed evolution of the universe.

We shall see that these assumptions are not necessarily true. To show this it will be sufficient to consider the simplest available minisuperspace model, the dust or radiation gas closed Friedmann universe, wherein the only degree of freedom to be quantized is R, the radius of the universe. The classical theory of such a cosmology will be reviewed at length in section 2, because although the Friedmann universe has been known for fifty years and has consequently been studied, several points which turn out to be crucial to the quantum theory of the model are obscured in the usual elementary treatments.

Interpreting the wave function of the universe requires adopting an interpretation of quantum mechanics. I shall adopt the Many-Worlds Interpretation (MWI), because this interpretation seems tailor made for quantum cosmology. Indeed, Hugh Everett, the inventor of the MWI, asserted in the first paragraph of his first paper on this interpretation, that a major motivation behind his development of the Many-Worlds formalism was to find a language in which the notion of the "wave function of the Universe" makes physical sense [2]. A similar justification for the MWI has been given by virtually all the proponents of the MWI at one time or another.

In spite of this, no one seems to have applied the MWI to quantum cosmology except in an off-hand manner [16]. Even Bryce DeWitt, the main proponent of the MWI and the physicist responsible for bringing the MWI to the attention of the general physics community, did not himself use the MWI to interpret the quantized Friedmann model which he developed. His only mention of Everett's theory in his classic paper "Quantum Theory of Gravity" occurred in the final philosophical section [3]. As a consequence, most cosmologists are unfamiliar with the MWI, except perhaps as a vague, almost philosophical idea. Even worse, I've discovered in conversations with my fellow cosmologists that the impressions which many of them have about what is implied by the MWI are actually false!

I shall endeavor to correct this unfortunate deficiency in section 3 by developing the MWI in detail by presenting a toy model of the measurement process according to the MWI, and illustrating the process by a concrete example: the Wilson cloud chamber measurement of particle tracks, a problem first considered by Heisenberg [4] and Mott [5]. This example highlights an important point mentioned above: the motion of a wave packet or the time evolution of the expectation value of the particle position in general may give a wholly misleading picture of what an apparatus designed to measure the motion of the particle would actually see. In the case of the Wilson cloud chamber, the wave function for the particle is a stationary plane wave, which is not normalizable and indeed has constant modulus over all space! So how do we see isolated particle tracks in the cloud chamber? All will be explained in section 3.

I shall develop in section 3 a formalism sufficiently general to discuss the problem of what an apparatus designed to measure the radius of the Universe at sequential times would actually see and record. We shall see that before the first two radius measurements are made, it is meaningless to say the Universe has a radius, while after these measurements are made, at best the Universe can be said to have all radii consistent with the support of the respective wave functions of the Universe and apparatus.

In section 4 I shall apply the formalism to a quantized version of the classical Friedmann universe discussed in section 2. A number of well-known technical problems arise in quantizing the Friedmann model, among them the choice of the time parameter, the way in which the constraint Einstein equations are to be taken into account, and the appropriate boundary conditions. These points will be covered in sections 2 and 4. I shall argue that conformal time is the most appropriate time parameter. I shall impose the physical condition that the domain of the Universal wave function be restricted to positive values of R, where R is the Friedmann scale factor (the spatial quantization variable). This condition has a considerable physical effect if we require the quantum Hamiltonian to be a self-adjoint operator, as has been previously noted by Gotay and Demaret [7]. I shall discuss this point at length in section 4. One rather surprising fact to emerge from this analysis is the effect of the $R = 0$ curvature singularity on the time evolution of the wave function: the singularity affects the time evolution at *all* times rather than just at the beginning and the end as is the case of the classical closed universe. Most papers on quantum cosmology attempt to determine the effect of the singularity by studying the time evolution of the expectation value $\langle R \rangle$ of the radius of the Universe. I shall show in section 4 using the formalism of section 3 that this is an incorrect procedure, for it is not possible to measure the expectation value of *any* Universal operator, even in principle. Similarly, the assertion that the Universe is in an eigenstate of the Universal Hamiltonian is physically meaningless.

Although the formalism of section 3 can be applied to any universal wave function to determine the observed time evolution, we are of course most interested in applying it to a wave function which hopefully is close to the true wave function, and this will be obtained by imposing the appropriate boundary conditions and deciding on the method of taking in account the constraint equations. It turns out that the problems of the boundary conditions and the problem of the constraints are intimately connected. I shall handle them by an unorthodox technique: I shall quantize an *unconstrained* Hamiltonian, and then take the constraints into account by imposing boundary conditions on the solutions to Schrödinger's equation. The reader is warned that this procedure may or may not give an accurate quantum model. I shall nevertheless follow this procedure because the resulting model is exceedingly simple and illustrates the interpretation formalism without introducing mathematical complexities. The formalism can be of course applied to quantum models obtained by other quantization procedures. In particular, I shall show in the Appendix that the ADM method gives exactly the same Schrödinger equation as my unorthodox method in the case of the radiation-filled closed Friedmann universe.

I shall argue that my method of imposing the constraints allows just one solution to the Schrödinger equation using the unconstrained Hamiltonian. In other words, my procedure of quantizing the universe has the property which Wheeler [33] and DeWitt [3] conjectured the correct quantum theory for the Universe must have: a single, unique solution for the cosmological Schrödinger equation. I shall discuss the various properties of this unique solution in section 4. Besides uniqueness, it has several other nice properties: (1) the time evolution of the universal radius as measured by an observer inside the universe is virtually identical to the classical motion even near the singularity, except of course for the inevitable but very small limitations imposed by the uncertainty principle; and (2), all classical closed Friedmann universes are equally probable in the sense that when the Universe is split into branches by the first two measurements of the Universal radius, the probability that we are in a given universe does not depend on the value of R. In section 4, I shall propose this unique universal wave function as a representation of *the* wave function of the Universe; more precisely, I shall argue that it probably models the features of the actual quantum Universe at least as well as the classical Friedmann universe models the observed structure of the branch of the total Universe we happen to be in. Even if my unorthodox quantization procedure leaves a lot to be desired, I expect the essential properties of the unique solution to persist in the wave function obtained by a more correct procedure. For evidence, I shall show in the Appendix that the same solution is the most physically natural solution of the Wheeler–DeWitt equation obtained via the ADM quantization method.

My proposed Universal wave function has another interesting property: the probability is essentially one that we are in a branch of the Universe in which the actual matter density is arbitrarily close to the critical density. This is true even though no inflationary phase is assumed to occur. The stress-energy tensor is assumed to be radiation for the entire evolutionary history of the Universe. I shall show in section 4 that this property – we might call it "inflation without inflation" – is a property of any wave function which satisfies a Quantum Copernican Cosmological Principle: all closed universes are equally probable.

In the concluding section 5, I shall summarise the conclusions implied by the Many-Worlds approach to quantum cosmology, and discuss possible directions of future work from this point of view. In particular, I shall touch on the Horizon Problem, the Magnetic Monopole Problem, and the Initial Perturbation Problem.

I shall be using standard Planck units ($c = h/2\pi = G = 1$) unless these constants are explicitly included in the equations. For example, the Einstein equations are $R_{ab} - \frac{1}{2}g_{ab}R = 8\pi T_{ab}$. I shall capitalize the word "Universe" when I use it to refer to the traditional object of study in cosmology, and leave it uncapitalized when I am referring to a classical model, or to system/apparatus collection. See section 3 for a thorough discussion of this point.

2. The classical closed Friedmann universe

The deepest insight into the significance of a physical theory is obtained by expressing it in terms of an action principle. The full action in Einstein's gravitational theory is

$$S = \frac{1}{16\pi} \int_{\mathrm{M}} \mathcal{R}\sqrt{-g}\, \mathrm{d}^4x + \frac{1}{8\pi} \int_{\partial\mathrm{M}} \mathrm{tr}\, K\sqrt{\gamma}\, \mathrm{d}^3x + \int_{\mathrm{M}} L_{\mathrm{m}}\sqrt{-g}\, \mathrm{d}^4x + C \tag{2.1}$$

where ∂M is the boundary of the four-dimensional region M, having extrinsic curvature K and intrinsic metric γ. The spacetime metric is g, from which the Ricci curvature $\mathcal{R}$ is obtained. The matter

Lagrangian is L_m. The constant C is a boundary term which must be considered in open universes, including asymptotically flat spacetimes [8], but may be set to zero in closed universes. The action (2.1) is a global object and is well-defined only if the global topology is fixed [9, 10, 11; see however, 1 and 12]. I shall consider here only the standard Friedmann closed universe, so the global topology is $S^3 \times \mathbb{R}^1$. The form of the action (2.1) makes it clear why theorists from Einstein [13] to Misner, Thorne and Wheeler [14] have regarded closed universes as more physically reasonable than open universes. The latter have the boundary term C to deal with, and the surface integral in (2.1) is also more complicated in open universes, since in this case it must contain timelike portions. For closed universes, the surface integral can be taken over two disjoint spacelike hypersurfaces, or removed altogether by collapsing the boundary ∂M onto the initial and final singularities. For closed universes one would have only the two volume integrals in (2.1), and these terms would be finite even if M were chosen to be the entire spacetime (for suitable choices of the matter Lagrangians). In the open universe case the boundary terms come into the theory in a fundamental way, and there is no good way to decide what these terms should be for arbitrary open universes [13, 14].

I shall be concerned in this paper only with closed Friedmann universes, which means I shall consider only variations in the action (2.1) which preserve isotropy and homogeneity. Taking the path integral view of quantum mechanics, this means we shall consider only those paths in which the radius of the universe varies; paths in which the homogeneity or isotropy varies will be omitted from the Feynmann sum. It is well-known that the metric for such a universe can be written

$$ds^2 = -dt^2 + R^2(t)\,[d\chi^2 + \sin^2\chi\,(d\theta^2 + \sin^2\theta\, d\phi^2)] \tag{2.2}$$

where the spatial variables have ranges $0 \le \chi \le \pi$, $0 \le \theta \le \pi$, $0 \le \phi \le 2\pi$. We immediately have $\sqrt{-g} = \sqrt{\gamma} = R^3(t)\sin^2\chi \sin\theta$, and

$$\mathcal{R} = 6[R^{-1}\ddot{R} + R^{-2}(1 + \dot{R}^2)] \tag{2.3}$$

with the dot denoting differentiation with respect to t, and

$$K_{ij} = -R^{-1}\dot{R}\delta_{ij}\,, \qquad \mathrm{tr}\,K = g^{ij}K_{ij} = \delta^{ij}K_{ij} = -3R^{-1}\dot{R} \tag{2.4}$$

in a local Euclidean frame, if we choose ∂M to be two disjoint hypersurfaces of homogeneity and isotropy. Consider first the purely gravitational part of (2.1):

$$S_{\mathrm{grav}} = \frac{1}{16\pi}\int_{M} \mathcal{R}\sqrt{-g}\, d^4x + \frac{1}{8\pi}\int_{\partial M} \mathrm{tr}\,K\sqrt{\gamma}\, d^3x\,. \tag{2.5}$$

A straightforward calculation, including an integration by parts of the first term in (2.5) shows that G_{grav} can be written

$$S_{\mathrm{grav}} = \frac{3V}{8\pi}\int_{t_1}^{t_2} (1 - \dot{R}^2)\, R\, dt = \frac{3}{8\pi}\int_{M} \left[\frac{1}{R^2} - \frac{\dot{R}^2}{R}\right]\sqrt{-g}\, d^4x \tag{2.6}$$

where t_1, t_2 are the values of the proper time on the future and past boundaries, respectively, and

$$V=\int_0^{\pi}\int_0^{\pi}\int_0^{2\pi}\sin^2\chi\sin\theta\,\mathrm{d}\chi\,\mathrm{d}\theta\,\mathrm{d}\phi=\int\sqrt{\gamma}\,\mathrm{d}^3x=2\pi^2$$

is the scale-factor independent volume of one of the spacelike hypersurfaces forming the boundary of the region M. Equation (2.6) shows very clearly that the effect of the boundary term is to remove second derivatives from the gravitational action. Like all other fundamental fields in physics, the gravitational field has an action which is a functional of the basic field – the metric – and its first derivatives only. It is this fact that allows the quantization of the gravitational field with standard techniques (at least in the minisuperspace case).

The gravitational action simplifies considerably if the time variable is chosen to be conformal time t, defined by

$$R(t)\,\mathrm{d}/\mathrm{d}t=\mathrm{d}/\mathrm{d}\tau$$

or

$$\mathrm{d}t=R(t(\tau))\,\mathrm{d}\tau\,. \tag{2.7}$$

In conformal time the gravitational part of the action becomes

$$S_{\mathrm{grav}}=\frac{3V}{8\pi}\int_{\tau_1}^{\tau_2}(R^2-R'^2)\,\mathrm{d}\tau \tag{2.8}$$

where the prime denotes differentiation with respect to conformal time. The parameter τ is called conformal time because in terms of τ the metric (2.2) can be written

$$\mathrm{d}s^2=R^2(\tau)\,[-\mathrm{d}\tau^2+\sin^2\chi\,(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\phi^2)] \tag{2.9}$$

so the metric (2.2) is conformally related to the metric of the Einstein static universe, given by the metric inside the brackets in (2.9); see [15] for a detailed discussion.

Let us now consider the action for the matter. I shall enormously simplify the analysis if I assume the matter to be an isentropic perfect fluid. The techniques for dealing with an isentropic perfect fluid action integral have been developed by Hawking and Ellis [15] section 3.3. With my normalization, the Lagrange density for such a fluid can be written

$$L_{\mathrm{m}}=-\rho(1-\varepsilon)=-\mu \tag{2.10}$$

where ρ is the mass density, $\varepsilon=\varepsilon(p)$ is the fluid internal energy, p is the pressure, and μ is the total energy density of the fluid, consisting of rest mass and internal energy. Hawking and Ellis point out that in order to obtain the correct dynamical equations from a variation of the action, it is necessary to require the current vector J^{a}, defined by $J^{\mathrm{a}}=\rho u^{\mathrm{a}}$ where u^{a} is the unit tangent vector to the fluid flow lines, to be covariantly conserved:

$$J^a{}_{;a} = \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^a} (\sqrt{-g} J^a) = 0 . \tag{2.11}$$

Since in our case we have frozen all degrees of freedom except the radius of the universe, there will be no variation of the action in any variable except $R(t)$. This means u^a will remain normal to the hypersurfaces of homogeneity and isotropy even along histories in which the action is not an extremum. Thus (2.11) becomes simply

$$\frac{d}{dt}(R^3(t)\rho) = 0 = \frac{d}{d\tau}(R^3(\tau)\rho) . \tag{2.12}$$

For an isentropic perfect fluid with polytropic index γ, the pressure and total energy density are related by

$$p = (\gamma - 1)\mu . \tag{2.13}$$

The pressure is given in terms of the internal energy and mass density by

$$p = \rho^2 \, d\varepsilon/d\rho . \tag{2.14}$$

Equating (2.13) and (2.14), and using (2.10) to eliminate μ, we get the differential equation $d\varepsilon/(1+\varepsilon) = (\gamma - 1)\, d\rho/\rho$, which can be immediately integrated to yield $1 + \varepsilon = \text{constant} \times \rho^{\gamma-1}$. Putting this expression back into (2.10) gives

$$\mu = \text{constant} \times \rho^{\gamma} . \tag{2.15}$$

But eq. (2.12) implies that $R^3(t)\rho$ is a constant, so if this fact is used to eliminate ρ from (2.15), we finally obtain for the matter Lagrange density

$$L_m = -\mu = \tilde{C} R^{-3\gamma} \tag{2.16}$$

where $\tilde{C}$ is a constant. The action for the matter is

$$S_{\text{matter}} = \int \tilde{C} R^{-3\gamma} \sqrt{-g}\, d^4x = V\tilde{C} \int_{t_1}^{t_2} R^{-3(\gamma-1)}\, dt = V\tilde{C} \int_{\tau_1}^{\tau_2} R^{-3\gamma+4}\, d\tau \tag{2.17}$$

so the total action is

$$S = S_{\text{grav}} + S_{\text{matter}} = \frac{-3V}{8\pi} \int_{\tau_1}^{\tau_2} [(R')^2 - R^2 + CR^{-3\gamma+4}]\, d\tau \tag{2.18}$$

where $C = -8\pi\tilde{C}/3$ is a constant. The total Lagrangian will be quadratic in three cases:

$$-3\gamma + 4 = \begin{cases} 0 \Rightarrow \gamma = \frac{4}{3}, \text{ radiation gas} \\ 1 \Rightarrow \gamma = 1, \text{ dust} \\ 2 \Rightarrow \gamma = \frac{2}{3}, \text{ unphysical, since it implies a negative pressure} . \end{cases}$$

For the radiation gas, varying with respect to the metric gives the Lagrange equation

$$\mathrm{d}^2R/\mathrm{d}\tau^2 + R = 0 \tag{2.19}$$

since the constant term in the Lagrangian can be omitted. The general solution to (2.19) is of course

$$R(\tau) = \tilde{R}\sin(\tau + \delta) . \tag{2.20}$$

The two integration constants in (2.20) can be evaluated in the following way. It is clear that all solutions (2.20) have zeros with the same period π. Since it is physically meaningless to continue a solution through a singularity which occurs at every zero, all solutions exist only over a τ-interval of length π. Thus for all solutions we can choose the phase delta so that for all solutions the zero of τ-time occurs at the beginning of the universe, at $R = 0$. This implies $\delta = 0$ for all solutions, in which case the remaining constant $\tilde{R}$ is seen to be the radius of the universe at maximum expansion:

$$R(\tau) = R^{\text{rad}}_{\text{max}} \sin\tau . \tag{2.21}$$

In the radiation gas case, all solutions are parameterized by a single number $R^{\text{rad}}_{\text{max}}$, the radius of the universe at maximum expansion. It is important to note I have obtained the standard result (2.21) without having to refer to the Friedmann constraint equation. Indeed, I obtained the dynamical equation (2.19) by an unconstrained variation of the Lagrangian (2.18); I obtained the correct dynamical equation and the correct solution even though I ignored the constraint. The constraint equation contained no information that was unavailable in the dynamical equation, obtained by unconstrained variation, except for the tacit assumption that $\rho \neq 0$. From the point of view of the dynamical equation, the vacuum "radiation gas" is an acceptable "solution". For a true radiation gas ($\rho \neq 0$) at least, ignoring the constraints is a legitimate procedure. It is well this is so, for I have precluded any possibility of obtaining the Friedmann constraint equation by fixing the lapse N before carrying out the variation (in effect I chose $N = R(\tau)$). The fact that the constraint can be ignored in the radiation case is important because quantizing a constrained system is loaded with ambiguities [3, 17]; indeed, the problem of quantizing Einstein's equations is mainly the problem of deciding what to do with the constraint equations [17], and these ambiguities do not arise in the unconstrained case (see [18] chapter 21 and the Appendix for a discussion of the relationship between the lapse and the Einstein constraint equations).

The constraint equation in the radiation gas case tells us two things: the density cannot be zero, and the solutions hit the singularity. Thus as long as these implications of the constraints are duly taken into account in some manner in the quantum theory, quantizing an unconstrained system should be a legitimate procedure, at least for a radiation gas.

It is not possible to ignore the constraint in the dust-filled universe. With $\gamma = 1$ an unconstrained variation with respect to the metric gives the Lagrange equation

$$\mathrm{d}^2R/\mathrm{d}\tau^2 + R = C/2 \tag{2.22}$$

for which the general solution is

$$R(\tau) = \tilde{R}\cos(\tau + \delta) + C/2\,. \tag{2.23}$$

As far as we could tell by our previous analysis, the constant C is restricted only by the condition $C > 0$, which comes from the strong energy condition [15]. Thus it would seem that a judicious choice of C would enable us to avoid zeros of $R(\tau)$ – i.e., singularities – altogether. Alas, we know this isn't true. It is the constraints that force all solutions to the dust dynamical equations through $R = 0$. The constraint equation provides a relation between $\tilde{R}$, C and δ.

One could obtain some information about this relation purely dynamically by varying the flow lines of the fluid, which can be regarded in the perfect fluid case as the analogue of varying the action with respect to the non-gravitational variables. Such a variation implies Raychaudhuri's equation and the geodesic nature of the flow lines [15]. For geodesic flow, Raychaudhuri's equation can be written $\ddot{R} + \tilde{C}R = 0$, which implies that all solutions (2.23) have periodic zeros. Requiring as in the radiation case that the zero of time in all solutions (2.23) is to be set at the initial singularity gives the relation $C = -2\tilde{R}\cos\delta$. The fact that we can set $\delta = 0$ and so obtain the complete set of Friedmann solutions in the form

$$R(\tau) = (R^{\mathrm{dust}}_{\mathrm{max}}/2)(1 - \cos\tau) \tag{2.24}$$

where $R^{\mathrm{dust}}_{\mathrm{max}}$ is the radius of the universe at maximum expansion, comes from the Friedmann constraint equation. Of course, once we know the answer (2.24) we can obtain that answer by starting with the solutions (2.23) on the unconstrained Lagrange equation and putting in by hand the requirements that all solutions have initial singularities and that all solutions have zero phase at the initial singularity. In short, the other dynamical equation and the constraint equation can be incorporated by a judicious choice of boundary conditions on the unconstrained dynamical equation (2.22).

The matter Lagrangian will still be quadratic (and so the unconstrained quantization will still be fairly straightforward) if it is the sum of two non-interacting perfect fluids, dust and radiation.

The total action for the two-fluid model is

$$S = -\frac{3V}{8\pi}\int_{\tau_1}^{\tau_2} [(R')^2 - R^2 + C^{\mathrm{dust}}R + C^{\mathrm{rad}}]\,\mathrm{d}\tau\,. \tag{2.25}$$

Since the constant C^{rad} gives no contribution to the unconstrained dynamical equation, the dynamical equation obtained by varying (2.25) with respect to the metric and ignoring the effect of constraints is just (2.22), with the general solution (2.23). As in the dust case, the constraint equation allows us to evaluate the constants $\tilde{R}$, δ and C^{dust}. The solution which takes into account the constraint can be written (ref. [18] p. 741):

$$R(\tau) = R^{\mathrm{dust\,\&\,rad}}_{\mathrm{max}}(1 - \sec\delta\cos(\tau + \delta))\,(1 + \sec\delta)^{-1} \tag{2.26}$$

where as before $R^{\mathrm{dust\,\&\,rad}}_{\mathrm{max}}$ is the radius of the universe at maximum expansion. In expressing the solution in the form (2.26), it is assumed that the initial singularity occurs at $\tau = 0$. We immediately obtain from (2.26) that the universe reaches its maximum radius at $\tau = \pi - \delta$, and it terminates in a final singularity at $\tau = 2(\pi - \delta)$. The phase angle delta is related to physical quantities via

$$\tan \delta = 3\mu_{\rm rad}(\tau)/[2\pi\mu^2_{\rm dust}(\tau)\, R^2(\tau)]^{1/2} \tag{2.27}$$

where $\mu_{\rm rad}(\tau)$ is the radiation density and $\mu_{\rm dust}(\tau)$ is the dust density. Thus $0 \le \delta \le \pi/2$. Also, since $\mu_{\rm rad}(\tau)$ is proportional to R^{-4} and $\mu_{\rm dust}(\tau)$ is proportional to R^{-3}, the right-hand side of (2.27), and hence the phase δ, is an invariant, independent of τ, which is determined by the universal photon per baryon ratio. For a constant τ, as $\mu_{\rm rad}/\mu^2_{\rm dust} \to 0$, $\delta \to 0$, and (2.26) reduces to (2.24). As $\mu_{\rm rad}/\mu^2_{\rm dust} \to +\infty$, $\delta \to \pi/2$, and (2.26) reduces to (2.21). Since the phase δ is determined by the photon per baryon ratio, it is an isentropic invariant, so cannot be expected to change in the isentropic models analyzed in this paper. This means that the length of time the model universe stays in existence, which is $2(\pi - \delta)$, is chosen a priori in both the classical models by selecting the initial photon per baryon ratio.

I have written the action in terms of conformal time because I ultimately intend to quantize in conformal time; i.e., I shall quantize by replacing the classical Hamiltonian implied by the Lagrangian in (2.18) or (2.25) by the "obvious" Hamiltonian operator and evolve the system via the Schrödinger equation

$$\mathrm{i}\, \partial\psi/\partial\tau = \hat{H}\psi \tag{2.28}$$

so that conformal time τ will be my time parameter. A few remarks need to be made about this quantization proposal. First of all, equation (2.28) is first order in time but second order in space. This is contrary to most quantization procedures for quantum gravity, which generally require time and space to enter the equation for the wave function on an equal footing. However, requiring space and time to appear on an equal footing either results in equations which are second order in time, which in turn means Universal wave functions with positive and negative frequencies, or equations which are first order in space, which in turn means dealing with Universal spinors rather than Universal scalar wave functions, or operators which are neither first nor second order in either space or time and have to be defined indirectly via the spectral theorem. In any of these cases both the mathematics and the wave function interpretation is very difficult. With a distinguished time coordinate, one can avoid these difficulties. Furthermore, general relativity, in contrast to special relativity, often picks out a preferred time coordinate. In virtually all physically realistic classical closed universes, the constant mean curvature foliation defines a unique preferred time coordinate [19]. In writing the Friedmann metric in the form (2.2), I have in fact chosen the constant mean curvature foliation by setting the shift vector to be zero. Thus we might expect quantum gravity to be similar to classical gravity in picking out a preferred time coordinate. In addition, equation (2.28) will be regarded in the preferred time quantum mechanics as a genuine dynamical equation. In many treatments of quantum gravity (e.g., the Appendix, [1] and [3]), the "evolution" equation is not an evolution equation at all but rather a constraint equation $\mathcal{H} = 0$, where $\mathcal{H}$ is the super-Hamiltonian, turned into a constraint on the wave function via the prescription $\hat{\mathcal{H}}\psi = 0$. There is no general principle of quantum mechanics that requires the constraints to be handled in this way. It is merely slavish devotion to the special relativistic (but not general relativistic!) principle that the space and time variables are to be treated in the same manner, and that all time coordinates are to be on an equal footing, combined with the fact that $\hat{\mathcal{H}} = 0$ gives no quantum mechanics at all [3]. And in the end, the symmetry between space and time is lost anyway in the $\hat{\mathcal{H}}\psi = 0$ theory because one must choose a preferred time coordinate in order to carry out the replacement $\mathcal{H} \to \hat{\mathcal{H}}$, and different preferred times give physically different quantum theories. However, for those who prefer the $\hat{\mathcal{H}}\psi = 0$ approach to quantum gravity, I shall show in the Appendix that in fact for the radiation-filled closed Friedmann universe, the equation $\hat{\mathcal{H}}\psi = 0$ is equivalent to (2.28), with $\hat{H}$ being the simple harmonic oscillator operator, provided the preferred time is conformal time.

Although the constant mean curvature foliation picks out a preferred time coordinate, it does not fix the scale of the time coordinate. The conformal time is but one scale of many; other choices are discussed in [17]. There are two reasons to chose the conformal time. First and foremost, such a choice simplifies the mathematics enormously. The Lagrangian (2.18) for the radiation gas is just the Lagrangian for a simple harmonic oscillator (SHO), and the more general Lagrangian (2.25) is not much more complicated. It would be difficult to find a simpler system to analyze in trying to interpret the wave function. Such a simple Lagrangian allows us to completely avoid the factor-ordering problems which bedevil other choices of the time variable [17].

There is a more physical reason for choosing the conformal time as the quantization time parameter. Whichever time variable is chosen, it will be an unmeasurable parameter in the quantum theory, because the time variable in Schrödinger's equation is a *parameter*, not an operator [20]. But according to the fundamental postulates of quantum mechanics, an observable – which is to say, anything that can be measured – must be represented by an operator [self-adjoint]. Hence, once a quantization time parameter is chosen, we must henceforth abandon all attempts to measure that variable. It is not a physical quantity; it just orders the states in time.

The conformal time is to be preferred over many other choices (such as proper time, or matter time [17]) because on the classical level, it is used in precisely the way the quantum mechanical time is used: to order the states. The conformal time is independent (at least in the pure dust or radiation cases) of the only physical constant in the system, the radius of maximum expansion. True physical time, for instance the proper time or the matter time, does depend on the maximum radius, since proper time (for example) is obtained via integration of eq. (2.7). From the point of view of the Many-Worlds Interpretation, a time parameter which treats all the classical worlds on an equal footing is required, so that all the classical worlds can be subsumed into a single wave function. This is possible to accomplish in the dust and radiation case, because δ is not quantized. The conformal time coordinate is perhaps not unique in having these characteristics, but it is the simplest choice which does.

The conformal time also has an advantage in closed universes over some intrinsic times, such as matter time (in which the time coordinate is set equal to the matter density), because it increases in both the expanding and contracting phase. The matter time, on the other hand, is double valued: in the closed Friedmann universe, the same matter density occurs twice – once in the expanding phase, and once in the contracting phase. This means that if matter time is used, the same time variable cannot be used in both phases, and it becomes very difficult to address global evolution questions in quantum cosmology.

With conformal time as the time variable, time unidirectionally increases in both the expanding and contracting phases of a closed universe. This unidirectionality of time may be lost if other time parameters are used. For example, time goes backward in the contracting phase of Hawking's quantum cosmology [1]. In order for the second law of thermodynamics to be globally valid, we must have temporal unidirectionality, and this is another argument for conformal time.

I am not the first to suggest quantizing in conformal time. J.V. Narlikar and his students have extensively developed a number of quantum cosmologies based on quantizing in conformal time; see [40] for a recent review of this work. I have two mathematical objections to their work. First of all, I do not think they have adequately considered the constraint equations in their model. They use the standard Friedmann metric with only the conformal degree of freedom, and put this into the standard Einstein action. However, as I discussed above, the variation of this metric will give only the dynamical

Einstein equations. The constraints must be either solved before quantization, or dealt with at the quantum level. A second mathematical objection is that they pay no attention to the domain of their operators. If, as I have argued, the domain must be restricted to $R > 0$, then the requirement of self-adjointness of the Hamiltonian places boundary conditions on the admissible wave functions at the origin. I see no evidence that they have imposed such boundary conditions.

But these mathematical objections are rather minor; the real objection I have to their work is the view of quantum mechanics underlying it. They regard the effect of quantum gravity as generating fluctuations around a *single* classical universe. The MWI suggests a radically different Universe: *all* classical universes are simultaneously present, and the appearance of a single classical universe arises from the measurement process. The MWI leads to a different view of the appropriate boundary conditions, and hence the MWI implies a *physically distinct* quantum cosmology.

The matter in my classical model is a perfect fluid. Since the matter tensor completely determines the geometry in the Friedmann universe, I shall, in effect, be quantizing the perfect fluid. Many workers in quantum cosmology worry about the legitimacy of quantizing such a "field", and so they decide to let the matter be a scalar field, for which the quantization procedures are well understood. However, the actual matter in the Universe is not now, and as far as we can tell, never has been, predominately a scalar field. This makes it difficult to apply the scalar field quantum cosmologies to the actual universe.

I personally believe directly quantizing a perfect fluid is justified on physical grounds. Perfect fluids have been directly quantized: Landau did so in order to study superfluids, and his quantum fluid model [31] has been found quite accurate in describing many (but not all) aspects of superfluidity. Landau quantized the fluid by replacing the classical density scalar and current vector with quantum operators which were subject to certain commutation relations. My quantization procedure will be somewhat analogous to his.

3. The Many-Worlds Interpretation

I shall begin the presentation of the Many-Worlds theory with a simple example which illustrates the salient features without any unnecessary mathematical complication. The simplest quantum mechanical system of interest is a system with two possible states. For definiteness let us think of these states as the two possible states of the vertical component of the spin of the electron, and they will be represented as $|\uparrow\rangle$ and $|\downarrow\rangle$, the former denoting spin up, and the latter spin down. Acorrding to the postulates of quantum mechanics, these two states form a basis for the state space of the electron-spin system and so any state $|\psi\rangle$ of this system can be written as a linear superposition

$$|\psi\rangle = a|\uparrow\rangle + b|\downarrow\rangle . \tag{3.1}$$

For reasons that will be apparent later, I shall not impose an a priori normalization condition on the constants a and b.

According to quantum theory, it is necessary to include some physics of the observer or measuring apparatus in the analysis if one wishes to talk about the result of a measurement on a system. A moment's reflection will show that the essential feature of a measuring apparatus is the ability to record the result of a measurement. The essence of a successful measurement is the transfer of information about the system being measured to the memory of the measuring apparatus, as Everett [2] and DeWitt [20] (see also [4] and [5]) were the first to show. Since in our simple system the spin of the electron can be spin up or spin down, we need an apparatus whose memory is sufficiently complex to record either possibility. We should also have an apparatus neutral state, corresponding to no measurement having been performed. The minimal apparatus required to measure the spin of the electron should have three memory states $|u\rangle$, $|d\rangle$, $|n\rangle$, which represent the memory recording the electron spin to be up, the electron spin to be down, and no recording yet having been made respectively. (The neutral state is not strictly speaking necessary, but it simplifies the analysis enormously. See Deutsch's important paper [49] for a discussion of this point.) The laws of quantum mechanics must be regarded as universal if we are to apply them to cosmology, so they must apply with equal force to the measuring apparatus as to the system being measured. Thus a general state $|\Phi\rangle$ of the measuring apparatus must also be a linear superposition of the three basis states. The general state $|\text{Cosmos}\rangle$ of the universe, which is defined to be everything – all systems and apparata – considered in the analysis, is a sum of tensor products of the basis states of the system with the basis states of the apparatus:

$$|\text{Cosmos}\rangle = \sum_i a_i\, |\psi_i\rangle\, |\Phi_i\rangle\,. \tag{3.2}$$

It is of course not a priori obvious that the universe as defined above can be sufficiently divided into systems and apparata to permit writing $|\text{Cosmos}\rangle$ as a sum of tensor products. However, such a division must be possible if the term "measurement" is to be meaningful, so we must assume the universe is sufficiently inhomogeneous to allow such a division. This assumption is called The Postulate of Complexity by DeWitt [21, 22]. It is an unobjectionable assumption in standard quantum mechanics, but it must be justified in cosmology, where the system is the Universe itself. The measurement corresponds to a change of state of the universe. According to laws of quantum mechanics, all changes of state are accomplished by linear unitary operators acting on the state. In our example, the appropriate state is $|\text{Cosmos}\rangle$, so the measurement must be represented as

$$M|\text{Cosmos(before)}\rangle = |\text{Cosmos(after)}\rangle \tag{3.3}$$

where M is the linear unitary operator, and $|\text{Cosmos(before)}\rangle$, $|\text{Cosmos(after)}\rangle$ are the states of the universe before and after the measurement is performed, respectively. It is very important to note we have eliminated all possibility of wave function reduction by our assumption that all apparata and systems are equally governed by the same quantum mechanical laws. Again, this assumption is essential if standard quantum mechanics is to be applied to the Universe as a whole. Since M is a linear operator, its action on $|\text{Cosmos}\rangle$ can be completely determined by its action on the tensor products of the system and apparatus basis states. For simplicity, we shall choose M to represent what DeWitt calls a "von Neumann measurement" [23], which is a measurement that has no effect on the system if the system is in an eigenstate of the observable measured by the particular apparatus used. In the electron spin example, if the apparatus is set to measure spin up or down, and the electron spin happens to be either up or down, then a von Neumann measurement is performed if the apparatus records spin up or down, respectively, and further the state of the electron is not changed by the measurement interaction. This measurement can be represented formally as

$$M|\uparrow\rangle|n\rangle = |\uparrow\rangle|u\rangle \tag{3.4a}$$

$$M|\downarrow\rangle|n\rangle = |\downarrow\rangle|d\rangle\,. \tag{3.4b}$$

Thus if the system is in an eigenstate of the system variable to be measured by the apparatus a von Neumann measurement does not disturb the system. The existence of such measurement operators surprises many physicists. Ever since Heisenberg used his gamma-ray microscope thought-experiment to demonstrate the Uncertainty Principle for the position and momentum of an electron, many have believed that a measurement on a system necessarily disturbs the system, and this disturbance is cause of the Uncertainty Principle. This is not true. The operator defined in (3.4) does not disturb the system (provided the system happens to be in an eigenstate of the component of spin measured by the apparatus). For virtually any variable, measurement operators can be defined which have the effect of recording the state of the system in the memory of the measuring apparatus without disturbing the system. As DeWitt has stated it "... if suitable devices are used, ... the apparatus can record what the value of the system observable would have been without the coupling [24]". In our simple two-spin-state electron example, the Stern–Gerlach apparatus can be regarded as a physical realization of such a von Neumann measuring apparatus, provided the vertical component of momentum of the atom is considered to be the memory trace of the apparatus, and spin precession is ignored. (See ref. [20] pp. 201–202 for a fuller discussion.)

The effect of a von Neumann measurement operator M acting on any state (3.2) where we set $|\mathrm{Cosmos(before)}\rangle = |\psi\rangle\,|\Phi\rangle$ for simplicity, and with $|\psi\rangle$ given by (3.1) and $|\Phi\rangle = |\mathrm{n}\rangle$ is then

$$\begin{aligned} M|\mathrm{Cosmos(before)}\rangle &= M(a|\uparrow\rangle + b|\downarrow\rangle)|\mathrm{n}\rangle \\ &= M(a|\uparrow\rangle|\mathrm{n}\rangle) + M(b|\downarrow\rangle|\mathrm{n}\rangle) \\ &= a|\uparrow\rangle|\mathrm{u}\rangle + b|\downarrow\rangle|\mathrm{d}\rangle \\ &= |\mathrm{Cosmos(after)}\rangle\,. \end{aligned} \tag{3.5}$$

We can assume that $\{|\uparrow\rangle|\mathrm{n}\rangle, |\downarrow\rangle|\mathrm{n}\rangle\}$ span the initial state space, for we shall assume the apparatus is always initially in the neutral position.

The fundamental problem in the quantum theory of measurement is deciding what the linear superposition of universe states in the third line of eq. (3.5) means. The advocates of the Many-Worlds Interpretation decide this question by arguing as follows. It is obvious that each element in the two cases (3.4) corresponds to a real physical state of some actual entity either associated with the system or the apparatus. If we grant that the state (3.1) also corresponds to an actual physical state – and we can justify this by reference either to innumerable experiments or to the superposition principle of quantum mechanics – and we grant that quantum evolution of everything in existence occurs via linear operators, then we are led necessarily to the conclusion that each term in (3.5) corresponds to an actual physical state. We are forced to say that the universe "splits" into two "worlds". In the first world, represented by the first term in (3.5), the electron has spin up, and its spin is measured to be spin up. In the second world, represented by the second term in (3.5), the electron has spin down, and its spin is measured to be spin down. Another way to express this is to say that all a quantum measurement does, or indeed can do, is establish a unique correlation between states of the system being measured and states of the measuring apparatus. In the above discussion, I qualified the statement that the operator (3.4) did not disturb the system with the proviso that the system be in a certain eigenstate. If the system is not in an eigenstate – as it is not in (3.5) – then the operator does effect the system. What the operator (3.4) does when the system is in a general state is establish correlations between the apparatus basis states and those system basis states which are selected by the choice of apparatus basis states. The existence of these correlations can be detected if the {system} + {apparatus} is measured by a second apparatus. For example, a short calculation would show that a measurement of the system by an apparatus with basis states corresponding to a measurement of spin in the horizontal rather than the vertical direction would give a different result if the system is measured by the second apparatus before the system interacts with the first apparatus, than the result which the second apparatus would obtain were it to measure the system after the system has been measured by the first apparatus. Needless to say, the practical importance of these correlations will depend on the size of the system and the measuring apparatus relative to Planck's constant, and in the situation where the system and the apparatus are both macroscopic objects (which is the case we are interested in here), the correlations can be effectively ignored.

There is a misconception in popular accounts about the MWI which must be cleared up before the MWI can be applied to cosmology. The misconception arises because the word "universe" is used in one sense in technical discussions about the MWI, and in another sense in non-technical discussions. We have said in our interpretation of (3.5), which is the state of the universe after the measurement, that the universe is split by the measurement. This is the standard terminology in the technical literature, but it is important to note this split is to be associated more with the measuring apparatus rather than with the system being measured. In the case of a von Neumann measurement, the system is not effected (again, with the exception of the correlations) by the measurement, so it is completely misleading to the system as splitting as a result of the measurement. On the other hand, as is obvious from (3.5), the measuring apparatus undergoes a tremendous change: it goes from $|\mathrm{n}\rangle$ to either $|\mathrm{u}\rangle$ or $|\mathrm{d}\rangle$ (or both). Of course, in measurements which are not of the von Neumann type, the system variables and not just the system/apparatus correlations will be changed by the measurement, but for macroscopic systems the change of the system variables are very small; measurements of such systems can be regarded as essentially von Neumann measurements. In particular, a measurement of the radius of the Universe can be considered a von Neumann measurement, and it would thus be more appropriate to regard the recording apparatus as splitting rather than the Universe, although the "universe" in the technical sense defined above does split. The "universe" in the technical sense includes just the system and the measuring apparatus, whereas the Universe in the non-technical sense includes these two entities, plus everything else in existence. I have made a distinction between the two uses of the word "universe" by capitalizing the word when it refers to the totality of everything in existence, and left it uncapitalized when it refers to just the system and the apparatus: i.e., to everything being considered in the analysis of the measurement. The other things in the Universe, those things which are not considered in the analysis of the measurement – the planets, stars and galaxies – are coupled only very weakly to the measuring apparatus. Thus these other items do not split when the apparatus does. Looking at the split from this point of view obviates one of the major objections to the MWI, which is that the MWI seems to require if not an actual infinity, then at least a large number of "Universes" (in the popular sense) to explain a measurement of some microscopic phenomena, and this is contrary to Occam's Razor. In the explanation of the MWI given above, there is only one Universe, but small parts of it – measuring apparata – split into several pieces. They split – or more precisely, they undergo a drastic change – upon the act of measurement because they are designed to do so. If they were not capable of registering changes on a macroscopic level they would be quite useless as measuring devices. This fact plus the linearity of quantum mechanical operators requires them to split.

Everett himself realized that it is more appropriate to think of the measuring apparatus rather than the Universe as splitting. In reply to a criticism by Einstein against quantum mechanics, to the effect that he (Einstein) "... could not believe ... a mouse could bring about drastic changes in the Universe simply by looking at it," Everett said, "... it is not so much the system which is affected by an observation as the observer ... ([20] p. 116).... The mouse does not affect the Universe – only the mouse is affected" ([20] p. 117).

We can see this formally by simply putting the non-interacting remainder of the universe in eq. (3.5):

$$\begin{aligned} M|\mathrm{Universe(before)}\rangle &= M(a|\uparrow\rangle + b|\downarrow\rangle)|\mathrm{n}\rangle\,|\mathrm{everything\ else}\rangle \\ &= a|\uparrow\rangle|\mathrm{u}\rangle\,|\mathrm{everything\ else}\rangle + b|\downarrow\rangle|\mathrm{d}\rangle\,|\mathrm{everything\ else}\rangle \\ &= (a|\uparrow\rangle|\mathrm{u}\rangle + b|\downarrow\rangle|\mathrm{d}\rangle)\,|\mathrm{everything\ else}\rangle\,. \end{aligned} \tag{3.6}$$

It is clear from (3.6) that "everything else" does not split.

Bryce DeWitt has occasionally described the Many-Worlds Interpretation in terms which suggest he disagrees with Everett's assessment quoted above. For instance, DeWitt has claimed: "The Universe is constantly splitting into a stupendous number of branches, all resulting from the measurement-like interactions between its myriads of components. Moreover, every quantum transition taking place on every star, in every galaxy, in every remote corner of the Universe is splitting our local world on Earth in myriads of copies of itself.... I still recall vividly the shock I experienced on first encountering this multiworld concept.

"The idea of 10^{100+} slightly imperfect copies of oneself all constantly splitting into further copies, which ultimately becomes unrecognizable, is not easy to reconcile with common sense. Here is schizophrenia with a vengeance ([20] p. 161)."

I fear I must stand with Everett against DeWitt; I believe Everett's word picture is closer to what the mathematics is telling us than the image invoked by DeWitt's phrasing. A human being can properly be said to split only if he undergoes an interaction with rest of the Universe which causes a change of state in his memory, in which different states of his memory are correlated after the interaction with different states of that portion of the rest of the Universe with which he interacted. The overwhelming majority of quantum transitions taking place in distant stars will not induce a change of state analogous to (3.4), with $|\Phi\rangle$ representing the human memory, even if the word "memory" is expanded to include any part of the human body that can record the effect of interacting with the external world. Thus these transitions will not induce a split in the human being. With respect to these distant quantum transitions, the human being is analogous to $|\text{everything else}\rangle$ in (3.6). The human being is not split by these transitions.

A human being, or indeed any measuring apparatus, would be unaware of, or in the case of an inanimate apparatus, could not detect, those splits which he/she/it does undergo. To detect the split would entail introducing a second observing apparatus into the universe which is capable of recording in its memory both worlds $|\mathrm{u}\rangle$ and $|\mathrm{d}\rangle$ of the split first apparatus. In the case of a human being, the two apparata could in principle be two sections of the human memory, the second of which observes the first. It is impossible to construct such a second apparatus if it is reasonably required that this second apparatus definitely records the first apparatus to be in the state $|\mathrm{u}\rangle$ if in fact it is, or in the state $|\mathrm{d}\rangle$ if in fact it is. We may as well let the second apparatus perform a von Neumann measurement on the system simultaneously with measuring the first apparatus, as a check. We require only that the second apparatus record the system as being in the state $|\uparrow\rangle$ if in fact it is in this state, and as being in the state $|\downarrow\rangle$ if in fact it is in this state. The state of the second apparatus, $|\mathrm{A}_2\rangle$, can thus be expanded in terms of basis states of the form $|a_1, a_2\rangle$, where a_1 records the value of the system variable and a_2 records the content of the first apparatus' memory. Both a_1 and a_2 can have the values n, u or d. Before the interaction between the second apparatus and the rest of the universe, we shall require the second apparatus to be in the state $|\mathrm{n}, \mathrm{n}\rangle$.

The above restrictions on what the second apparatus must record uniquely defines the second apparatus interaction operator M_2 acting on the basis states of the universe. We have

$$M_2|\uparrow\rangle|\mathrm{u}\rangle|\mathrm{n}, \mathrm{n}\rangle = |\uparrow\rangle|\mathrm{u}\rangle|\mathrm{u}, \mathrm{u}\rangle \tag{3.7a}$$

$$M_2|\downarrow\rangle|\mathrm{d}\rangle|\mathrm{n}, \mathrm{n}\rangle = |\downarrow\rangle|\mathrm{d}\rangle|\mathrm{d}, \mathrm{d}\rangle \tag{3.7b}$$

$$M_2|\uparrow\rangle|\mathrm{n}\rangle|\mathrm{n}, \mathrm{n}\rangle = |\uparrow\rangle|\mathrm{n}\rangle|\mathrm{u}, \mathrm{n}\rangle \tag{3.7c}$$

$$M_2|\downarrow\rangle|\mathrm{n}\rangle|\mathrm{n}, \mathrm{n}\rangle = |\downarrow\rangle|\mathrm{n}\rangle|\mathrm{d}, \mathrm{n}\rangle\,. \tag{3.7d}$$

The last two entries in (3.7) are effective only if we were to interact the second apparatus with the rest of the universe before the first apparatus has measured the state of the system. Before any measurements by any apparatus are performed, the state of the universe is

$$|\text{Cosmos(before)}\rangle = |\psi\rangle|\text{n}\rangle|\text{n, n}\rangle\,. \tag{3.8}$$

A measurement of the state of the system by the first apparatus, followed by measurements of the state of the system and the state of the first apparatus is thus represented as:

$$M_2M_1|\text{Cosmos(before)}\rangle = M_2M_1(a|\uparrow\rangle + b|\downarrow\rangle)|\text{n}\rangle|\text{n, n}\rangle$$

$$= M_2(a|\uparrow\rangle|\text{u}\rangle|\text{n, n}\rangle + b|\downarrow\rangle|\text{d}\rangle|\text{n, n}\rangle) \tag{3.9a}$$

$$= a|\uparrow\rangle|\text{u}\rangle|\text{u, u}\rangle + b|\downarrow\rangle|\text{d}\rangle|\text{d, d}\rangle\,. \tag{3.9b}$$

It is clear from (3.9) that the first apparatus is the apparatus responsible for the splitting of the universe. More precisely, it is the first apparatus that is responsible for splitting itself and the second apparatus. The second apparatus splits, but the split just follows the original split of the first apparatus, as is apparent in (3.9b). As a consequence, the second apparatus does not detect the splitting of the first apparatus. Again, the impossibility of split detection is a consequence of two assumptions: first, the linearity of the quantum operators M_2 and M_1; second, the requirement that M_2 measure the appropriate basis states of the system and the apparatus correctly. The second requirement is formalized by (3.7). Again, in words this requirement says that if the system and first apparatus are in eigenstates, then the second apparatus had better record this fact correctly.

Our ultimate goal is to develop a formalism which will tell us what we will actually observe when we measure an observable of a system while the system state is changing with time. One lesson from the above analysis on quantum mechanics from the Many-Worlds point of view is that to measure anything it is necessary to set up an apparatus which will record the result of that measurement. To have the possibility of observing a change of some observable with time requires an apparatus which can record the results of measuring that observable at sequential times. To make n sequential measurements requires an apparatus with n sequential memory slots in its state representation. At first we will just consider the simple system (3.1) that we have analyzed before, so the time evolution measurement apparatus has the state $|E\rangle$, which can be written as a linear superposition of basis states of the form

$$|a_1, a_2, \ldots, a_n\rangle \tag{3.10}$$

where each entry a_j can have the value n, u or d, as before. The jth measurement of the system state is represented by the operator M_j, defined by

$$M_j|\uparrow\rangle|a_1, a_2, \ldots, a_j, \ldots, a_n\rangle = |\uparrow\rangle|a_1, a_2, \ldots, \text{u}, \ldots, a_n\rangle \tag{3.11a}$$

$$M_j|\downarrow\rangle|a_1, a_2, \ldots, a_j, \ldots, a_n\rangle = |\downarrow\rangle|a_1, a_2, \ldots, \text{d}, \ldots, a_n\rangle\,. \tag{3.11b}$$

As before, the initial state of the apparatus will be assumed to be $|\text{n, n}, \ldots, \text{n}\rangle$. The measurement is a von Neumann measurement.

Time evolution will be generated by a time evolution operator $T(t)$. It is a crucial assumption that

$T(t)$ act only on the system, and not have any effect on the apparatus that will measure the time evolution. In other words, we shall assume the basis states (3.10) are not effected by the operator $T(t)$. This is a standard and indeed an essential requirement imposed on instruments that measure changes in time. If the record of the values of some observable changed on time scales comparable with the rate of change of the observable, it would be impossible to disentangle the change of the observable from the change of the record of the change. When we measure the motion of a planet, we record its positions from day to day, assuming (with justification!) that our records of its position at various times are not changing. If we write the apparatus state as $|\Phi\rangle$, the effect of a general time evolution operator $T(t)$ on the basis states of the system can be written as

$$T(t)|\uparrow\rangle|\Phi\rangle = (a_{11}(t)|\uparrow\rangle + a_{12}(t)|\downarrow\rangle)|\Phi\rangle \qquad (3.12a)$$

$$T(t)|\downarrow\rangle|\Phi\rangle = (a_{21}(t)|\uparrow\rangle + a_{22}(t)|\downarrow\rangle)|\Phi\rangle\,. \qquad (3.12b)$$

Unitarity of $T(t)$ imposes some restrictions on the a_{ij}'s, but we won't have to worry about these. Interpreting the result of a measurement on the system in an initially arbitrary state after an arbitrary amount of time has passed would require knowing how to interpret the a_{ij}'s, and as yet we have not outlined the meaning of these in the MWI. So let us for the moment analyze a very simplified type of time evolution. Suppose that we measure the state of the system every unit amount of time; that is, at $t = 1, 2, 3, \ldots$, etc. Since time operators satisfy $T(t)\,T(t') = T(t+t')$, the evolution of the system from $t = 0$ to $t = n$ is given by $[T(1)]^n$. Again for simplicity, we shall assume $a_{11}(1) = a_{22}(1) = 0$, $a_{12}(1) = a_{21}(1) = 1$. This choice will give a unitary $T(t)$. We have

$$T(1)\,|\uparrow\rangle\,|\Phi\rangle = |\downarrow\rangle\,|\Phi\rangle \qquad (3.13a)$$

$$T(1)\,|\downarrow\rangle\,|\Phi\rangle = |\uparrow\rangle\,|\Phi\rangle\,. \qquad (3.13b)$$

All that happens is that if the electron spin happens to be in an eigenstate, that spin is flipped from one unit of time to the next, with $[T(1)]^{2n} = I$, the identity operator.

After every unit of time we shall measure the state of the system. The time evolution and measurement processes together will be represented by a multiplicative sequence of operators acting on the universe as follows:

$$M_n\,T(1)\,M_{n-1}\,T(1)\cdots M_2\,T(1)\,M_1\,|\psi\rangle\,|\mathrm{n, n, \ldots, n}\rangle \qquad (3.14a)$$

$$= M_n\,T(1)\,M_{n-1}\,T(1)\cdots M_2\,T(1)\,[M_1(a|\uparrow\rangle + b|\downarrow\rangle)]|\mathrm{n, n, \ldots, n}\rangle \qquad (3.14b)$$

$$= M_n\,T(1)\,M_{n-1}\,T(1)\cdots M_2\,T(1)\,(a|\uparrow\rangle|\mathrm{u, n, \ldots, n}\rangle + b|\downarrow\rangle|\mathrm{d, n, \ldots, n}\rangle) \qquad (3.14c)$$

$$= M_n\,T(1)\,M_{n-1}\,T(1)\cdots M_2\,(a|\downarrow\rangle\,|\mathrm{u, n, \ldots, n}\rangle + b|\uparrow\rangle\,|\mathrm{d, n, \ldots, n}\rangle) \qquad (3.14d)$$

$$= M_n\,T(1)\,M_{n-1}\,T(1)\cdots M_3\,T(1)\,(a|\downarrow\rangle\,|\mathrm{u, d, n}\ldots\rangle + b|\uparrow\rangle\,|\mathrm{d, u, n,}\ldots\rangle) \qquad (3.14e)$$

and so on.

The particularly interesting steps in the above algebra are (3.14c) and (3.14e). The first measurement of the state of the system splits the universe (or more precisely, the apparatus) into two worlds. In each world, the evolution proceeds as if the other world did not exist. The first measurement, M_1, splits the apparatus into the world in which the spin is initially up and the world in which the spin is initially down. Thereafter each world evolves as if the spin of the entire system were initially up or down respectively.

If we were to choose $a = b$, then $T(1)|\psi\rangle|\Phi\rangle = |\psi\rangle|\Phi\rangle$; i.e., the state of the system in the absence of a measurement would not change with time. It would be a stationary state. If the system were macroscopic – for instance, if it were the Universe – then even after the measurement the Universe would be almost stationary; the very small change in the state of a macroscopic system can be ignored. Nevertheless, the worlds would change with time. An observer who was capable of distinguishing the basis states would see a considerable amount of time evolution even though the actual, total state of the macroscopic system is essentially stationary. Whether or not time evolution will be observed depends more on the details of the interaction between the system and the observer trying to see if the change occurs in the system, rather than on what changes are actually occurring in the system.

In order to interpret the constants a, b in (3.1), or the a_{ij}'s in (3.12), it has been shown by Hartle [28], Finkelstein [29] and Graham [30] (see also DeWitt's articles in [20]) that it is necessary to use an apparatus which makes repeated measurements on not merely a single state of a system, but rather on an ensemble of identical systems. The initial ensemble state has the form:

$$|\text{Cosmos(before)}\rangle = (|\psi\rangle)^m |\text{n}, \text{n}, \text{n}, \ldots, \text{n}\rangle \tag{3.15}$$

where there are m slots in the apparatus memory state $|\text{n}, \text{n}, \ldots, \text{n}\rangle$. The kth slot records the measured state of the kth system in $(|\psi\rangle)^m$. The kth slot is changed by the measuring apparatus operator M_k, which acts as follows on the basis states of the kth $|\psi\rangle$:

$$M_k|\psi\rangle \cdots |\psi\rangle|\text{u}\rangle|\psi\rangle \cdots |\psi\rangle|\text{n}, \text{n}, \ldots, \text{n}\rangle = |\psi\rangle \cdots |\psi\rangle|\text{u}\rangle|\psi\rangle \cdots |\psi\rangle|\text{n}, \ldots, \text{n}, \text{u}, \text{n}, \ldots, \text{n}\rangle \tag{3.16a}$$

$$M_k|\psi\rangle \cdots |\psi\rangle|\text{d}\rangle|\psi\rangle \cdots |\psi\rangle|\text{n}, \text{n}, \ldots, \text{n}\rangle = |\psi\rangle \cdots |\psi\rangle|\text{d}\rangle|\psi\rangle \cdots |\psi\rangle|\text{n}, \text{n}, \ldots, \text{n}, \text{d}, \text{n}, \ldots, \text{n}\rangle\,. \tag{3.16b}$$

The M_k operator effects only the kth slot of the apparatus memory. It has no other effect on either the system ensemble or the other memory slots.

If we perform m state measurements on the ensemble $(|\psi\rangle)^m$, an operation which would be carried out by the operator $M_m M_{m-1} \cdots M_2 M_1$. The result is

$$\begin{aligned}
&M_m M_{m-1} \cdots M_2[M_1(a|\uparrow\rangle + b|\downarrow\rangle)](|\psi\rangle)^{m-1}|\text{n}, \text{n}, \ldots, \text{n}\rangle \\
&\quad = M_m M_{m-1} \cdots M_3 M_2(|\psi\rangle)^{m-1}(a|\uparrow\rangle|\text{u}, \text{n}, \ldots, \text{n}\rangle + b|\downarrow\rangle|\text{d}, \text{n}, \ldots, \text{n}\rangle) \\
&\quad = M_m M_{m-1} \cdots M_3(|\psi\rangle)^{m-2}(aM_2|\psi\rangle|\uparrow\rangle|\text{u}, \text{n}, \ldots, \text{n}\rangle + bM_2|\psi\rangle|\downarrow\rangle|\text{d}, \text{n}, \ldots, \text{n}\rangle) \\
&\quad = M_m \cdots M_4(|\psi\rangle)^{m-3}(M_3|\psi\rangle)(a^2|\uparrow\rangle|\uparrow\rangle|\text{u}, \text{u}, \text{n}, \ldots, \text{n}\rangle \\
&\qquad + ab|\uparrow\rangle|\downarrow\rangle|\text{u}, \text{d}, \text{n}, \ldots, \text{n}\rangle + ba|\downarrow\rangle|\uparrow\rangle|\text{d}, \text{u}, \text{n}, \ldots, \text{n}\rangle + b^2|\downarrow\rangle|\downarrow\rangle|\text{d}, \text{d}, \text{n}, \ldots, \text{n}\rangle) \\
&\quad = \sum a^j b^{m-j}(|\uparrow\rangle)^j(|\downarrow\rangle)^{m-j}|s_1, s_2, \ldots, s_m\rangle
\end{aligned}$$

where the s_i's represent either u or d, and the sum is over all possible permutations of u's and d's in the memory basis state $|s_1, s_2, \ldots, s_m\rangle$. All possible sequences of u's and d's are represented in the sum. The measurement operator $M_m \cdots M_1$ splits the apparatus into 2^m worlds. In this situation we have m

systems rather than one, so each measurement splits the apparatus (or equivalently, the universe). Each measurement splits each previous world in two.

In each world, we now calculate the relative frequency of the u's and d's. Hartle [28], Finkelstein [29] and Graham [30] have independently shown that if a, b are defined by $a = \langle\psi|\uparrow\rangle$ and $b = \langle\psi|\downarrow\rangle$, then as m approaches infinity, the relative frequency of the u's approaches $|a|^2/(|a|^2+|b|^2)$, and the relative frequency of the d's approaches $|b|^2/(|a|^2+|b|^2)$ in the Hilbert space for which the scalar product defines $\langle\psi|\uparrow\rangle$ and $\langle\psi|\downarrow\rangle$, except for a set of worlds of measure zero in the Hilbert space. It is only at this stage, where a and b are to be interpreted, that it is necessary to assume $|\psi\rangle$ is a vector in a Hilbert space. For the discussion of universe splitting, it is sufficient to regard $|\psi\rangle$ as a vector in a linear space with $|\psi\rangle$ and $c|\psi\rangle$, for any complex constant c, being physically equivalent. If we impose the normalization condition $|a|^2+|b|^2=1$, then $|a|^2$ and $|b|^2$ will be the usual probabilities of measuring the state $|\psi\rangle$ in the state $|\uparrow\rangle$ or $|\downarrow\rangle$, respectively. It is not essential to impose the normalization condition even to interpret a and b. For example, $|a|^2/(|a|^2+|b|^2)$ would represent the relative probability of the subspace $\langle\psi|\uparrow\rangle$ as opposed to $\langle\psi|\downarrow\rangle$ even if we expanded $|\psi\rangle$ to include other states, enough to make $|\psi\rangle$ itself non-normalizable. This point is important because it allows us to consider non-normalizable Universal wave functions. In the quantization procedure I shall develop in the next section, the ability to handle non-normalizable wave functions will be absolutely essential, for the universal wave function generated by my quantization method will in fact be non-normalizable.

One key point can be made now: since there is only *one* Universe represented by only *one* unique wave function $|\Psi\rangle$, the ensemble necessary to measure $|\langle a|\Psi\rangle|^2$ cannot exist for *any* state $|a\rangle$. Thus, being unmeasurable, the quantities $|\langle a|\Psi\rangle|^2$ have no direct physical meaning. We can at best *assume* $|a|^2/(|a|^2+|b|^2)$ measures relative probability. But there is still absolutely no reason to assume that $|\Psi\rangle$ is normalizable.

We will now consider wave packet spreading from the Many-Worlds point of view. A simple system which will show the essential features has four degrees of freedom, labeled by the basis states $|\uparrow\rangle$, $|\downarrow\rangle$, $|\rightarrow\rangle$ and $|\leftarrow\rangle$. As before, we shall need a measuring apparatus to record the state of the system if we are to say anything about the state of the system. Since we are interested in measuring time evolution, say at m separate times (which will be assumed to be multiples of unit time, as before), we shall need an apparatus state with m slots: $|\mathrm{n}, \mathrm{n}, \ldots, \mathrm{n}\rangle$, where the n denotes the initial "no record" recording. The kth measurement of the system state will be carried out by the operator M_k, which changes the kth slot from n to u, d, r or l, depending on whether the state of the system is $|\uparrow\rangle$, $|\downarrow\rangle$, $|\rightarrow\rangle$ or $|\leftarrow\rangle$, respectively. The time evolution operator $T(t)$ will not effect the apparatus state, and its effect on the system basis states is as follows:

$$T(1)|\uparrow\rangle = a_{\uparrow\rightarrow}|\rightarrow\rangle + a_{\uparrow\downarrow}|\downarrow\rangle \tag{3.17a}$$

$$T(1)|\downarrow\rangle = a_{\downarrow\leftarrow}|\leftarrow\rangle + a_{\downarrow\uparrow}|\uparrow\rangle \tag{3.17b}$$

$$T(1)|\leftarrow\rangle = a_{\leftarrow\uparrow}|\uparrow\rangle + a_{\leftarrow\rightarrow}|\rightarrow\rangle \tag{3.17c}$$

$$T(1)|\rightarrow\rangle = a_{\rightarrow\downarrow}|\downarrow\rangle + a_{\rightarrow\leftarrow}|\leftarrow\rangle\,. \tag{3.17d}$$

The effect of the time evolution operator is easily visualized by regarding the arrow which labels the four basis states of the system as a hand of a clock. If the hand is initially at 12 o'clock (basis state $|\uparrow\rangle$), the operator $T(1)$ carries the hand clockwise to 3 o'clock (basis state $|\rightarrow\rangle$), and to 6 o'clock (basis state

$|\downarrow\rangle$). More generally for any basis state, the operator $T(1)$ carries the basis state (thought of a clock hand at 12, 1, 6 or 9 o'clock) clockwise one quarter and one half the way around the clock. We shall imagine that

$$|a_{ij}|^2 \gg |a_{ik}|^2 \tag{3.18}$$

if $j = i+1$, and $k = i+2$, where $i+n$ means carrying the arrow clockwise around n quarters from the ith clock hand position. The condition (3.18) means roughly that "most" of the wave packet initially at one definite clock position is carried to the immediately adjacent position in the clockwise direction, with a small amount of spreading into the position halfway around the clock. In addition to satisfying (3.18), the constants a_{ij} must be chosen to preserve the unitarity of $T(t)$, and let us suppose they will have the property $T(t)|\psi\rangle = |\psi\rangle$, where $|\psi\rangle = |\uparrow\rangle + |\rightarrow\rangle + |\downarrow\rangle + |\leftarrow\rangle$. The measured time evolution of the state $|\uparrow\rangle$ through three time units is then

$$M_4T_3M_3T_2M_2T_1M_1|\uparrow\rangle|\mathrm{n,n,n,n}\rangle = M_4T_3M_3T_2M_2T_1|\uparrow\rangle|\mathrm{u,n,n,n}\rangle \tag{3.19a}$$

$$= M_4T_3M_3T_2M_2(a_{\uparrow\rightarrow}|\rightarrow\rangle + a_{\uparrow\downarrow}|\downarrow\rangle)|\mathrm{u,n,n,n}\rangle \tag{3.19b}$$

$$= M_4T_3M_3T_2(a_{\uparrow\rightarrow}|\rightarrow\rangle|\mathrm{u,r,n,n}\rangle + a_{\uparrow\downarrow}|\downarrow\rangle|\mathrm{u,d,n,n}\rangle) \tag{3.19c}$$

$$= M_4T_3M_3[a_{\uparrow\rightarrow}(a_{\rightarrow\downarrow}|\downarrow\rangle + a_{\rightarrow\leftarrow}|\leftarrow\rangle)|\mathrm{u,r,n,n}\rangle + a_{\uparrow\downarrow}(a_{\downarrow\leftarrow}|\leftarrow\rangle + a_{\downarrow\uparrow}|\uparrow\rangle)|\mathrm{u,d,n,n}\rangle] \tag{3.19d}$$

$$\begin{aligned} &= M_4T_3[a_{\uparrow\rightarrow}a_{\rightarrow\downarrow}|\downarrow\rangle|\mathrm{u,r,d,n}\rangle + a_{\uparrow\rightarrow}a_{\rightarrow\leftarrow}|\leftarrow\rangle|\mathrm{u,r,l,n}\rangle \\ &\quad + a_{\uparrow\downarrow}a_{\downarrow\leftarrow}|\leftarrow\rangle|\mathrm{u,d,l,n}\rangle + a_{\uparrow\downarrow}a_{\downarrow\uparrow}|\uparrow\rangle|\mathrm{u,d,u,n}\rangle] \end{aligned} \tag{3.19e}$$

$$\begin{aligned} &= M_4[a_{\uparrow\rightarrow}a_{\rightarrow\downarrow}(a_{\downarrow\leftarrow}|\leftarrow\rangle + a_{\downarrow\uparrow}|\uparrow\rangle)|\mathrm{u,r,d,n}\rangle + a_{\uparrow\rightarrow}a_{\rightarrow\leftarrow}(a_{\leftarrow\uparrow}|\uparrow\rangle + a_{\leftarrow\rightarrow}|\rightarrow\rangle)|\mathrm{u,r,l,n}\rangle \\ &\quad + a_{\uparrow\downarrow}a_{\downarrow\leftarrow}(a_{\leftarrow\uparrow}|\uparrow\rangle + a_{\leftarrow\rightarrow}|\rightarrow\rangle)|\mathrm{u,d,l,n}\rangle + a_{\uparrow\downarrow}a_{\downarrow\uparrow}(a_{\uparrow\rightarrow}|\rightarrow\rangle + a_{\uparrow\downarrow}|\downarrow\rangle)|\mathrm{u,d,u,n}\rangle] \end{aligned} \tag{3.19f}$$

$$= a_{\uparrow\rightarrow}a_{\rightarrow\downarrow}a_{\downarrow\leftarrow}|\leftarrow\rangle|\mathrm{u,r,d,l}\rangle + a_{\uparrow\rightarrow}a_{\rightarrow\downarrow}a_{\downarrow\uparrow}|\uparrow\rangle|\mathrm{u,r,d,u}\rangle + \text{six other terms}\,. \tag{3.19g}$$

Thus each measurement splits each previous world into two; the number of branches of the universe doubles upon each measurement. The time evolution however, does not split worlds, for only measurements can do that. Each world is defined by a definite sequence of measured system basis states in the apparatus memory. (This is another indication that it is more appropriate to regard the apparatus as splitting rather than the system.) Every possible sequence of records allowed by the time evolution operator is represented in the universe after each of the three measurements. However, because of condition (3.18) and the probability interpretation of the constants a_{ij}, some of the worlds are much more probable than others. The first world in the list in (3.19g), the world $|\leftarrow\rangle|\mathrm{u,r,d,l}\rangle$ is the most probable world to be in at the end of three time periods and four measurements, since the coefficient of this world has the largest relative modulus. When condition (3.18) is imposed, the time evolution operator is most likely to carry the system state into the clockwise adjacent state, and indeed this is what is recorded in the memory sequence of the most probable final state of the universe. We might regard this sequence as the "classical" evolutionary sequence, because it is both the sequence of the peak of the wave packet initially in the state $|\uparrow\rangle$, and, as a consequence, the most probable final state. It is possible, of course, to have a memory sequence corresponding to a "non-classical" world: one in which the observed motion is not from i to $i+1$. The most probable of the "non-classical" worlds are those which have only one memory slot entry out of the classical sequence, so if one did not observe a purely "classical" evolution, the most likely one to see is one of the ones which is as close to "classical" as possible.

For all worlds – memory sequences – there is no overlap between the worlds, even though by the second time period the wave packets of the system have begun to overlap one another. This is a general property which is a consequence only of the linearity of the operators, the assumption that the time evolution does not effect the apparatus memory, and the assumption that the measurement is a von Neumann measurement.

If we had evolved and measured the time evolution of a general system state $|\psi\rangle$, the results would have been broadly speaking the same. For example, if we had chosen $|\psi\rangle = |\uparrow\rangle + |\rightarrow\rangle + |\downarrow\rangle + |\leftarrow\rangle$, then if the a_{ij}'s were chosen as in (3.19) but with the added proviso that $T(1)|\psi\rangle = |\psi\rangle$, then there would have been four maximum probability worlds, each of which would be observed to evolve "classically", as we have defined it. The "classical" worlds would be defined by the initial state recorded in the first memory slot: for each "classical" world, the recorded value i in the first slot would be u, r, d or l, and the value recorded in the kth slot would be $i + k$. The overlap between the system wave packets would be enormous, but the overlap would not be seen by the measurement apparatus.

We have hitherto assumed in our analysis that the eigenspectra of both the apparatus and the system are discrete. This is a convenient but not essential assumption. If one wishes to have the universe be split cleanly by a measurement into distinct, non-overlapping worlds then it will be necessary to assume that at least one of the system and apparatus have a discrete spectrum. It need not be both that have a discrete spectrum, but one must. To see this, we shall use a notation for the MWI measurement developed by DeWitt [20]. DeWitt's basis states are $|s\rangle|A\rangle$, where s labels the system variable (which he assumes to be discrete), and A labels the apparatus variable (which he assumes to be continuous). A measurement of the system variable s is accomplished by the operator M, which has the effect

$$M|s\rangle|A\rangle = |s\rangle|A + gs\rangle \tag{3.20}$$

where g is some small coupling constant. The result of a measurement on a general universe state $|\Psi\rangle = |\psi\rangle|\Phi\rangle$, where $|\psi\rangle$ is a general state of the system and $|\Phi\rangle$ is a general state of the apparatus, is thus

$$M|\Phi\rangle = \sum_s \langle s|\psi\rangle \int \langle A|\Phi|A + gs\rangle \, \mathrm{d}A \, . \tag{3.21}$$

In order for (3.21) to make sense, the system and apparatus state spaces must be Hilbert spaces. DeWitt normalizes the bases as follows:

$$\begin{gathered} \langle s \,|\, s'\rangle = \delta_{ss'}, \qquad \langle A \,|\, A'\rangle = \delta(A - A') \\ \sum_s \int |s\rangle|A\rangle\langle s|\langle A| \, \mathrm{d}A = 1 \, . \end{gathered} \tag{3.22}$$

Since the apparatus variable A is continuous, the measurement (3.20) will not split the apparatus into clearly distinct apparatus states unless we restrict the support of the initial apparatus state $|\Phi\rangle$. We require $|\Phi\rangle$ to satisfy

$$\Delta A \ll g \, \Delta s \tag{3.23}$$

where Δs is the spacing of the discrete system variable s (that is, the distance between adjacent s

values), and ΔA is the root mean square deviation of A from its average value defined by the wave function $\langle A \mid \Phi \rangle$:

$$\Delta A^2 = \frac{\int (A - \langle A \rangle)^2 \, |\langle A \mid \Phi \rangle|^2 \, \mathrm{d}A}{\int |\langle A \mid \Phi \rangle|^2 \, \mathrm{d}A} \tag{3.24}$$

$$\langle A \rangle = \langle \Psi | A | \Psi \rangle / \langle \Psi \mid \Psi \rangle = \langle \Phi | A | \Phi \rangle / \langle \Phi \mid \Phi \rangle \,.$$

To see the effect of condition (3.23), it is useful to imagine the function $|\langle A \mid \Phi \rangle|^2$ to be a Gaussian distribution with standard deviation ΔA, and peak at $\langle A \rangle$. The result of measuring a system which happened to be in a system basis state would be, from (3.21),

$$\begin{aligned} M|s\rangle|\Phi\rangle &= |s\rangle \int \langle A \mid \Phi \rangle \, |A + gs\rangle \, \mathrm{d}A \\ &= |s\rangle \int \langle A - gs \mid \Phi \rangle \, |A\rangle \, \mathrm{d}A \end{aligned}$$

which means that the peak of the Gaussian is moved from $\langle A \rangle$ to $\langle A \rangle - gs$. If condition (3.23) holds, then there will be essentially no overlap between the Gaussian in its shifted position, and the Gaussian placed in its original position.

With condition (3.21) the effect of the measurement operator on a general system state $|\psi\rangle = \Sigma \, |s\rangle \langle s \mid \psi \rangle$ is to split the apparatus wave function into a number of non-overlapping wave functions, one for each value of the system variable s. This sum of non-overlapping wave functions is essentially what (3.21) denotes. The single Gaussian wave packet going to several non-overlapping wave packets is analogous to the transition $|\mathrm{n}\rangle \to \{|\mathrm{u}\rangle, |\mathrm{d}\rangle\}$ which occurred in the discrete variable apparatus we discussed above. Condition (3.21) insures that the basis states of the continuous variable apparatus do not interfere after the measurement. Such a condition is not necessary if the apparatus memory has discrete variables only.

If both systems and apparatus variables are continuous, then the sum in (3.20) must be replaced by an integral over s. There is no obvious replacement for (3.21); we cannot restrict $g^2 \Delta s^2 = g^2 \int (s - \langle s \rangle)^2 |\langle s \mid \psi \rangle|^2 \, \mathrm{d}s$, because we want to measure system states more general than narrow Gaussian wave packets. Whatever the form of $\langle s \mid \psi \rangle$, the effect of the measurement is to correlate each s with an A. The original narrow Gaussian wave packet $|\langle A \mid \Phi \rangle|^2$ is smeared out by the measurement operator; it is no longer a narrow Gaussian, but a wide one, and it cannot be represented as a superposition of non-overlapping wave packets. It will prove instructive to demonstrate this in detail. Before the measurement the universe is in the state

$$|\Psi\rangle = |\psi\rangle|\Phi\rangle = \int \int \mathrm{d}s \, \mathrm{d}A \, |s\rangle \, |A\rangle \, \langle s \mid \psi \rangle \langle A \mid \Phi \rangle$$

where the continuous variables s, A will be assumed to go from $-\infty$ to $+\infty$, and $|\langle A \mid \Phi \rangle|^2$ will be assumed to be a narrow Gaussian in A with peak at $\langle A \rangle$ and standard deviation ΔA. The measurement operator will have the effect $M|s\rangle|A\rangle = |s\rangle|A+gs\rangle$ on the basis states. A measurement on the universal state $|\Psi\rangle$ will yield a state

$$M|\Psi\rangle = \int\int \mathrm{d}s\,\mathrm{d}A\,|s\rangle|A+gs\rangle\langle s|\psi\rangle\langle A|\Phi\rangle \tag{3.25a}$$

$$= \int \mathrm{d}s\,\langle s|\psi\rangle|s\rangle\left[\int \mathrm{d}A\,|A\rangle\langle A-gs|\Phi\rangle\right] \tag{3.25b}$$

$$= \int \mathrm{d}A\,|A\rangle\left[\int \mathrm{d}s\,|s\rangle\langle s|\psi\rangle\langle A-gs|\Phi\rangle\right] \tag{3.25c}$$

$$= \int \mathrm{d}A\,|A\rangle|\psi\rangle_A \tag{3.25d}$$

where (3.25b) has been obtained by a change of apparatus variable. The state $|\psi\rangle_A$, which is defined to the state in brackets in (3.25c), represents the superposition of system basis states which have been correlated with a given apparatus variable A. The wave function $\langle A-gs|\Phi\rangle$ will be a Gaussian peaked at the value $A-gs=\langle A\rangle$. This implies that the value of s in $|\psi\rangle_A$ (wherein A is fixed) will be peaked at $s=(A-\langle A\rangle)/g$, with a spread of

$$\Delta s = \Delta((A-\langle A\rangle)/g) = \Delta A/g\,.$$

But in the total universal wave function (3.25d) the value of A is not fixed: it varies from $-\infty$ to $+\infty$. This means that the Gaussian in s, which is peaked at $(A-\langle A\rangle)/g$, is smeared out as far as the effective support of $\langle s|\psi\rangle$ will permit; if $\langle s|\psi\rangle$ = constant, then the Gaussian will be spread infinitely wide and will be non-normalizable. If s were discrete, then this would not occur because in this case $\langle s|\psi\rangle$ would in effect be very narrow Gaussians peaked at the integral values of s. These narrow Gaussians in the state $|\psi\rangle_A$ would cause the integrand in (3.25d) to have effective support only in separated regions of the variable A, and so the integral would be replaced by a sum. There would be a clear split in the apparatus states.

DeWitt [32] has called a superposition with system and apparatus variables both continuous, an "imperfect measurement", because there is no clear split in the apparatus memory states. However, DeWitt [32] claims a split can be forced in an imperfect measurement by introducing a second apparatus to measure the state of the first apparatus.

Let us investigate DeWitt's claim by calculating the state of the universe which consists of a system and two apparata after the first apparatus with initial state $|\Phi\rangle = \int \mathrm{d}A\,|A\rangle\langle A|\Phi\rangle$, measures the system state and after the second apparatus, with initial state $|\Gamma\rangle = \int \mathrm{d}B\,|B\rangle\langle B|\Gamma\rangle$, has measured the state of the first apparatus. The measurement operators are defined by:

$$M_A\,|s\rangle|A\rangle|B\rangle = |s\rangle|A+gs\rangle|B\rangle \tag{3.26a}$$

$$M_B\,|s\rangle|A\rangle|B\rangle = |s\rangle|A\rangle|B+hA\rangle\,. \tag{3.26b}$$

The effect of a measurement on a general universal state $|\Psi\rangle = |\psi\rangle|\Phi\rangle|\Gamma\rangle$ is then:

$$M_B M_A |\psi\rangle|\Phi\rangle|\Gamma\rangle = M_B \int\int \mathrm{d}A\, \mathrm{d}s\, |s\rangle|A\rangle\langle s|\psi\rangle\langle A - gs|\Phi\rangle \int \mathrm{d}B\, |B\rangle\langle B|\Gamma\rangle \tag{3.27a}$$

$$= \int\int \mathrm{d}A\, \mathrm{d}s\, |s\rangle|A\rangle\langle s|\psi\rangle\langle A - gs|\Phi\rangle \int \mathrm{d}B\, |B + hA\rangle\langle B|\Gamma\rangle \tag{3.27b}$$

$$= \int\int \mathrm{d}A\, \mathrm{d}s\, |s\rangle|A\rangle\langle s|\psi\rangle\langle A - gs|\Phi\rangle \int \mathrm{d}B\, |B\rangle\langle B - hA|\Gamma\rangle \tag{3.27c}$$

$$= \int \mathrm{d}B\, |B\rangle|\psi\rangle_B \tag{3.27d}$$

where we have defined

$$|\psi\rangle_B = \int\int \mathrm{d}A\, \mathrm{d}s\, |s\rangle|A\rangle\langle s|\psi\rangle\langle A - gs|\Phi\rangle\langle B - hA|\Gamma\rangle\,.$$

We shall assume as before that $\{s, A, B\}$ are all continuous variables with ranges from $-\infty$ to $+\infty$, and that $\langle A|\Phi\rangle$, $\langle B|\Gamma\rangle$ are both Gaussian wave packets, with peaks at $\langle A\rangle$ and $\langle B\rangle$, and standard deviation ΔA and ΔB, respectively.

To assess the splitting effect of the second apparatus, let us assume the system wave function to be flat: $\langle s|\psi\rangle = 1$. In this case, a close inspection of (3.27c) and (3.27d) shows that $|\psi\rangle_B$ does not have peaks, just certain distinct values of B. Rather, it is a state which is just as smeared out as $|\psi\rangle_A$ since the wave function product

$$\langle A - gs|\Phi\rangle\langle B - hA|\Gamma\rangle$$

will peak for any B for a suitable choice of A and s. In short, a second apparatus with continuous variables will not generate a measurement split in the universe. I fear I must disagree with DeWitt's claim to the contrary. A clear split will occur only if the system or apparatus variable is discrete, or if both are discrete. A consistent formulation of the Many-Worlds Interpretation requires that continuous variables be measured by discrete variables, at least if a "good measurement" is defined to be one in which there is a clear and unequivocal split.

One may get confused if one consults Everett's original paper on the question of measuring continuous variables, for he discussed [34] at length the "measurement" of the position of a particle by recording it in the position of a pointer, with both the pointer and particle position being regarded as continuous. However, it should be kept in mind that Everett did not define a "good measurement", a concept introduced by DeWitt [20]. I believe this concept is essential if the nature of the measurement process is to be understood. A "good measurement" is one which results in a clear split.

Since we shall be interested in measuring the radius of the universe, which is a continuous variable, it will be useful to formulate the measurement of a continuous system variable by a discrete apparatus variable. The state of the universe before the measurement will be denoted

$$|\Psi\rangle = |\psi\rangle|\Phi\rangle = \sum_A |A\rangle \int \mathrm{d}s\, |s\rangle\langle s|\psi\rangle\langle A|\Phi\rangle\,. \tag{3.28}$$

For concreteness, we will restrict the continuous variable s to $(0, +\infty)$ – the range of possible values of the radius of the Universe – and the range of the discrete variable A to the non-negative integers. Since the variable A is discrete, it will be able to distinguish values of s only if these values are sufficiently far apart. We will imagine that the measurement registers in the memory of the apparatus only the integer portion of the positive real number s, written as a decimal fraction. Thus actual $s = 78.879$ and 78.001 will be registered as $s = 78$, etc. Hence the effect of the measurement operation on the universe basis states will be

$$M\,|s\rangle\,|A\rangle = |s\rangle\,|A + s'\rangle, \qquad \text{where } s' = \text{integer}(s)\,. \tag{3.29}$$

For simplicity we shall assume $|\Phi\rangle = |0\rangle$, with $\langle 0\,|\,0\rangle = 1$. The effect of measuring a general system state is then

$$M\,|\psi\rangle\,|0\rangle = M \int \mathrm{d}s\,|s\rangle\,\langle s\,|\,\psi\rangle\,|0\rangle = \sum_A \int \mathrm{d}s\,|s\rangle\,\langle s\,|\,\psi\rangle\,|A + s'\rangle \tag{3.30}$$

where the sum is over all integers A consistent with the support of the wave function $\langle s\,|\,\psi\rangle$. In (3.30) we have a clear split; the measurement is "good".

A particularly instructive example of a good measurement of a continuous variable by a discrete variable apparatus is the Wilson cloud chamber experiment, which was first analyzed quantum mechanically by Mott and Heisenberg. In this experiment, the system variable is the position of a charged particle, an alpha particle, say, and this position is measured by exciting a series of atoms in a three dimensional array. Since an atom has a non-zero size a, the apparatus will not be able to measure the location of the alpha particle at any given time closer than a. This limitation is essentially the same as pointed out in the simple model above.

The alpha particle wave function will be a spherical wave outgoing from the nucleus from which it is emitted. By the time it reaches the cloud chamber, it can be approximated very accurately by a plane wave. The theory of measurement must explain how a plane wave function, which is spread out all over space, can give the localized straight line motion actually observed.

The explanation was given by Mott and Heisenberg (we shall follow the presentation of J.S. Bell [35]). The initial wave function of the alpha particle is

$$\psi(\boldsymbol{r}) = \exp(\mathrm{i}k|\boldsymbol{r} - \boldsymbol{r}_0|) \tag{3.31}$$

and ϕ_0 will denote the ground state of the array of atoms. Let

$$\phi(n_1, n_2, \ldots) \tag{3.32}$$

denote a state of the array in which atoms $n_1, n_2, \ldots$ are excited. If no alpha particle were present, the universal state would be the product of (3.31) and (3.32). Because of the interaction between the alpha particle and the atoms of the array, the universal wave function will be the sum of this product and the scattered waves produced by the interaction. In a multiple scattering approximation the scattered waves are

$$\sum_N \sum_{n_1, n_2, \ldots, n_N} [\phi(n_1, n_2, \ldots, n_N) \exp(\mathrm{i}k_N|\boldsymbol{r} - \boldsymbol{r}_N|)\, f_N(\theta_N)/|\boldsymbol{r} - \boldsymbol{r}_N|]$$
$$\times [\exp(\mathrm{i}k_{N-1}|\boldsymbol{r}_N - \boldsymbol{r}_{N-1}|)\, f_{N-1}(\theta_{N-1})/|\boldsymbol{r}_N - \boldsymbol{r}_{N-1}|] \cdots [\exp(\mathrm{i}k_0|\boldsymbol{r}_1 - \boldsymbol{r}_0|)/|\boldsymbol{r}_1 - \boldsymbol{r}_0|]\,. \tag{3.33}$$

The general term in (3.33) is a sum over all possible sequences of N atoms in the three dimensional array. The position of the n_j atom is denoted by $\boldsymbol{r}_j$; $k_j = (k_{j-1}^2/2m - e)^{1/2}$, where e is the atomic excitation energy; θ_j is the angle between $\boldsymbol{r}_j - \boldsymbol{r}_{j-1}$, and $\boldsymbol{r}_{j+1} - \boldsymbol{r}_j$ (or $\boldsymbol{r} - \boldsymbol{r}_N$ for $n = N$); $f_j(\theta)$ is the inelastic scattering amplitude for an alpha particle of momentum k_{j-1} incident on a single atom.

An explicit formula for $f(\theta)$ can be calculated in the Born approximation in terms of atomic wave functions, and it is found that $f(\theta)$ peaks in the direction of the incident alpha particle momentum k, with angular spread $\Delta\theta = (ka)^{-1}$. This means that the relative probability of observing a sequence of excited atoms $n_1, n_2, \ldots,$ will be greatest if these atoms lie essentially in a straight line, or rather in a cone of opening angle $\Delta\theta$. For an alpha particle of energy ~ 1 MeV, and with a typical atomic size of $\sim 10^{-8}$ cm, we have $\Delta\theta \sim 10^{-5}$ radians, so it is easy to see why we see the alpha particle track as a straight line.

However, it is not just a single straight line we should see. The relative probabilities for observing a sequence of atoms $n_1, n_2, \ldots,$ are given by the squares of the moduli of the coefficients of $\phi(n_1, n_2, \ldots)$, and there are many straight line sequences of atoms in the sum (3.33), each having approximately the same probability. It is clear from our previous discussion of the MWI how to interpret this: the universe is split by the first stack of atoms in the array, and subsequent excitations respect the original split. Any other measuring apparatus we could bring in to measure the excitations (e.g., ourselves) would also respect the split, as discussed above, and so we see a single straight line alpha particle track in the cloud chamber. The first atom in the array to be excited could be any atom, located at any point in the array, so there will be an enormous number of worlds in the universe. The split of the universe into clearly distinct straight lines will occur only if a, the atomic radius, is non-zero, for were a to be small in comparison to the alpha particle momentum, the opening angle would be so large that no single particle track would be apparent. This illustrates our previous assertion that a continuous variable can be measured only by a discrete variable if one wants a clean split between the worlds.

The above analysis is static since it is concerned with the spatial shape of the alpha particle tracks. However, a dynamical analysis [35] shows just what one would expect: the straight lines develop in time. It is worth considering in some detail the quantum dynamics for a one dimensional array of atoms and an alpha particle moving in one dimension, for this situation is very closely analogous to the problem of measuring the radius of the Universe in the Friedmann universe. The static wave function for the array and the alpha particle in one dimension is the same as (3.33), except that the factors $|\boldsymbol{r}_j - \boldsymbol{r}_{j-1}|$ in the denominator are removed. The array wave function $\phi(n_1, n_2, \ldots)$ now refers to a sequence of atoms whose positions are given by a single coordinate x. It will be useful to distinguish unexcited and excited atoms in the sequence, so we shall denote an unexcited atom in jth position in the array by 0_j, and an excited atom in the jth position by e_j. For example, with four atoms in the array the wave function for the second and fourth atoms unexcited and the other atoms excited would be $\phi(\mathrm{e}_1, 0_2, \mathrm{e}_3, 0_4)$. Initially the universal wave function is

$$\Psi_{\mathrm{i}} = \phi(0_1, 0_2, \ldots, 0_N) \exp(\mathrm{i}[kr - Et]) . \tag{3.34}$$

The interaction will be turned on at $t = t_0$, after which the wave function of the universe becomes

$$\Psi_{\mathrm{i}} + \sum_{N} \sum_{n_1, n_2, \ldots} \phi(n_1, n_2, \ldots) \exp(\mathrm{i}[k_N|x - x_N| - k_N^2 t/2m]) f_N(x, t)$$
$$\times \exp(\mathrm{i}[k_{N-1}|x_N - x_{N-1}| - k_{N-1}^2 t/2m]) f_{N-1}(x, t) \cdots \exp(\mathrm{i}[k_0|x_1 - x_0| - k_0^2 t/2m]) \tag{3.35}$$

where the $f_i(x, t)$ are the time dependent inelastic scattering amplitudes, m is the mass of the alpha particle, and the other symbols are defined as in (3.33). As in the static case the scattering amplitudes can be calculated in time dependent perturbation theory, and the most probable atomic states at any time $t > t_0$ are illustrated in fig. 1.

Since initially the alpha particle wave function is spread out equally over all space – that is, its squared modulus is independent of the position coordinate x – the first atom to become excited is equally likely anywhere in the one dimensional array. This first atom to be excited defines a branch of the universe for all succeeding time, and each atom in the array defines such a world. In each such world – the world defined by the ith atom being the first to be excited, say – the atom most likely to become excited next is in the $i + 1$ position and the most likely time of its excitation is $t = k_i/mL + t_i$, where L is the spacing between the array atoms, and t_i is the time at which the ith atom becomes excited. Figure 1 shows four such worlds, in which the first atoms to be excited are adjacent atoms, and the time is such that three further atoms along the line have become excited. The direction of the

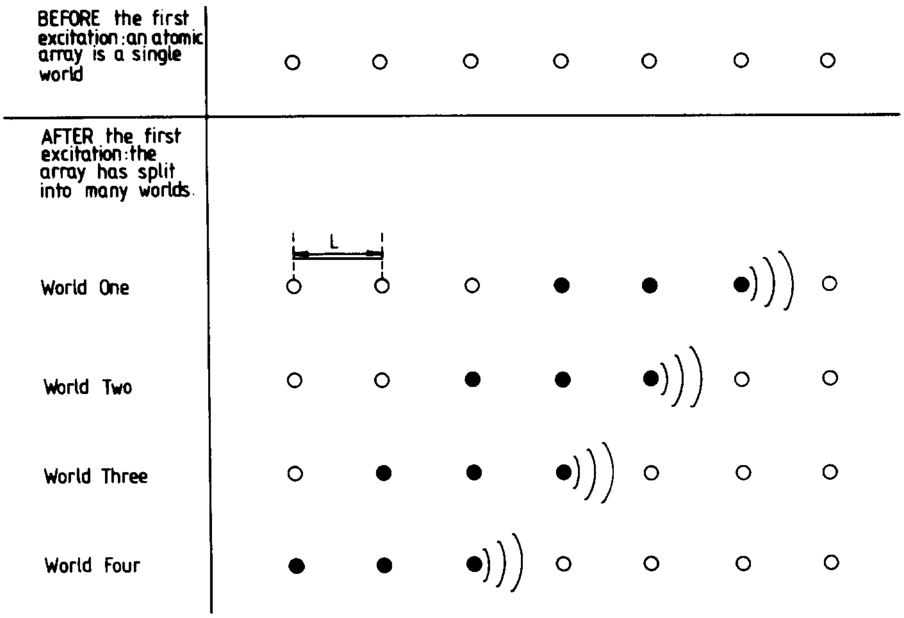

Fig. 1. *The Splitting of an Apparatus Designed to Measure the Position of an Alpha Particle as a Function of Time.* The apparatus consists of a one dimensional array of atoms which become excited by the passage of an alpha particle with definite energy. A darkened circle denotes an atom that has become excited, while an empty circle denotes an unexcited atom. The alpha particle momentum points from left to right. Each world is defined by the first atom to the left to become excited. Before the excitation of the first atom, the atomic array defines only one world, denoted by the single unexcited array at the top of the page. The universe is split into a large number of worlds by the first excitation. In the figure, four such worlds are shown. Each world pictured is the most probable world defined by the leftmost excited atom, wherein the next atom to be excited is the adjacent atom which is excited at time k_i/mL after the excitation of the first atom. (There would actually be one such most probable world defined by each atom in the array, and many worlds of lesser probability. The worlds of lesser probability are those in which excited atoms are interspersed with unexcited atoms.) In each world an outgoing wave packet is pictured moving to the right. The unexcited atom immediately to the right of the last excited atom is most likely to be excited when the packet reaches it. The above picture assumes that the alpha particle wave function is initially a plane wave, and that there is no interaction between the alpha particle and the array until $t = t_0$. In a more realistic model, the alpha particle would be a wave packet always in interaction with the atoms of the array, in which case the most probable worlds would be those whose first excited atoms are near the left hand side of the array. The splitting of the Universe into branch universes, however, is closer to the "unrealistic" analysis pictured above.

propagation of the excitation is pictured in fig. 1 by the direction of the outgoing waves from the third atom in each world. Again, we must emphasize that only one of the four worlds pictured in fig. 1 would be seen by a human observer, because he himself would split into four branches if he were to try to measure the state of the array of atoms at the given time.

It is important to note that the most probable time for the $i+1$ atom to be excited in the branch defined by the ith atom is given by the time a wave packet, of energy $k_i^2/2m$, would take to travel between the two atoms. This means that to investigate the most probable time evolution of a single branch (which is all we are physically capable of doing when we try to determine the time evolution of the entire Universe), it is sufficient to study the time evolution of a single wave packet of the appropriate characteristics outgoing from the first interaction center which measures the radius of the Universe.

We have assumed in the above analysis that the alpha particle had a single definite energy, which means a wave function spread out over all space. However, the essential features would remain if we were to analyze the measurement of an alpha particle wave packet which is localized in a region of physical space, and hence spread out in momentum space; i.e., being a superposition of plane wave functions with a range of energies. The splitting into worlds would be a bit more complicated, for the energy of a particle is determined by two position measurements: one at one time and another at some later time. Two position measurements, in other words, would determine a world, rather than a single position measurement in the single energy universe.

An incoming alpha particle wave packet would cause the atomic array to be split by the first two atomic excitations into worlds with all energies consistent with the support of the alpha particle wave packet and the discrete energy resolution of the atomic array.

The time evolution seen by the atomic array is in all essentials the same as that seen in our discrete time evolution model developed at length above. In both cases the splits occur for the dominant worlds in the first few interactions. There is, however, subsequent splitting into improbable branches at every measurement interaction. The most probable worlds will be those which evolve classically, as the Heisenberg–Mott analysis shows, and so in what follows I shall focus attention on them. When in doubt about what is going on in the more realistic continuous variable models, return to the toy discrete model.

4. The quantized Friedmann universe

We shall now quantize the closed Friedmann universe whose classical theory was developed in section 2. As discussed in section 2, our quantization procedure will be to quantize the unconstrained Hamiltonian, and attempt to take into account the constraints via boundary conditions.

The quantization of the unconstrained Hamiltonian is on the face of it quite simple since in all three cases considered in section 2 – dust, radiation, and non-interacting dust and radiation – the Hamiltonian is either the Hamiltonian for a SHO (the radiation case) or can easily be transformed into such a Hamiltonian by standard techniques. The only subtlety we must consider is the domain of the Hamiltonian operator: since a negative radius for the Universe makes no obvious sense, we shall restrict the domain to $R > 0$.

The quantum mechanics of the SHO on the domain $(0, +\infty)$ has been studied extensively [36, 37]. The key problem one faces on this domain is the problem of which boundary conditions to impose at the singularity $R = 0$. A straightforward calculation shows that in order for the operator $-\mathrm{d}^2/\mathrm{d}R^2 +$

$V(R)$, where the time independent potential and acts on functions which make it L^2 on $(0, +\infty)$, is regular at the origin, to be self-adjoint on $(0, +\infty)$, the operator must be restricted to those functions which satisfy one of the following boundary conditions:

either

$$\psi(R = 0, t) = 0 \tag{4.1}$$

or

$$\psi'(R = 0, t) + A\,\psi(R = 0, t) = 0\,. \tag{4.2}$$

Condition (4.1) is a boundary condition which Bryce DeWitt argued must be imposed on the wave function of the Universe, for it has the effect of keeping wave packets away from the singularity. I shall therefore call condition (4.1) the DeWitt boundary condition. Condition (4.2) has an arbitrary constant A. Were condition (4.2) the appropriate boundary condition to impose on the Universal wave function, then this constant would be a new fundamental physical constant. We could then avoid introducing a new physical constant only by requiring it to be zero; i.e., by imposing the boundary condition

$$\psi'(R = 0, t) = 0\,. \tag{4.3}$$

Both the DeWitt boundary condition and (4.2) tell what happens to wave packets when they hit the singularity at $R = 0$. It should be emphasized that in either case, the singularity is a real entity which influences the evolution of the Universe (or more precisely, its wave function) at all times via the boundary condition at the origin. In the classical universe, the singularity is present only at the end and at the beginning of time, so in a sense the singularity is even more noticeable in quantum cosmology than in classical cosmology.

Because they are the only boundary conditions which do not introduce a new physical constant, the DeWitt boundary condition or (4.3) are the most natural boundary conditions to impose. I shall henceforth restrict attention to these conditions only. The wave function of the Universe, $\Psi(R, \tau)$, can be expressed in terms of the boundary conditions $\Psi(R, \tau = 0)$ imposed at the beginning of time and the Green's function $G(R, \tilde{R}, \tau)$ via

$$\Psi(R, \tau) = \int_0^{+\infty} \mathrm{d}\tilde{R}\,\Psi(\tilde{R}, \tau = 0)\,G(R, \tilde{R}, \tau)\,. \tag{4.4}$$

The initial conditions $\psi(\tilde{R}, \tau = 0)$ are determined by the hitherto ignored constraint equations. I argued in section 2 that the effect of the constraint equations in the classical case was to require all classical solutions to pass through the singularity $R = 0$ when $\tau = 0$. I shall therefore attempt to include the constraints in my quantum model by requiring all quantum universes to do the same. The only way this can occur is if

$$\Psi(\tilde{R}, \tau = 0) = f(\tilde{R})\,\delta(\tilde{R})\,. \tag{4.5}$$

From the properties of the delta function, the functional form of $f(\tilde{R})$ is irrelevant since only $f(0)$ gives a contribution. The value of the constant $f(0)$ cannot be measured, even in principle, for it is normalization constant for the Universal wave function, and I argued in section 3 that such a constant is not measurable. Therefore for mathematical simplicity I shall set $f(0) = 1$.

The Green's function for the SHO on the domain $(-\infty, +\infty)$ can be found in many textbooks (e.g., [38]). If the boundary conditions at $R = 0$ are (4.1) or (4.3), then the Green's function for the SHO on $(0, +\infty)$ can be obtained from the Green's function on $(-\infty, +\infty)$ by linear superposition. If $\tilde{G}(R, \tilde{R}, \tau)$ denotes the Green's function on $(-\infty, +\infty)$, and $G(R, \tilde{R}, \tau)$ is the Green's function satisfying the appropriate boundary conditions at the origin, then for $\psi(\tilde{R}, \tau = 0)$ being an $L^2[0, +\infty]$ function:

$$G(R, \tilde{R}, \tau) = \tilde{G}(R, \tilde{R}, \tau) - \tilde{G}(R, -\tilde{R}, \tau) \tag{4.6}$$

if the boundary condition at the singularity is (4.1), and

$$G(R, \tilde{R}, \tau) = \tilde{G}(R, \tilde{R}, \tau) + \tilde{G}(R, -\tilde{R}, \tau) \tag{4.7}$$

if the boundary condition at the singularity is (4.3). I have been unable to obtain an expression in closed form for the boundary condition (4.2).

If the boundary condition $\Psi(\tilde{R}, \tau = 0)$ is not a smooth function but a distribution* with support at $R = 0$ – the situation we wish to consider – the appropriate Green's function for boundary condition (4.3) is

$$G(R, \tilde{R}, \tau) = \{\tilde{G}(R, \tilde{R}, \tau) + \tilde{G}(R, -\tilde{R}, \tau)\}/2\,. \tag{4.8}$$

The DeWitt boundary condition at the singularity is inconsistent with the initial boundary condition (4.5), as a simple calculation using (4.1) and (4.6) (or (4.6) times some constant) will show. Therefore, the boundary condition (4.3) is the appropriate singularity boundary condition to use to obtain the Universal Green's function.

Putting (4.8) into (4.1) and using the Hamiltonian obtained from (2.18) to generate the Green's function $\tilde{G}(R, \tilde{R}, \tau)$, we get for the wave function of a radiation-dominated Friedmann universe:

$$\Psi(R, \tau) = [3i/4L_{\text{Planck}} \sin \tau]^{1/2} \exp[(3\pi/4i)(\cot \tau)(R/L_{\text{Planck}})^2] \tag{4.9}$$

where I have put the units back in to show the scale dependence: L_{Planck} is the Planck length. The wave function (4.9) is actually just the Green's function $\tilde{G}(R, 0, \tau)$.

The wave function (4.9) not only begins as a delta function at $\tau = 0$, it recombines into a second delta function $\delta(R)$ when $\tau = \pi$; in other words, all quantum worlds terminate in a second singularity at $\tau = \pi$, just as all the classical closed Friedmann universes do. This shows that the initial boundary condition (4.5) is consistent, for the logic used to derive (4.5) requires that all the quantum universes terminate in a final singularity at $\tau = \pi$.

Although the wave function is scaled by the Planck length, as a quantum cosmology should be scaled, the scale only effects the wave function phase. The wave function modulus is independent of the radius of the Universe R, except at $\tau = 0$ and $\tau = \pi$. Since at the initial instant the Universal wave function (or more precisely, wave functional) is concentrated entirely at $R = 0$, it has all values of momenta initially.

* The quantum mechanics of distributions can be treated rigorously by regarding distributions and functions as both elements of a rigged Hilbert space; see [50] and [51] for a discussion.

These momenta cause the wave function to explosively spread out from the singularity to $+\infty$ the instant after $\tau = 0$.

The physical interpretation of this Universal wave function is essentially the same as that given a highly localized alpha particle wave packet in section 3. The first two measurements of the radius of the Universe will split the Universe into a large number of worlds, in each of which an almost classical motion will be observed. As we have shown at length in section 3, the measurement of any variable requires a physical variable wherein the measurement is recorded, and our simple Friedmann model contains no such variable. In the actual Universe the "measuring device" would be some non-gravitational field in the early universe which could define a scale length. The radiation gas is conformally invariant and so defines no intrinsic length, but a conformally invariant field can be used to define a non-intrinsic length: an electromagnetic wave Gaussian packet has its standard deviation as a length scale. The first such field to couple to the radius of the Universe, and which retains the result of the coupling on time scales long compared to the expansion of a given branch of the Universe was the actual "apparatus" that initially split the Universe. Whatever generated the perturbation spectrum that eventually gave rise to the galaxies would be a candidate for such a field, or it could be that the spectrum is a relict of the initial split. As with the localized alpha particle, the first two such field interactions would define the branches of the Universe. At the present time the galaxies themselves would serve as benchmarks for the radius measurements, as they do in classical cosmology for measurements of the radius of the Universe [14]. The observed motion of the alpha particle is approximately classical in each branch and is determined by the motion of the wave packets scattered from each atom in the array. Similarly, the observed motion of the Universe in each branch would be approximately classical (so long as the observer is far from the final singularity), but the scattered wave packets will be evolving in an harmonic oscillator potential. This will give the usual sinusoidal motion of a radiation universe in each branch, for the motion of $\langle R\rangle$ of wave packets in such a potential satisfies the harmonic oscillator equation. This interpretation of the Universal wave function is shown in fig. 2.

The Universe is split into branches by the first two measurements. Say this first measurement occurred at $\tau = \tau_{\text{initial}}$. Since the wave function modulus is independent of R, this means the probability of being in a world with radius R at τ_{initial} is independent of R: all classical universes are equally likely. In particular, there is no tendency for the worlds to be typically a Planck length in size at the time of maximum expansion, in this quantum cosmology at least. Some relativists have argued on dimensional grounds that such a tendency should be a generic property of quantum cosmologies. The Planck length is indeed a scale, but it scales the phase of the wave function, not the overall size of the Universe.

Another consequence of $|\Psi(R, \tau_{\text{initial}})|^2$ being independent of R is that it is overwhelmingly probable the particular world we happen to be in will have an enormous radius of maximum expansion. That is, the probability is $1-\varepsilon$ that the density parameter Ω in our particular branch of the Universe equals $1+\delta$, where ε and δ are true infinitesimals. To see this, we need only recall the discussion in section 3 about the meaning of relative probabilities calculated from non-normalizable wave functions: the probability of $|A\rangle$ relative to $|B\rangle$ is given by $|\langle\psi|A\rangle|^2/(|\langle\psi|A\rangle|^2 + |\langle\psi|B\rangle|^2)$ even if $|\psi\rangle$ is not normalizable. In the case of the radius of the Universe at τ_{initial}, the probability that the radius is smaller than a given radius R_1 relative to the probability that it is larger than R_1 is

$$\int_0^{R_1} |\Psi(R, \tau_{\text{initial}})|^2 \,\mathrm{d}R \Big/ \left(\int_0^{R_1} |\Psi(R, \tau_{\text{initial}})|^2 \,\mathrm{d}R + \int_{R_1}^{+\infty} |\Psi(R, \tau_{\text{initial}}|^2 \,\mathrm{d}R \right). \tag{4.10}$$

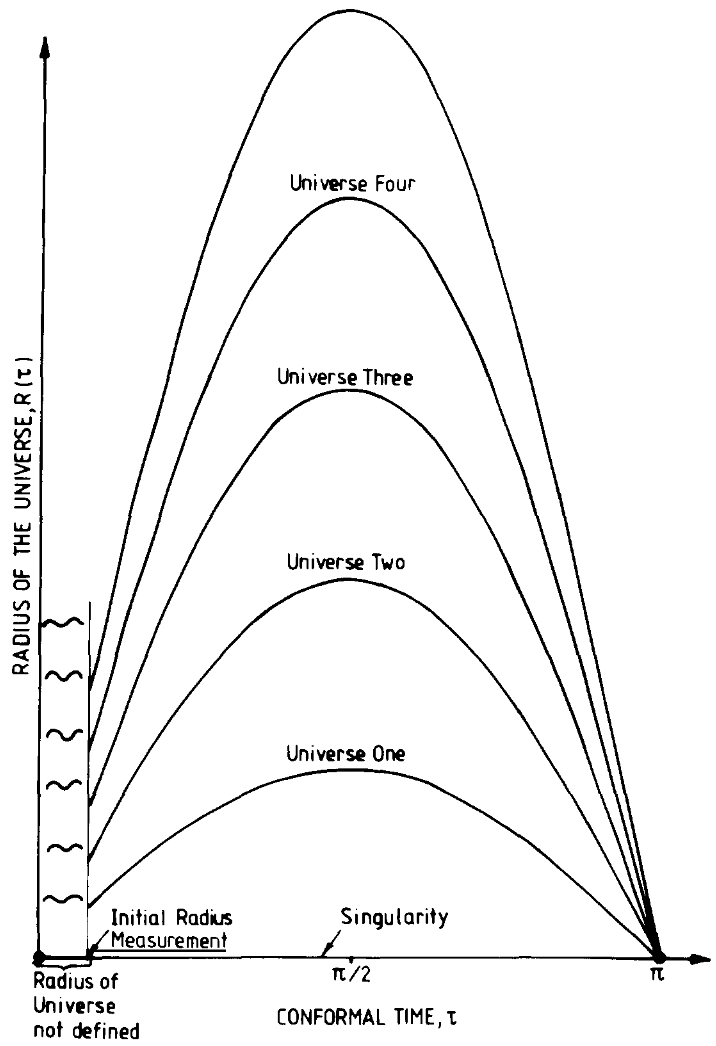

Fig. 2. *The Branching of a Quantum Universe.* Before the first interaction occurs that can encode a scale measurement, the Universe, represented before this interaction occurs as a series of wavy lines, has no radius. After the first two scaled interactions have occurred, the Universe has been split by the interactions into a large number of branches, in each of which an essentially classical evolution is seen. These branches are represented in the figure by the sine curves, each of which goes through the final singularity at $\tau = \pi$. The collection of all sine curves are all the classical radiation gas Friedmann models. Each curve is defined by R_{max}, the radius of the universe at maximum expansion. In the quantum Universe, all the classical universes are present, one classical universe defining a single branch. The classical universes are equally probable. Five such classical universes are pictured.

But this is zero, which gives the result claimed. To put it simply, if we pick a single integer (= radius of Universe to the nearest parsec) at random from the set of all positive integers, and if all integers are equally probable, then it is overwhelmingly probable the integer we pick will be an extremely large integer.

We do not actually need the various worlds to be equally probable in order for the value of Ω we would measure to be infinitesimally close to 1. The expression (4.10) would be zero for any non-normalizable wave function, which is regular at the origin, since with such a wave function the second term in the denominator would be infinite. Thus whatever the actual probability distribution as a function of R, any non-normalizable wave function would yield an overwhelmingly most probable value of Ω of $1+\delta$. To the best of my knowledge, Narlikar and Padmanabham [40] were the first to suggest that quantum gravity might naturally lead to the prediction $\Omega = 1$.

It is a general rule in physics that what is not forbidden is compulsory, and we showed in section 3 that there was no physical reason to require the wave function of the Universe to be normalizable. Therefore, we would expect that the correct quantum gravity theory would yield a non-normalizable Universal wave function. The general Universal wave function (not the special Friedmann model) advanced by Hartle and Hawking [1] probably has this property. The non-normalizability of the Hartle–Hawking wave function arises from the desire of its creators to include all possibilities in the Feynman sum over histories.

Another general rule in cosmology is the Copernican Principle: our place in the Universe is typical. In standard cosmology the word "place" is interpreted to mean position in space: the Universe on a sufficiently large scale ought to have the same properties independently of position. But even in classical cosmology there is another possible meaning to the word "place" – position in initial data space. In the case of the radiation-dominated Friedmann universe, the initial data space is parameterized by one variable, which can be chosen to be the radius of the Universe at a set time τ_1 (it is conventional in classical cosmology to pick τ_1 to be $\pi/2$, the time of maximum expansion). The Flatness Problem is essentially the problem of explaining why, out of all possible points in the one dimensional initial data space of the Friedmann universe, we happen to live in a very special point corresponding to a very large radius at maximum expansion. (This is equivalent to asking why Ω is extremely close to 1.) In classical cosmology the only possible answer to this question is to say that we have been misinformed as to the size of the initial data space: there are more forces governing the expansion of the Universe than a radiation gas coupled to gravity, and these other forces restrict the actual Friedmann initial data space to a narrow range around $\Omega = 1$, at least in our neighborhood in space. Such an answer to the Flatness Problem is the one provided by the inflationary universe model [39]. However, the inflationary model does not provide a unique value for the Universal initial condition. Although the initial data space is reduced in size, it is still not reduced to a single point, and so the question of why we happen to live in a very special Universe defined by a definite particular value of the radius at maximum expansion is left unanswered by the inflationary model. Indeed, any classical cosmological model must leave this question unanswered. It would also remain unanswered in any interpretation of quantum mechanics that has some force responsible for wave function reduction.

But it has an answer in quantum mechanics if we accept the Many Worlds Interpretation, for here we have the possibility of having many universes, each defined by a different radius at any given time, existing simultaneously. The whole of initial data space can be spanned by the various universes. Each point in the initial data space would be as real as the points in the sensible three dimensional physical space. Thus we should expect the Copernican Principle to apply to the initial data space as it applies to three dimensional physical space.

A *Quantum Copernican Principle* would require that all the points in classical initial data space would be equally probable; we would be no more likely to find ourselves in one classical universe than another. We have seen that the quantum cosmological model defined above has this property, but I would expect that any accurate model of the Universe would have this property if the Quantum Copernican Principle were true. A consequence of the Quantum Copernican Principle is a non-normalizable wave function if the wave function domain is $(0, +\infty)$, which leads to an $\Omega = 1+\delta$ prediction.

It is a standard procedure in quantum mechanics to expand a wave function in terms of the eigenfunctions of a self-adjoint operator $\hat{H}$, and the expectation value of the operator $\hat{H}$ with respect to any wave function $|\psi\rangle$ is then written in terms of the eigenfunctions of $\hat{H}$. I have used this procedure extensively in section 3. However, in standard quantum mechanics each term in the mathematical series for the expectation values can actually be directly measured, since

$$\langle\psi|\hat{H}|\psi\rangle = \sum H\,|\langle\psi\,|\,H\rangle|^2 \tag{4.11}$$

and the relative frequency $|\langle\psi\,|\,H\rangle|^2$ can be measured by conducting a series of experiments on an ensemble $|\psi\rangle\,|\psi\rangle\ldots$ as discussed in section 3.

However, in quantum cosmology, this series of experiments cannot be carried out, for there is only one Universe: it is not possible to form an ensemble of Universe states $|\Psi\rangle\,|\Psi\rangle\ldots$. The only "ensemble" available in quantum cosmology is the collection of various branches into which the Universe is split by the first measurements. This "ensemble" is utterly different from the ensemble needed to define the terms in (4.11). Thus no expectation value in quantum cosmology can be measured physically, and this is true even if the wave function of the Universe were normalizable. Thus expectation values are without physical significance in quantum cosmology, even though the concept is useful mathematically.

Similarly, although many treatments of quantum cosmology imagine the Universe to be in an eigenstate of the Hamiltonian, there is no way of measuring the Universal wave function to see if in fact it is such an eigenstate. As shown above, the measurements of position which determine the branch of the Universe we happen to be in would tell us that the branch is approximately in an eigenstate, whatever the initial spectral shape of the Universal wave function.

I should add one caveat to this criticism of the use of expectation values. The evolution of each branch of the Universe is closely modeled by the evolution of a single wave packet, by the process described above. Thus the evolution of $\langle R\rangle$ where the expectation value is taken with respect to the wave packet normalizable wave function, will give us a picture of how a branch evolves. But it will not give a picture of the evolution of the Universal wave function.

If expectation values, which are the means in standard quantum mechanics by which operators are given a physical interpretation, have no physical meaning in quantum cosmology, the question arises of why the Hamiltonian operator should be required to be self-adjoint. Self-adjointness is usually required in order to guarantee the existence of a complete set of operator eigenstates, so the series (4.11) would be defined. The reality of the eigenvalues is assured by Hermiticity; the vastly more powerful assumption of self-adjointness is not needed to obtain real eigenvalues. The self-adjointness of the Hamiltonian is needed if the time evolution it defines is to be unitary – if the wave function norm is to be conserved. But again, if these norms cannot be measured, why should we require them to be conserved?

I think the self-adjointness requirement should be retained in quantum cosmology for two reasons: first, the complete set of states was very important in determining that the Universe actually split into various branches and for observing the evolution of each branch by following the evolution of wave packets – it is not clear that the Universe would split if this complete set did not exist. Second, without the spectral theorem for self-adjoint operators, we shall be unable to define functions of operators. For example, on the domain $(0, +\infty)$ the momentum operator $\hat{P}$ can be defined as a unique self-adjoint operator only via $\hat{P} = (\hat{P}^2)^{1/2}$, for $\hat{P}^2$ can be defined on this domain as a self-adjoint operator even if $\hat{P} = \mathrm{i}\,\mathrm{d}/\mathrm{d}R$ is not a self-adjoint operator, and cannot be extended to one. If we give up self-adjointness, we give up uniqueness. Gotey and Demaret [7] have discussed whether the Universal Hamiltonian should be self-adjoint at some length. They consider the time evolution of wave packets evolved by various non-self-adjoint Hamiltonian operators.

I have hitherto concentrated attention on the radiation gas quantum cosmology. There is a major difficulty in the interpretation if we try to quantize the dust model or the two-fluid model (2.25): in both cases the Lagrangian contains a potential term which is not conformally invariant. The potential term contains a scale constant C^{dust} which does not come from the branching interaction, but exists in the underlying Friedmann model. The Green's functions for the dust and two-fluid Lagrangians can be written in closed form; for example, if we choose the scale to be the Planck length, the Green's function at the origin for dust is

$$\tilde{G}(R, 0, \tau) = (3\mathrm{i}/4L_{\text{Planck}} \sin \tau)^{1/2} \exp[(3\pi/4\mathrm{i})\phi]$$

where

$$\phi = (R/L_{\text{Planck}})^2 \cot \tau - (R/L_{\text{Planck}})\,(1 - \cos \tau)/\sin \tau - \tau\,.$$

The scale factor does not appear in the modulus. This means that it is impossible to choose the scale factor so that the second singularity occurs at $\tau = 2\pi$, as it must if the quantization procedure is to be consistent. The second singularity occurs at $\tau = \pi$, as in the radiation gas universe. This could just show that my quantization procedure is incorrect, but I fear the trouble is deeper than that. Any quantization procedure applied to a Lagrangian with an intrinsic length scale would single out this length in the wave function evolution; it would be impossible to retain the nice democracy of branch universes which is permitted in the interpretation of the radiation gas wave function. A Quantum Copernican Principle would be impossible.

I therefore suggest that whatever the quantization procedure, the only allowed other fields must be radiation fields; the scales must arise by the branching into worlds, and must not be in the basic quantized physical fields.

It is well-known that classically, the Universe cannot now be radiation dominated, for a radiation dominated Universe today would cause the expansion at the nucleosynthesis era to be too slow to give the correct helium abundance. However, the expansion rate in a branch could appear to be that of a dust dominated classical universe even though the quantum Hamiltonian which is controlling the overall evolution is a radiation gas Hamiltonian.

To see this, recall that the observed time evolution in a branch is roughly the same as the time evolution of the position expectation value of a wave packet scattered by an "apparatus" interaction site, the cosmological analogue of an atom in the cloud chamber array. The time evolution of $\langle R \rangle$ is given by Ehrenfest's equation, modified by the boundary conditions imposed at $R = 0$ in order to make the SHO Hamiltonian a self-adjoint operator. This equation can be written

$$\mathrm{d}^2\langle R\rangle/\mathrm{d}\tau^2 = -(1/m)\,\langle \mathrm{d}V/\mathrm{d}x\rangle + F(R = 0)\,. \tag{4.12}$$

The F term in (4.12) is an extra term which arises from having the boundary at a finite distance rather than infinity. The detailed functional form of this term is not important for our present purposes; it depends on the form of the boundary conditions at the origin. It is sufficient to note that the term acts as a repulsive force on wave packets at the origin to keep them from passing through $R = 0$. If a normalized wave packet were to pass through $R = 0$, the norm could not be preserved, which a self-adjoint Hamiltonian would preserve. The motion of $\langle R \rangle$ would still be sinusoidal for a harmonic oscillator potential if F were constant, but the sine curve would be shifted above the $R = 0$ line with the effect that the time evolution of $\langle R \rangle$ would look more like a cycloid than a sine curve, and a cycloid is the shape of a dust dominated universe as opposed to the sine shape of the radiation dominated universe.

Whether this suggestion as to how dust dominated evolution would be observed in a radiation dominated quantum universe would actually work would require a detailed analysis of the interaction between the apparatus and the Universe, and is beyond the scope of the present paper. I offer it as a suggestion only. If the suggestion pans out it could provide another argument for self-adjointness, for this requirement is the source of the boundary term.

5. Conclusions and suggestions for further research

I have attempted in this paper to give a systematic treatment of how the quantum theory of measurement should be applied to quantum cosmology. The ideas were illustrated with a specific quantized Friedmann model, but they are sufficiently general to tell us what to expect in most models obtained by other quantization techniques. In summary, my analysis shows that whatever the model and whatever the initial wave function of the Universe, the Universe at very early times cannot be said to have a radius. At some point in the early universe, the Universe is split into a large number of branches, in each of which an essentially classical evolution is seen by the observers in that branch. At this stage in the evolution of the Universe, the Universe can be said to have not merely one radius, but rather all radii consistent with the support of the wave function and the resolution of the forces responsible for the split.

I have been rather vague about the details of these forces. This is, I am afraid, inevitable in the present stage of physics, for the splitting forces would have done their work in the very early universe, where our knowledge of the interactions is quite limited. Nevertheless, the existence of the split and the subsequent time evolution of the branches should not depend on the details of the forces responsible for the splits, any more than the splitting of an atomic array into various worlds in the Wilson cloud chamber depends on the details of the atoms.

But the details are obviously crucial in understanding the future evolution of the atoms. Since the analogue of the atomic array is the inhomogeneous material in the Universe, for only such material can record scales, the details which have been ignored in this paper are going to determine the origin and subsequent evolution of the Perturbation Spectrum. It would appear that one of the major problems in cosmology, accounting for the amplitude and shape of the initial perturbations, would be solved if one understood the details of the initial split. It could be that the split would not occur if the inhomogeneous perturbations were sufficiently large, and this could be the solution of the Horizon Problem. Another possible solution is that a quantum cosmology with more degrees of freedom put in would lead to the isotropic worlds being having a probability of essentially one. Such a solution has been proposed by Narlikar and Padmanabham [40], and Hawking and Luttrell [42]. This solution is analogous to the solution of the Flatness Problem first proposed by Narlikar and Padmanabham, and defended in this paper.

This solution to the Flatness Problem requires that the density parameter Ω must be only infinitesimally greater than 1. As shown at length in the body of this paper, such a prediction is a consequence of any quantum cosmology which has a non-normalizable wave function. Thus it is not necessary to invoke inflation in order to account for $\Omega = 1$. In principle, $\Omega = 1$ arising from quantum gravity could be distinguished experimentally from that arising from inflation, for the effective cosmological constant in inflation does not quite inflate the Universe completely down to $\Omega = 1$; in the new inflationary universe, for example, the density parameter would actually be $1 + 10^{-6}$. In practice, it is quite impossible to distinguish these two cases by astronomical observation. However, the inflationary universe has come under severe attack recently; it seems that inflation cannot after all explain isotropy [41], and it is not at all clear that a spontaneous symmetry-breaking scalar field will lead to inflation [43]. Thus quantum gravity may be the only remaining candidate capable of solving the above cosmological problems.

The Many Worlds cosmology doesn't seem to throw any light on the Magnetic Monopole Problem. Thus it would appear that if a non-normalizable wave function is responsible for the density parameter being 1, then the explanation for the absence of huge numbers of monopoles in the Universe must come from a study of local interactions in particle physics, not cosmology.

A Many-Worlds quantum cosmology does put one powerful restriction on the matter content: the fundamental Lagrangian must be conformally invariant, because the scale of the universe we see must arise from the splitting into branches. It cannot be present in the Lagrangian. This is quite different from the point of view arising from classical cosmology and indeed from other interpretations of quantum mechanics, in which the Universe as a whole defines a scale, the present radius of the Universe. In the MWI view, the universe we see is but one of many branches, each being equally real. The requirement that all these classical branches included in the wave function governs the choice of time parameters, the boundary conditions on the wave function, and the way the constraints are handled. Thus a MWI approach to quantum gravity yields a class of models which are physically distinct from the cosmological models suggested by other interpretations. Compare, for example, my model with that of Narlikar [40]. Furthermore, I shall show in the Appendix that the MWI can single out a preferred time for the ADM quantization procedure. Since the models implied by the interpretation are different, it becomes possible in principle to distinguish the interpretations experimentally in cosmology.

I have provided in this paper a very rough picture of what a quantum Universe would look like from the Many Worlds point of view. I have discussed a number of esthetic reasons for preferring the Many Worlds Interpretation over other interpretations, most notably the fact that with the Many Worlds Interpretation, which assumes the various branches are real, the problem of the boundary conditions restricting the classical Universe is obviated. I hope others will be encouraged to think about cosmology from the Many Worlds point of view.

Acknowledgments

I should like to thank for helpful discussions J. Barrow, B. Berger, D. Deutsch, G. Ellis, J. Goldstein, K. Kuchar, R. Matzner, D. Page, R. Penrose and J. Wheeler. This work was supported in part by the National Science Foundation under grant number PHY84-09672.

References

[1] J. Hartle and S.W. Hawking, Phys. Rev. D 28 (1983) 2960.
[2] H. Everett, Rev. Mod. Phys. 29 (1957) 454.
[3] B.S. DeWitt, Phys. Rev. 160 (1967) 1113.
[4] W. Heisenberg, The Physical Principles of Quantum Theory (University of Chicago Press, Chicago, 1930) pp. 66–76.

[5] N.F. Mott, Proc. Roy. Soc. A126 (1929) 76; reprinted in [6].
[6] J.A. Wheeler and W.H. Zurek, Quantum Theory and Measurement (Princeton University Press, Princeton, 1983).
[7] M.J. Gotay and J. Demaret, Phys. Rev. D 28 (1983) 2402.
[8] S.W. Hawking, in: General Relativity: An Einstein Centenary Survey, eds. S.W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979) p. 746.
[9] B.S. DeWitt, in: Quantum Gravity 2, eds. C.J. Isham, R. Penrose and D.W. Sciama (Oxford University Press, Oxford, 1981) p. 449.
[10] B.S. DeWitt, Sci. Am. 249, #6 (1983) 112.
[11] F.J. Tipler, Gen. Rel. Grav. 15 (1983) 1139.
[12] S.W. Hawking, D.N. Page and C.N. Pope, Nucl. Phys. B170 (1980) 283.
[13] A. Einstein, in: The Principle of Relativity, ed. A. Einstein (Dover Publications, New York, 1923) pp. 177–183.
[14] C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (Freeman, San Francisco, 1973).
[15] S.W. Hawking and G.F.R. Ellis, The Large Scale Structure of Space-Time (Cambridge University Press, Cambridge, 1973).
[16] M.A.H. MacCallum, in: Quantum Gravity, eds. C.C. Isham, R. Penrose and D.W. Sciama (Oxford University Press, Oxford, 1975) p. 174.
[17] W.F. Blyth and C.J. Isham, Phys. Rev. D.11 (1975) 768.
[18] J.E. Marsden and F.J. Tipler, Phys. Reports 66 (1980) 109.
[19] M. Jammer, The Philosophy of Quantum Mechanics (Wiley, New York, 1974).
[20] B.S. DeWitt and N. Graham, The Many-Worlds Interpretation of Quantum Mechanics (Princeton University Press, Princeton, 1973).
[21] B.S. DeWitt, in ref. [20] p. 168.
[22] B.S. DeWitt, in: Battelle Rencontres: 1967 Lectures in Mathematics and Physics, eds. C. DeWitt and J.A. Wheeler (W.A. Bejamin, New York, 1968).
[23] Ref. [20] p. 143.
[24] Ref. [20] p. 157.
[25] Ref. [20] p. 116.
[26] Ref. [20] p. 117.
[27] Ref. [20] p. 161.
[28] J. Hartle, Am. J. Phys. 36 (1968) 704.
[29] D. Finkelstein, Trans. N.Y. Acad. Sci. 25 (1963) 621.
[30] N. Graham, in ref. [20].
[31] I.M. Khalatnikov, An Introduction to the Theory of Super-fluidity (W.A. Benjamin, New York, 1965).
[32] B.S. DeWitt, ref. [20] p. 210.
[33] J.A. Wheeler, The Monist 47 (1962) 40.
[34] H. Everett [2], reprinted in [20] p. 143.
[35] J.S. Bell, in: Quantum Gravity 2: A Second Oxford Symposium, eds. C.J. Isham, R. Penrose and D.W. Sciama (Oxford University Press, Oxford, 1981) p. 611.
[36] M. Reed and B. Simon, Methods of Modern Mathematical Physics, Volume II: Fourier Analysis, Self-Adjointness (Academic Press, New York, 1975) chapter 10.
[37] B. Simon, Quantum Mechanics for Hamiltonians Defined as Quadratic Forms (Princeton University Press, Princeton, 1971).
[38] L.S. Schulman, Techniques and Applications of Path Integration (John Wiley, New York, 1981) chapter 6.
[39] A. Guth, Phys. Rev. D23 (1981) 347.
[40] J.V. Narlikar and T. Padmanabham, Phys. Reports 100 (1983) 151.
[41] J.D. Barrow and M.S. Turner, Nature 292 (1981) 35.
[42] S.W. Hawking and J.C. Luttrell, Nucl. Phys. B247 (1984) 250.
[43] R. Wald, W. Unruh and G. Mazenko, Phys. Rev. D31 (1985) 273.
[44] V.G. Lapchinskii and V.A. Rubakov, Theo. Math. Phys. 33 (1977) 1076.
[45] B.F. Schutz, Phys. Rev. D2 (1970) 2762.
[46] B.F. Schutz, Phys. Rev. D4 (1971) 3559.
[47] R. Abraham and J.E. Marsden, Foundations of Mechanics (W.A. Benjamin, New York, 1978).
[48] C.W. Misner, in: Magic Without Magic: John Archibald Wheeler, ed. J.R. Klauder (Freeman, San Francisco, 1972).
[49] D. Deutsch, Int. J. Theo. Phys. 24 (1985) 1.
[50] A. Bohm, Quantum Mechanics (Springer-Verlag, Berlin, 1979).
[51] A. Bohm, Rigged Hilbert Space and Quantum Mechanics: Lecture Notes in Physics #78 (Springer-Verlag, Berlin, 1978).

Appendix: The many-worlds interpretation in the ADM formalism

I shall demonstrate in this appendix that if the time coordinate is chosen to be conformal time, then the ADM formalism also implies that for quantized radiation-filled closed Friedmann universes, the evolution is governed by eq. (2.28), where H is, up to unimportant numerical constants, the SHO operator $d^2/dR^2 + R$. In other words, I shall show that the quantization method of section 2 yields exactly the same quantum equation as the more standard ADM method of handling the general relativity constraints. In the ADM method, however, one does not obtain the initial boundary condition $\Psi(0,0) = \delta(R)$, which must be put in by hand. My discussion will follow that of Lapchinskii and Rubakov [44], who were the first to show that the quantum mechanics of radiation-filled closed Friedmann universes was that of the SHO. I shall point out that this result can probably be extended to all isentropic perfect fluids, provided the MWI is used to determine the quantization variables. In particular, I shall show that Misner's quantization of the radiation-filled Friedmann universe is not allowed by the MWI.

The ADM formalism for the Friedmann universe is based not on the metric (2.2), but rather on

$$ds^2 = -N^2(t)\,dt^2 + R(t)\,[d\chi^2 + \sin^2\chi\,(d\theta^2 + \sin^2\theta\,d\phi^2)] \tag{A.1}$$

where the function $N(t)$, called the shift, ultimately will be determined after the variation by the choice of time variable. The shift cannot be fixed prior to the variation because in the ADM formalism, it plays the role of a Lagrange multiplier. Furthermore, in standard ADM theory the two boundary terms in (2.1) are set equal to zero. In the presence of matter in the form of a perfect fluid, the ADM action is written:

$$S = \int (R + p)(\sqrt{-g})\,d^4x = \int L_{\mathrm{ADM}}\,dt \tag{A.2}$$

where p is the pressure of the fluid [7]. The matter Lagrangian is thus $L_m = p$, rather than $-\mu$ as in Hawking and Ellis (see eq. (2.10)). For simplicity, the factor $(16\pi)^{-1}$ has been suppressed. Now perfect fluids are difficult to handle in a variational principle formulation of general relativity, because if one tries to use only the usual fluid variables, the variation must be constrained. In section 2, the constraint was the conservation of the current vector: the variation was restricted by eq. (2.11). As I discussed in section 2, one wishes in action principles to minimize the constraints. For this reason, it has become standard practice when dealing with perfect fluids in the ADM formalism to regard the perfect fluid variables μ and p as functions of scalar fields ϕ, θ and s. The variable s is the specific entropy, but ϕ and θ have no direct physical significance. The variables ϕ, θ, s are called the Schutz velocity potentials, after B. Schutz who showed [45, 46] that perfect fluids which depend on μ, p, and the specific enthalpy h have an elegant Hamiltonian formulation if the independent variables are ϕ, θ, s. These fields are called velocity potentials because the four-velocity u^a of the fluid is written in terms of these potentials:

$$u^a = h^{-1}(\phi_{,a} + \theta s_{,a})\,. \tag{A.3}$$

The equation of state is assumed to be of the form $p = p(h, s)$, and the specific enthalpy h is determined as a function of the three Schutz potentials via the normalization condition $u^a u^b g_{ab} = -1$.

Using the first and second laws of thermodynamics, it can be shown [44] that the perfect isentropic equation of state $p = (\gamma - 1) = p(h, s)$ can be written

$$p = (\gamma - 1)(h/\gamma)^{\gamma/(\gamma-1)} \exp[s/(1-\gamma)]\,. \tag{A.4}$$

The canonical variables are (R, ϕ, θ, s), with conjugate momenta $(p_R, p_\phi, p_\theta, p_s)$ respectively.

However, there are two initial value constraints

$$p_\theta = 0\,, \qquad p_s = \theta p_\phi \tag{A.5}$$

on the fluid variables, which can be used to eliminate the canonical pair (θ, p_θ) from the Lagrangian. The ADM Lagrangian with the pair (θ, p_θ) removed is thus [7, 44]:

$$L_{\mathrm{ADM}} = p_R R' + p_\phi \phi' + p_s s' - N(H_{\mathrm{g}} + H_{\mathrm{m}}) \tag{A.6}$$

where the prime denotes the time derivative,

$$H_{\mathrm{g}} = -[p_R^2/(24R) + 6kR] \tag{A.7}$$

is the purely gravitational super-Hamiltonian, and

$$H_{\mathrm{m}} = N^2R^3[(\rho + p)(u^0)^2 + pg^{00}] = p_\phi^\gamma R^{-3(\gamma-1)}\,\mathrm{e}^s \tag{A.8}$$

is both the coordinate energy density measured by a comoving observer, and the super-Hamiltonian of the matter. The momentum conjugate to the radius of the universe R has a simple expression in terms of R and R':

$$p_R = -12RR'/N\,. \tag{A.9}$$

The constraint equation for the Friedmann universe is obtained by substituting (A.6)–(A.8) into (A.1) and varying the lapse N. The result is the super-Hamiltonian constraint:

$$\mathscr{H} = H_{\mathrm{g}} + H_{\mathrm{m}} = 0 = -p_R^2/(24R) - 6kR + p_\phi^\gamma R^{-3(\gamma-1)}\,\mathrm{e}^s\,. \tag{A.10}$$

In the case of the Friedmann universe, all the information on the time evolution is contained in the super-Hamiltonian constraint (A.10); in fact the ADM analogue of Schrödinger's equation (2.28) is the so-called Wheeler–DeWitt equation:

$$\hat{\mathscr{H}}\psi = (\hat{H}_{\mathrm{g}} + \hat{H}_{\mathrm{m}})\psi = 0\,. \tag{A.11}$$

In the ADM formalism, some combination of the canonical variables is chosen to be p_T, the momentum conjugate to the time T, and a canonical transformation of (A.10) is carried out to a new set of canonical variables of which (T, p_T) is one pair. The ADM Hamiltonian H_{ADM} is then equated to p_T. Quantization is accomplished by treating the time variable on an equal footing with the spatial variables, and replacing p_T with $-\mathrm{i}\,\partial/\partial T$. The trick is to choose the time variable so that (A.11) becomes (2.28) with H being the ADM Hamiltonian. In general – with a poor choice of the time variable – p_T could appear in (A.10) to second order, and then we would be faced with quantizing either a Klein–Gordon equation or a first order equation in which the spatial momenta appear in a square root. Needless to say, the solutions to such an equation would be exceedingly difficult to interpret. In the general case, it may not be possible to find a canonical transformation to a time variable which avoids this problem.

However, in the case of the radiation-filled closed Friedmann universe, there *is* a suitable choice of the time which reduces (A.11) to (2.28), with $H_{\rm ADM}$ being the Hamiltonian for the SHO. With $\gamma = \frac{4}{3}$, $H_{\rm m} = p_\phi^{4/3}\,{\rm e}^s/R$, and thus the natural choice for the momentum conjugate to the time is

$$p_T = p_\phi^{4/3}\,{\rm e}^s \tag{A.12}$$

which involves the matter variables alone. Thus we want the canonical transformation to involve the matter variables alone. If for simplicity we retain the original momentum p as a canonical matter momentum – that is, the canonical transformation for the momenta is $(p_R, p_\phi, p_s) \to (p_R, p_\Phi, p_T)$, where $p_\phi = p_\Phi$ – we obtain from the so-called direct conditions of a canonical transformation the new variables

$$T = -p_s\,{\rm e}^{-s} p_\phi^{-4/3}\,, \qquad R = R\,, \qquad \Phi = \phi + 4p_s/3p \tag{A.13}$$

and the super-Hamiltonian constraint (A.10) takes the form

$$\mathcal{H} = 0 = -p_R^2/24R - 6kR + p_T/R \tag{A.14}$$

which from $H_{\rm ADM} = p_T$ gives

$$H_{\rm ADM} = p_R^2/24 + 6kR^2 \tag{A.15}$$

which is the SHO Hamiltonian.

The lapse N is fixed by solving the Hamilton's equation $T' = 1 = \partial(N[H_{\rm g} + H_{\rm m}])/\partial p_T = N/R$, which gives $N = R$; i.e., the time coordinate T is just conformal time.

With the standard quantization method $p_T \to \hat{p}_T = -{\rm i}\,\partial/\partial T$ and $p_R \to \hat{p}_R = -{\rm i}\,\partial/\partial R$, (and a reversal of the direction of time $T \to -T$) (A.11) and (A.14) give us the Schrödinger equation of the SHO, as I wished to show.

In the ADM approach, in contrast to the approach I developed in the body of this paper, the constraint equation gives us no information about the boundary conditions to be imposed on the wave function of the Universe. However, as I discussed in the body of this paper, Hartle and Hawking [1] have presented a very strong argument that we should regard a quantum universe as coming out of, or exploding from, "nothing". This means that at $T = 0$, the wave function of the Universe was concentrated at the initial singularity: that is, the initial boundary condition is just $\Psi(R = 0, T = 0) = \delta(0)$, where δ is the delta function. Thus with the Hartle–Hawking boundary condition, the entire discussion in section 4 is reproduced in the ADM formalism.

This equivalence depends, of course, on the choice of time parameter in the ADM formalism. If a different time coordinate had been chosen, I would not have obtained the SHO oscillator Hamiltonian; the quantum mechanics will be different for different choices of the time parameter. (This ambiguity is equivalent to the so-called factor ordering problem in quantum mechanics; see [1], [7], [17] and [47] for a discussion.) However, as I remarked at the end of section 4, the MWI strongly suggests that the fundamental Hamiltonian ought not to pick out a length scale which is special to one branch universe. This suggestion can be used in the ADM to rule out certain choices of the time parameter for radiation-filled Friedmann universes.

For example, it rules out Misner's choice of the time parameter in his quantum model of a closed radiation-filled Friedmann universe [48]. Misner expressed the scale factor as $R = {\rm e}^{\Omega}$, and he chose Ω as his time parameter. By (2.16), we have for the matter density $\mu = \Gamma R^{-4} = \Gamma\, e^{-4\Omega}$, where Γ is a constant. It is easy to show [48] that with this matter density and the time parameter Ω, (A.10) becomes

$$-d^2\psi/d\Omega^2 + (e^{-4\Omega} - \Gamma\, e^{-2\Omega})\psi = 0\,. \tag{A.16}$$

The problem with this equation from the MW point of view is that it contains the constant Γ. Classically, this constant is determined in a time-independent manner by measuring the density of radiation at the time of maximum expansion and setting $\mu(t_{max}) = \Gamma R^{-4}(t_{max})$. But this would give a different value of Γ for each classical universe, and thus Γ cannot be the same for all branch universes. But it must be the same for all branch universes by (A.16) and the MWI. Hence (A.16) is inconsistent with the MWI. This is the same scale factor problem I discussed in section 4. For radiation-filled Friedmann universes, the only choice of time parameter that treats the branch universes the same is conformal time, and the choice of conformal time fixes the momentum conjugate to the time to be (A.12), however the scale factor R is parameterized.

It is likely that in the ADM formalism, a canonical transformation involving the scale factor R can be found which transforms (A.11) for any value of the polytropic index γ into a form which treats all branch universes equally. To see that such a canonical transformation might exist, note that by (2.16), the mass density is $\mu = \Gamma R^{-3\gamma}$. If we define a new scale factor y by

$$y = R^{(3\gamma-2)/2} \tag{A.17}$$

then it is easy to show that the super-Hamiltonian constraint (A.10) can be written

$$\mathcal{H} = 0 = -6R^{-3(\gamma-1)}\,[2/(3\gamma-2)]^2\,[(y')^2 + k\,\{(3\gamma-2)/2\}^2\, y^2 - \{(3\gamma-2)/2\}^2\,\Gamma] \tag{A.18}$$

provided the time variable is conformal time ($N = R$); the prime denotes differentiation with respect to conformal time. (I am grateful to Dr. J.D. Barrow for pointing out the coordinate transformation (A.18) to me.) If (A.18) is differentiated with respect to conformal time, then $d\mathcal{H}/dt = 0$ and $\mathcal{H} = 0$ together imply

$$y'' + k[(3\gamma-2)/2]^2\, y = 0 \tag{A.19}$$

which is the SHO if $k = +1$. All the dynamics is contained in (A.19): if the space variable is y and the time variable is conformal time, then the dynamics is that of a SHO. This result strongly suggests that in the ADM formalism it may be possible to reduce the quantum mechanics of a Friedmann universe filled with a perfect fluid of any polytropic index to the quantum mechanics of a SHO. Note also that there is no reference to the troublesome constant Γ in (A.19); in the variable y, all branch universes are on an equal footing. This means – if we accept the MWI – that the appropriate spatial coordinate for quantizing a Friedmann universe with general Γ is not R but y (of course, $y = R$ if, and only if, $\gamma = \frac{4}{3}$). The details of quantizing a Friedmann universe with general polytropic index γ will be discussed elsewhere. As in the case of $\gamma = \frac{4}{3}$, the idea will be to absorb the constant Γ into the momentum conjugate to the time; i.e., into the ADM Hamiltonian.

For models with more than one type of fluid – dust and radiation, say – then the arguments I gave at the end of section 4 show that it will be difficult even in the ADM formalism to find a time parameter which treats the branch universes equally; in particular, those arguments show that conformal time is not an appropriate choice. The existence or non-existence of such a time parameter is a subject for future research. Nevertheless, the reasons which I gave in section 4 for believing that the fundamental Hamiltonian must contain no fundamental length scales also apply to the ADM Hamiltonian: if the MWI is true, then the ADM Hamiltonian must not single out a class of branch universes.