

On Gravity's Role in Quantum State Reduction

Roger Penrose^{1,2}

Received August 22, 1995. Rev. version December 12, 1995

The stability of a quantum superposition of two different stationary mass distributions is examined, where the perturbing effect of each distribution on the space-time structure is taken into account, in accordance with the principles of general relativity. It is argued that the definition of the time-translation operator for the superposed space-times involves an inherent ill-definedness, leading to an essential uncertainty in the energy of the superposed state which, in the Newtonian limit, is proportional to the gravitational self-energy E_{Δ} of the difference between the two mass distributions. This is consistent with a suggested finite lifetime of the order of \hbar/E_{Δ} for the superposed state, in agreement with a certain proposal made by the author for a gravitationally induced spontaneous quantum state reduction, and with closely related earlier suggestions by Diósi and by Ghirardi *et al.*

1. THE PROBLEM OF QUANTUM STATE REDUCTION

The fundamental problem of quantum mechanics, as that theory is presently understood, is to make sense of the *reduction of the state vector* (i.e. *collapse of the wavefunction*), denoted here by R. This issue is usually addressed in terms of the "quantum measurement problem", which is to comprehend how, upon *measurement* of a quantum system, this (seemingly) discontinuous R-process can come about. A measurement, after all, merely consists of the quantum state under consideration becoming entangled with a more extended part of the physical universe, e.g. with a

¹ Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK
² Center for Gravitational Physics and Geometry, Pennsylvania State University, University Park, Pennsylvania 16801, USA

measuring apparatus. This measuring apparatus — together with the observing physicist and their common environment — should, according to conventional understanding, all also have some quantum description. Accordingly, there should be a quantum description of this *entire* quantum state, involving not only the original system under consideration but also the apparatus, physicist, and remaining environment — and this entire state would be expected to evolve continuously, solely according to the Schrödinger equation (unitary evolution), here denoted by the symbol U .

Numerous different attitudes to R have been expressed over many years, ever since quantum mechanics was first clearly formulated. The most influential viewpoint has been the “Copenhagen interpretation” of Niels Bohr, according to which the state vector $|\psi\rangle$ is not to be taken seriously as describing a quantum-level physical reality, but is to be regarded as merely referring to our (maximal) “knowledge” of a physical system, and whose ultimate role is simply to provide us with a means to calculate probabilities when a measurement is performed on the system. That “measurement” would be taken to come about when that system interacts with a *classical* measuring apparatus. Since “our knowledge” of a physical system can undergo discontinuous jumps, there is no reason to be surprised — so it would be argued — when $|\psi\rangle$ undergoes discontinuous jumps also!

Closely related is the environment-entanglement “FAPP” (“for all practical purposes”) point of view (e.g. Ref. 43; cf. Ref. 3), according to which the R -process is taken to be some kind of approximation to the U -evolution of the system together with its environment, and R is viewed as having taken place in the system itself “for all practical purposes” [3]. The essential idea is that the environment involves enormously numerous random degrees of freedom, and these become entangled with the limited number of degrees of freedom in the system itself. Accordingly, the delicate phase relations between the system’s degrees of freedom become irretrievably lost in these entanglements with the environment. It is argued that this effective loss of phase coherence in the system itself gives rise to R , FAPP, although certain additional assumptions are needed in order for this conclusion to be entertained.

There are also viewpoints — referred to as the “many worlds”, or “many minds”, or the Everett interpretation (Ref. 13; cf. Ref. 10) — whereby it is accepted that all the different macroscopic alternatives which (according to U) must remain superposed, actually co-exist *in reality*. It can be convincingly argued that if $|\psi\rangle$ is indeed taken to represent an *actual* reality at all levels, and if $|\psi\rangle$ evolves *precisely* according to U , then something of this nature must hold. However, by itself, and without fur-

ther assumptions, this would offer no explanation as to why the "illusion" of merely one world presents itself to our consciousnesses, nor why the correct quantum-mechanical probabilities come about (Refs. 38,39; cf. also Ref. 36, sec.6.2,6.7).

Other alternative interpretations have been put forward, such as those of de Broglie [9], Bohm [5], Bohm and Hiley [6], Haag [19], Griffiths [18], Omnès [25], and Gell-Mann and Hartle [15], in which the standard quantum procedures are reformulated in a different mathematical framework. The authors of these proposals do not normally take the view that any experimentally testable deviations from standard quantum mechanics can arise within these schemes.

Set against all these are proposals of a different nature, according to which it is argued that present-day quantum mechanics is a limiting case of some more unified scheme, whereby the U and R procedures are *both* to be approximations to some new theory of physical reality. Such a theory would have to provide, as an appropriate limit, something equivalent to a unitarily evolving state vector $|\psi\rangle$. Indeed, many of these schemes, such as that of Pearle [27,28], Bialynicki-Birula and Mycielski [4], Ghirardi, Rimini, and Weber [16], and Weinberg [41], use a quantum-state description $|\psi\rangle$, just as in standard quantum mechanics, but where the evolution of the state deviates by a tiny amount from the precise Schrödinger (or Heisenberg) evolution U. Related to these are proposals which posit that it is in the behaviour of *conscious beings* that deviations from precise U-evolution are to be found (cf. Ref. 42). But in the more "physical" such schemes, the suggested deviations from standard U-evolution become noticeable merely when the system becomes "large", in some appropriate sense. This "largeness" need not refer to physical dimension, but it might, for example, be the number of particles in the system that is relevant (such as in the specific scheme put forward in Ref. 16). In other proposals, it is considered that it is the mass, or mass distribution, that is all important. In schemes of the latter nature, it is normally taken that it is *gravity* that provides the influence that introduces deviations from the standard quantum rules. (See Refs. 20–24,11,17,29,37,30–36, for proposals of this nature.)

In this paper, a new argument is given that explicitly supports a gravitational role in state-vector reduction. Most particularly, it is consistent with a particular criterion, explicitly put forward in [35,36], according to which a macroscopic quantum superposition of two differing mass distributions is *unstable* (analogous to an unstable particle). Accordingly, such a state would decay, after a characteristic lifetime T , into one or the other of the two states. To compute T , we take the *difference* between the two mass distributions under consideration (so that one counts positively and

the other negatively) and compute its gravitational self-energy E_{Δ} . The idea is that we then have

$$T \simeq \hbar/E_{\Delta}.$$

This criterion is also very close to one put forward earlier by Diósi [11], and as modified by Ghirardi, Grassi, and Rimini [17], but their point of view is different from the one used here, proposed by Penrose [34–36]. In particular, in their type of scheme, there are violations of energy conservation which could — in principle or in practice — be experimentally detected. With the type of proposal in mind here, the idea would be that such violations of energy conservation ought to be absent, owing to the fundamental involvement of some of the basic principles of general relativity. (There is a brief discussion of this point in Ref. 36, p.345.) In this article it is indicated that there is a basic conflict between Einstein's general covariance principle and the basic principles of quantum theory, as they relate to stationary states of superposed gravitational fields. It is argued that this conflict can be resolved within the framework of the specific state-reduction proposal of [35,36], according to which such superposed gravitational fields are essentially *unstable*.

It should be made clear, however, that this proposal does not provide a *theory* of quantum state reduction. It merely indicates the level at which deviations from standard linear Schrödinger (unitary) evolution are to be expected owing to gravitational effects. Indeed, it is this author's personal opinion that the correct theory uniting general relativity with quantum mechanics will involve a major change in our physical world-view — of a magnitude at least comparable with that involved in the shift from Newtonian to Einsteinian gravitational physics. The present paper makes no pretensions about even aiming us in the right direction in this regard. Its purpose is a different one, namely to show that even within the framework of completely conventional quantum theory, there is a fundamental issue to be faced, when gravitational effects begin to become important. Standard theory does not provide a clear answer; moreover, it allows room for the type of instability in superposed states that would be consistent with the proposals of Refs. 11,17 and 34–36.

2. STATIONARY STATES

Let us consider the following situation. We suppose that a quantum superposition of two states has been set up, where each individual state has a well-defined static mass distribution, but where the mass distributions differ from one state to the other. For example, we could have a rigid lump of material which we contrive to place in a quantum superposition of two

different locations. Such a superposed state could be achieved by having a photon simultaneously transmitted through and reflected off a half-silvered mirror placed at a distance from the lump, where the transmitted part of the photon's state then triggers a device which slowly moves the lump from its initial location to somewhere nearby, but the reflected part leaves the lump alone. (This is an inanimate version of "Schrödinger's cat".)

If we ignore any gravitational effects, the two alternative locations of the lump will each be stationary states, and will therefore have wavefunctions, say $|\psi\rangle$ and $|\chi\rangle$, that are eigenfunctions of the energy operator H :

$$H|\psi\rangle = v|\psi\rangle, \quad H|\chi\rangle = v|\chi\rangle.$$

Here, the energy eigenvalue v is the same in each case, because the mass is simply displaced from one location to the other. (We ignore the energy of the photon in this, the lump itself being supposed to have enormously greater mass-energy, and also any energy involvement in the device that moves the lump.) It is clear from this that there is a complete degeneracy for linear superpositions of $|\psi\rangle$ and $|\chi\rangle$, and any linear combination

$$\lambda|\psi\rangle + \mu|\chi\rangle$$

(with λ and μ complex constants) will also be an eigenstate of H with the same eigenvalue v . Each such linear combination will also be a stationary state, with the same energy eigenvalue v . Thus, any one of these combinations would be just as stable as the original two, and must therefore persist unchanged for all time.

We shall shortly consider the delicate theoretical issues that arise when one considers the (superposed) *gravitational fields* of each of the two instances of the lump also to be involved in the superposition. In accordance with this, it will be appropriate to be rather careful about the kinds of issues that become relevant when quantum theory and general relativity are considered together.

3. PRELIMINARY CONSIDERATIONS

In particular, there is a point of subtlety which needs to be addressed, even before we consider the details of any gravitational effects. We take note of the fact that, in the absence of any spatial inhomogeneity in the background potentials (gravitational or otherwise), there is nothing in the intrinsic nature of one lump location that allows us to distinguish it from any other lump location. Thus, we might choose to adopt the standpoint

that there is really no physical difference at all between the various states of location of the lump and, accordingly, take the view that all these seemingly different states of the lump are actually all the same state! Indeed, this would be the normal standpoint of (quantum) general relativity. Each space-time geometry arising from of each separate lump location would be identical. The principle of general covariance forbids us to assign a meaningful (coordinate) label to each individual point, and there is no coordinate-independent way of saying that the lump occupies a different location in each of the two configurations under consideration.

However, this is not the standard attitude in ordinary quantum mechanics. If the various position states of a single quantum particle were considered to be all the same, then we would not be able to construct all the many different wave functions for that particle which are, after all, simply superpositions of states in which the particle occupies different locations. In standard quantum mechanics, all the different particle locations correspond to different quantum states whose superpositions can be independently involved.

It is worth while, in this context, to consider how standard quantum mechanics treats *composite* particles. Here one may choose to factor out by this translational freedom. For example, the wavefunction of a hydrogen atom, in an eigenstate of the energy and angular momentum operators, is still completely degenerate with respect to the spatial location of its centre of mass. In considering the quantum mechanics of a hydrogen atom, the normal procedure is (implicitly or otherwise) to factor out by the degrees of freedom that specify the location of the mass centre (or, as a simplifying approximation, merely the location of its proton), and consider the quantum mechanics only of what remains: essentially the location of the orbiting electron, this being what is fixed by the energy and angular momentum.

In the present context, with the two superposed lump locations, we are trying to examine precisely the translational degrees of freedom that had been "factored out" in the case of the hydrogen atom. In fact there is something of a physical inconsistency in simply considering our lump to be displaced from one location to another in the two states under superposition. There is a conservation law which requires that the mass centre remain fixed in space (or move uniformly in a straight line). This would have been violated unless the lump displacement is compensated by the displacement of some other massive object in the opposite direction. We shall consider that there is such another object, and that it is enormously more massive than is the lump itself. Let us call this other object "the Earth". To move the lump from one location to the other, we simply allow

the Earth to move by a very tiny amount, so as to allow the mass centre to remain fixed; and since the Earth is so very much more massive than the lump, we can consider that in practice the Earth does not move at all.

The presence of the Earth in these considerations allows us to circumvent the problem that we had previously considered. For the Earth (assumed to have some large but static finite irregular shape) serves to establish a "frame of reference" against which the motion of the lump can be considered to be taking place. The space-time geometries corresponding to all the various lump locations are now all different from one another, and "general covariance" does not prevent us from considering these various states to be all distinct.

At this point, another complication appears to arise because the Earth's gravitational field has to be taken into account. In fact, we shall not be much concerned with the Earth's field. It will be the differences in the gravitational fields in the various *lump* locations that will have essential relevance. In fact, there is no problem, in the considerations which follow, in allowing the Earth's gravitational field to be involved also. But we may prefer to avoid the complication of the Earth's field, if we choose, by supposing that our "experiment" is to be set up within a spherical cavity situated at the centre of the Earth. Then the Earth's actual field could be eliminated completely. Yet the Earth would still serve to establish a "frame of reference" against which the motion of the lump could be considered to be taking place.

On the other hand, we shall prefer to consider that the lumps are actually sitting on the surface of the Earth. The Earth's (effectively) constant gravitational field does not have any significant influence on the considerations of relevance here. There is no problem about having the lump in a stationary state on the surface of the Earth, provided that some appropriate upward forces are introduced on the lump, to allow it to remain at rest with respect to the Earth.

If we were to consider a situation in which the two superposed lump positions are at different heights from the ground, then we would have to consider the effects of the energy expended in raising the lump in the Earth's field. This would introduce an additional factor, but would not seriously affect the situation that we are concerned with here. If desired, all these considerations can be evaded if we regard our lump movements to be taking place entirely within our spherical cavity at the centre of the Earth. However, it will not be necessary to pass to this extreme idealization for the considerations of this paper.

4. SUPERPOSED GRAVITATIONAL FIELDS

Let us now try to consider how the gravitational field of the lump itself affects our superposed state. Each lump location is accompanied by the static gravitational field produced by the lump in that location. We must envisage that the superposed state is now an entangled one,

$$\lambda|\psi\rangle|G_\psi\rangle + \mu|\chi\rangle|G_\chi\rangle,$$

where $|G_\psi\rangle$ and $|G_\chi\rangle$ are the quantum states of the gravitational fields of the lump locations corresponding to $|\psi\rangle$ and $|\chi\rangle$, respectively. We might choose to think of $|G_\psi\rangle$ and $|G_\chi\rangle$ as *coherent states*, if that is the appropriate description. In any case, whatever is to be meant by the quantum state of a stationary gravitational field — including all the internal degrees of freedom of the field — would be supposed to be incorporated into $|G_\psi\rangle$ or $|G_\chi\rangle$. But it does not greatly matter, for our present purposes, what the “correct” quantum-mechanical descriptions of macroscopic gravitational fields actually are. We must suppose, however, that whatever this description is, it closely accords in its physical interpretation, for a *single* unsuperposed lump, with the classical gravitational field of that lump according to the descriptions of Einstein’s general relativity.

In fact, the “entangled” nature of the superposed state will not be of importance for us here. We simply take the two states under superposition to be

$$|\Psi\rangle = |\psi\rangle|G_\psi\rangle \quad \text{and} \quad |X\rangle = |\chi\rangle|G_\chi\rangle.$$

The essential point is that each of the two states concerned must involve a reasonably well-defined (stationary) space-time geometry, where these two space-time geometries differ significantly from each other. We must now raise the question: is this *superposed* state still a stationary state?

We have to consider carefully what a “stationary state” means in a context such as this. In a stationary space-time, we have a well-defined concept of “stationary” for a quantum state in that background, because there is a Killing vector T in the space-time that generates the time-translations. Regarding T as a differential operator (the “ $\partial/\partial t$ ” for the space-time), we simply ask for the quantum states that are eigenstates of T , and these will be the stationary states, i.e. states with well-defined energy values. In each case the energy value of the state in question, $|\Psi\rangle$, would be (essentially) the eigenvalue E_Ψ of T corresponding to that state:

$$T|\Psi\rangle = -i\hbar E_\Psi|\Psi\rangle.$$

However, for the superposed state we are considering here we have a serious problem. For we do not now have a specific space-time, but a superposition of two slightly differing space-times. How are we to regard such a "superposition of space-times"? Is there an operator that we can use to describe "time-translation" in such a superposed space-time? Such an operator would be needed so that we can identify the "stationary states" as its eigenvectors, these being the states with definite energy. It will be shown that there is a fundamental difficulty with these concepts, and that the notion of time-translation operator is *essentially* ill defined. Moreover, it will be possible to define a clear-cut measure of the *degree* of this ill-definedness for such a superposed state. Accordingly, there is, in particular, an essential uncertainty or "fuzziness" in the very concept of energy for such a state, and the degree of this uncertainty can be estimated in a clear-cut way. This is consistent with the view that such a superposed state is *unstable*, and the lifetime of the state will be given by \hbar divided by this measure of energy uncertainty, in accordance with the way in which Heisenberg's uncertainty principle is employed in the theory of unstable particles, where the particle's lifetime is related to its mass-energy uncertainty.

How are we to regard a quantum superposition of two space-times? It is not sufficient to take a completely formal attitude to such matters, as is common in discussions of quantum gravity. According to the sorts of procedure that are often adopted in quantum gravity, the superposition of different space-times is indeed treated in a very formal way, in terms of complex functions on the space of 3-geometries (or 4-geometries), for example, where there is no pretence at a pointwise identification of the different geometries under superposition. A difficulty with such formal procedures arises, however, if we attempt to discuss the physics that takes place *within* such a formal superposition of spaces, as is the case with the type of situation under consideration here (cf. also Ref. 1).

Indeed, in the case of the two minutely differing space-times that occur in our situation with the two superposed lump locations, there would be no obvious way to register the fact that the lump is actually in a different place, in each of the two configurations under superposition, unless there is some sort of (approximate) identification. (Of course, in this identification, since the lump itself is supposed to be "moved", corresponding points of the lump are not identified, but the corresponding points of the Earth are, since the Earth is not considered to have significantly moved.)

Nevertheless, it is clear from the principles of general relativity that it is not appropriate, in general, to make a precise identification between points of one space-time and corresponding points of the other. The *gauge*

freedom of general relativity — as reflected in “the principle of general covariance” (or, equivalently, “diffeomorphism invariance”) — is precisely the freedom that forbids a meaningful precise labelling of individual points in a space-time. Accordingly there is generally no precise meaningful pointwise identification between different space-times. (In special cases, it may be possible to circumvent this problem; see Refs. 1,2, where issues of this nature are examined, and interesting effects are anticipated, in relation, in particular, to the quantization of a cosmic string.) In the general case, all that we can expect will be some kind of *approximate* pointwise identification.

5. THE SEMICLASSICAL APPROACH

One possible way to address this difficulty might be to adopt a “semi-classical approximation” to quantum gravity, according to which it would be the *expectation value* of the (quantum) energy-momentum tensor that serves to specify the right-hand side of Einstein’s equations

$$R_{ab} - \frac{R}{2} g_{ab} = -8\pi G T_{ab}.$$

In such an approximation the superposition of our pair of states in which one lump is in two spatially displaced locations would have a gravitational field which is merely the average field of the fields of each lump individually.

In fact, the semiclassical approximation is not really physically consistent. In particular, it allows superluminary communication (see Ref. 29); and, in a certain interpretation, it is grossly inconsistent with observation [26]. These difficulties might be avoided if the semiclassical interpretation can be combined with some scheme of gravitationally induced state-vector reduction, as is argued by Pearle and Squires (Ref. 29; cf. also Ref. 23).

From the point of view of the present paper, there is an additional difficulty with the semiclassical approximation. The semiclassical description would provide a space-time containing the mass distributions of two spatially displaced lumps with vacuum ($R_{ab} = 0$) between them. This gravitational field — of a pair of spatially displaced lumps — is not really the same as a linear superposition of the two fields, each describing one of the two lump locations individually. The gravitational interaction effects between the pair of lumps would have to be taken into account, according to the non-linear effects of general relativity. Assume that the actual lump, in its two displaced locations, sits on a smooth horizontal table (“the Earth”). The “semiclassical” state would represent a pair of spatially separated massive lumps, and would not actually possess an

exact Killing vector which could play the role of T ($=\partial/\partial t$) because the "semiclassical" lumps would fall towards each other along the table in accordance with their gravitational attractions, leading to a non-static space-time geometry.

For reasons such as these, it is proposed not to follow a semiclassical description here. Instead, we shall follow the route of supposing that there is some kind of *approximate sense* in which the two superposed space-times can be pointwise identified. Without some kind of (approximate) identification between the space-times, we do not seem even to be able to express the fact that the quantum state of the lump is a superposition of two distinct locations, in which the degree of displacement between the two lump locations is reasonably well defined.

6. APPROXIMATE SPACE-TIME POINT IDENTIFICATION

The basic principles of general relativity — as encompassed in the term "the principle of general covariance" (and also "principle of equivalence") — tell us that there is no natural way to identify the points of one space-time with corresponding points of another. Consider our quantum superposition between two different space-times (here, the fields of two alternative lump locations). If we are to attempt to make a pointwise identification between these two space-times, we can do this in a way that would be only *approximately* meaningful. Let us try to obtain some measure of this degree of approximation.

In order to proceed to a reasonably explicit expression of this measure, it will be helpful to make the assumption that a Newtonian approximation to the gravitational fields of each lump location is adequate. Indeed, in any plausible practical situation in which a quantum superposition of lump locations might have to be considered, this would certainly be the case. Thus, there will be spatial sections of the two space-times in question which are Euclidean 3-spaces, and each space-time will possess a well-defined time coordinate whose constant values define these Euclidean spatial sections. We shall suppose that the time coordinates for each space-time can be naturally identified with each other, so we have a single time parameter t common to the two space-times. This would seem reasonable for the Newtonian situation under consideration, although there would be essential subtleties arising when the two gravitational fields are treated according to full general relativity.

It should be made clear, however, that passing to the Newtonian limit does not remove the difficulties that the principle of general covariance — or the principle of equivalence — presents in relation to the quantum

superposition of gravitational fields. As may be recalled, there is a space-time formulation of Newtonian gravitational theory originally provided by Cartan [7], and further studied independently by Friedrichs [14] and Trautman [40], in which the Newtonian version of the equivalence principle is directly incorporated into the geometrical description. (See Ref. 12 for an up-to-date account.) The Cartan geometric formulation of Newtonian gravitational theory is indeed the appropriate one for taking into account the subtleties of (what remains of) the principle of general covariance and the principle of equivalence. (For example, the Newton-Cartan space-time of a *constant* non-zero Newtonian gravitational field has an identical geometry to that of the zero gravitational field, but it differs from a non-uniform Newtonian field which produces tidal effects.) It has been pointed out by Christian [8] that the Newton-Cartan framework provides a valuable setting for exploring some of the fundamental problems of unifying quantum theory with gravitational theory without, at this stage, the more severe difficulties of general relativity having to be faced. Christian argues that this framework indeed sheds important light on the role of gravity in the measurement problem.

As it turns out, the criterion for quantum state reduction that we shall be led to here is independent of the value of the speed of light c . It will thus have a well-defined non-trivial Newtonian limit, and it can be expressed within the Newton-Cartan framework. However, although it will be valuable to bear this framework in mind here, we shall not actually use its specific mathematical details.

The essential point about superposing a pair of Newton-Cartan space-times is that whereas we are allowing that the time-coordinate t can be identified in the two space-times — and so there is a canonical correspondence between the various space sections of one space-time with those of the other — there is no canonical way of identifying the individual *points* of a section of one space-time with corresponding points of the other. It is this lack of a definite pointwise identification between the spatial sections of the two space-times which will lead us to an essential ill-definedness of the notion of time-translation — and therefore of the notion of stationarity — for the quantum-superposed state.

The reader might wonder why it is *time-translation* that should encounter problems, when, in our Newtonian context, there is no problem with the time-coordinate t . However, this is how it should be; for “time-translation” is something which is represented by the operator “ $\partial/\partial t$ ”, and the meaning of this operator is really concerned more with the choice of the *remaining* variables x, y, z , (those parameters to be held fixed in the definition of $\partial/\partial t$) than it is with t itself. An uncertainty in the *pointwise*

identification of the spatial sections of one space-time with the spatial sections of the other will indeed show up in an uncertainty in the definition of time-translation. Suppose that we use the coordinates x, y, z , and t for the points of one space-time, and x', y', z' , and t' for the points of the other (where, of course, $t = t'$); then we have

$$\begin{aligned}\frac{\partial}{\partial t'} &= \frac{\partial}{\partial t} + \frac{\partial x}{\partial t'} \frac{\partial}{\partial x} + \frac{\partial y}{\partial t'} \frac{\partial}{\partial y} + \frac{\partial z}{\partial t'} \frac{\partial}{\partial z} \\ &= \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\end{aligned}$$

where \mathbf{v} is the spatial velocity, as described with respect to the unprimed coordinate system, of a point fixed in the primed coordinate system. (In the (x, y, z, t) -system, the 3-vector \mathbf{v} has components $(\partial x/\partial t', \partial y/\partial t', \partial z/\partial t')$, i.e. $(-\partial x'/\partial t, -\partial y'/\partial t, -\partial z'/\partial t)$, and the operator ∇ has components $(\partial/\partial x, \partial/\partial y, \partial/\partial z)$.) If we were to attempt to identify the point with coordinates (x, y, z, t) in one of the space-times under consideration, with the point with coordinates (x', y', z', t') in the other, then we should encounter an incompatibility between their notions of time-evolution unless \mathbf{v} vanishes everywhere.

Of course, in this Newtonian limit, it is possible to arrange that $\mathbf{v} = 0$ everywhere, simply by taking ordinary static, non-rotating Newtonian/Cartesian coordinates for the two space-times, related by a constant spatial displacement between them. But this would be to go against the spirit of what is entailed, in the present context, by Einstein's principle of general covariance. It is the special nature of the Newtonian limit that provides us with flat Euclidean spatial sections whose local motions are determined directly by what happens at infinity. In this Newtonian limit, the "coordinate freedom" can indeed be eliminated if we nail things down at infinity, because in Newton-Cartan geometry, there is exact spatial rigidity. However, within the more general context of curved (and not necessarily stationary) space-times, this is not at all appropriate. Our physical expectations would be for the criteria characterizing the nature of a localized quantum superposition to depend on reasonably local criteria. The principles of general relativity are antagonistic towards the idea of identifying individual local points in a precise way, in terms of the situation at infinity.

Similar remarks also apply to certain proposals within the general programme of quantizing general relativity proper, according to which the time-translation operator would not be taken to have local relevance, but

to refer merely to the symmetries at spatial infinity — it being supposed that only asymptotically flat space-times are to be considered in quantum superposition. It seems, however, for situations such as those under consideration here where we are concerned with a reasonably local problem, that it is quite inappropriate to define the notion of time-translation merely in terms of what is going on at spatial infinity. We do not know, after all, what the geometry of the actual universe is on a very large scale (spatially asymptotically spherical or hyperbolic, for example, either of which would lead to different notions of “time-translation” from that which is normally considered). It is most unlikely that, for a reasonably local quantum problem such as this, Nature should really “care” what is going on at infinity.

We shall suppose instead that, in the particular Newtonian quantum superposition under consideration, it is appropriate to demand an approximate spatial identification between the two space-times, where the degree of approximation is governed by local (or quasi-local) considerations in the vicinity of the region where the identification occurs. In accordance with the principle of equivalence, it is the notion of free fall which is locally defined, so the most natural local identification between a local region of one space-time and a corresponding local region of the other would be that in which the free falls (i.e. space-time geodesics) agree. However, in the superposition under consideration, there is no way to make the spatial identifications so that the free falls agree everywhere throughout the space-times. The best that one can do is to try to minimize the amount of the *difference* between free fall motions.

How are we to express this difference mathematically? (Let us assume that the two quantum amplitudes assigned to the two superposed states are about the same size. Then we may take it that neither space-time's geometry dominates the other.) Let \mathbf{f} and \mathbf{f}' be the acceleration 3-vectors of the free-fall motions in the respective space-times, at some identified point (where the accelerations can be taken with respect to the appropriate local identified coordinates). In fact, \mathbf{f} and \mathbf{f}' will be the Newtonian gravitational force-per-unit-test-mass, at that point, in each space-time. Let us take the scalar quantity

$$(\mathbf{f} - \mathbf{f}')^2 = (\mathbf{f} - \mathbf{f}') \cdot (\mathbf{f} - \mathbf{f}')$$

as the measure of incompatibility of the identification — and, accordingly, of the “uncertainty” involved in this identification. It may be noted that if we apply an *acceleration* to the common coordinates to the two space-times, while keeping the actual point identification unaltered, then the difference $\mathbf{f} - \mathbf{f}'$ does not change. In fact, the expression $(\mathbf{f} - \mathbf{f}')^2$

is coordinate-independent (assuming that the "dot" \cdot refers to the actual 3-dimensional spatial metric).

For the total measure of incompatibility (or "uncertainty") Δ at a particular instant, given by a particular t -value, it is proposed to integrate this quantity (with respect to the spatial 3-volume element d^3x) over the $t = \text{constant}$ 3-space:

$$\begin{aligned}\Delta &= \int (\mathbf{f} - \mathbf{f}')^2 d^3x \\ &= \int (\nabla\Phi - \nabla\Phi')^2 d^3x \\ &= \int (\nabla\Phi - \nabla\Phi') \cdot (\nabla\Phi - \nabla\Phi') d^3x \\ &= - \int (\Phi - \Phi') (\nabla^2\Phi - \nabla^2\Phi') d^3x\end{aligned}$$

where $\mathbf{f} = -\nabla\Phi$ and $\mathbf{f}' = -\nabla\Phi'$, Φ and Φ' being the respective gravitational potentials for the two space-times. By Poisson's formula

$$\nabla^2\Phi = -4\pi G\rho,$$

we obtain

$$\Delta = 4\pi G \int (\Phi - \Phi') (\rho - \rho') d^3x,$$

where ρ and ρ' are the respective mass densities. Using the integral formula

$$\Phi(x) = - \int \rho(y) / |x - y| d^3y,$$

we get

$$\Delta = -4\pi G \iint (\rho(x) - \rho'(x)) (\rho(y) - \rho'(y)) / |x - y| d^3x d^3y,$$

which is basically just the gravitational self-energy of the *difference* between the mass distributions of each of the two lump locations.

How does this relate to the uncertainty in the time-translation Killing vector referred to at the beginning of this section? That there should be some relationship follows from general considerations, but the exact form that this relationship should take seems to depend upon the specific model that is used to describe the uncertainty of space-time identification. (As

long ago as 1966, Károlyházy provided some fairly closely related considerations, cf. Ref. 20 and also Refs. 21,22.) To understand what is involved in the present context, we consider three successive time slices, given by $t = \tau$, $t = \tau + \Delta t$, and $t = \tau + 2\Delta t$. We imagine that there is some uncertainty in the identification between the two space-times under superposition at each of these three times. The measure of relative “acceleration uncertainty” that we have just been considering, and which seems to be forced upon us from consideration of the principles of equivalence and general covariance, has to do with the difference between the error in spatial identification at $t = \tau + \Delta t$ and the average of those at $t = \tau$ and at $t = \tau + 2\Delta t$. The overall uncertainty in the quantity $\mathbf{v} \cdot \nabla$, which appears as the uncertainty in the time-translation Killing vector for the superposed space-times is a “velocity uncertainty”, and it has to do with the difference between the error in spatial identification at $t = \tau + \Delta t$ and that at $t = \tau$ (or else between the error at $t = \tau + 2\Delta t$ and that at $t = \tau + \Delta t$). The precise relationship between the “acceleration uncertainty” and “velocity uncertainty” seems to depend upon the way that this uncertainty is actually modelled. The direct “position uncertainty” does not feature in these considerations; it is the time-evolution of the error in spatial identification that has importance for us here.

A further ingredient of possible importance is an error in the identification in the actual time coordinate t for the two space-times. We have ignored this complication here, but it clearly has relevance in full general relativity. Perhaps it should also be taken into account in the Newtonian limit.

In view of these various complicating issues, no attempt will be made to formulate a definitive statement of the precise measure of uncertainty that is to be assigned to the “superposed Killing vector” and to the corresponding notion of “stationarity” for the superposed space-time. However, it is strongly indicated by the above considerations that the quantity Δ , as defined above, gives a very plausible (though provisional) estimate of this uncertainty. Hence, we can use Δ , or some simple multiple of this quantity, as a measure of the fundamental energy uncertainty “ E_Δ ” of the superposed state. Accordingly, the superposed state would not be exactly stationary, but it is consistent with the above considerations that it should have a lifetime of the general order of \hbar/E_Δ .

It is reassuring that basically the same expression is seen to arise from certain other considerations. For example, it was pointed out in [35,36] that the gravitational self-energy involved in a quantum superposition of a pair of differing space-time geometries should involve an essential uncertainty, owing to the fact that even in classical general relativity there is

difficulty with the energy concept for gravity. (There is no local expression for gravitational energy.) Moreover, considerations of the symplectic structure of linearized gravity [34] lead to something very similar, and so also does the earlier model of Diósi [11], although in these papers the suggestion is that the expression

$$G \iint \rho(x)\rho'(y) / |x - y| d^3x d^3y,$$

which is the *gravitational interaction energy*, should be the relevant measure of the required "energy uncertainty" E_Δ . In fact (apart from the factor 4π), it does not make any difference which expression is used, provided that — as is the case here — the two individual states under superposition each have the same gravitational self-energy. (A situation for which the difference in the expressions might be important would be in a cloud chamber, where the two states in superposition might be "droplet forming" and "droplet not forming". In this case the gravitational self-energy of the droplet would contribute a difference between the two expressions.)

7. FURTHER CONSIDERATIONS

It is clear that the above arguments provide only a very preliminary analysis of the difficulties involved in the notion of "stationarity" for a quantum superposition of differing stationary space-times. However, even just these preliminary considerations seem to suggest that such a superposition would be unstable, and that a very plausible expression for the order of magnitude of the lifetime of such a superposed state (where the two relevant amplitudes are roughly equal) is \hbar/E_Δ , where E_Δ is the gravitational self-energy of the *difference* between the two mass distributions involved.

In any case, it should be clear from what has been said above that *conventional quantum theory* provides no clear answer (in the absence of a satisfactory theory of quantum gravity) to the problem of the stability of a quantum superposition of two different gravitating states. We do not need to appeal to the contentious conceptual issues inherent in the measurement problem of quantum mechanics for some motivation for believing that such superpositions should be unstable.

In the above considerations, we have restricted attention to cases where the superposition involves just *two* different states, each of which individually has a well-defined space-time geometry, and where the two amplitudes assigned to each of these two constituent states are of about the same size. This does not tell us, for example, what to do about the

wavefunction of a single isolated proton, as it spreads throughout space in accordance with the Schrödinger equation. In the GRW scheme [16] there would be a small probability for this state to reduce spontaneously to a more localized state. The present considerations, however, are not adequate to provide us with an expectation as to whether (or how frequently) such a spontaneous state reduction should occur, nor do they tell us what kind of state should be the result of such a spontaneous state reduction. We might think that the “natural” states for the proton, having “well-defined space-time geometries”, are those for which the proton is reasonably localized, say to a region of roughly its Compton wavelength. But if we adopt such a view, we are driven to consider that most wavefunctions are superpositions of a great many of these natural states, and with widely differing amplitudes. A more detailed theory is clearly needed if such questions are to be addressed adequately.

Despite these uncertainties, it is still possible, in many different circumstances, to estimate the expected order of magnitude of the rate of gravitationally induced state-vector reduction according to this scheme. For a single proton, we may expect that a superposed state of two separated spatial locations will decay to one or the other location in something of the order of a few million years. For a water speck 10^{-5} cm in radius, the timescale would be about an hour or so; for a speck 10^{-3} cm in radius, something like a millionth of a second. These results indeed seem reasonable, and if confirmed would supply a very plausible solution to the quantum measurement problem, but for the moment they appear to be rather beyond what can be experimentally tested.

It should be emphasized that none of the considerations of the present paper give any clear indication of the mathematical nature of the theory that would be required to incorporate a plausible gravitationally induced spontaneous state-vector reduction. In all probability, such mathematical considerations would have to come from quite other directions. Indeed, this author's own expectations are that no fully satisfactory theory will be forthcoming until there is a revolution in the description of quantum phenomena that is of as great a magnitude as that which Einstein introduced (in the description of gravitational phenomena) with his general theory of relativity.

ACKNOWLEDGEMENTS

I am grateful to Abhay Ashtekar, Jeeva Anandan, Joy Christian, Ted Newman, and Lee Smolin and others for some very helpful discussions.

I am also grateful to NSF for support under research contract PHY 93-96246.

REFERENCES

1. Anandan, J. (1994). *Gen. Rel. Grav.* **26**, 125.
2. Anandan, J. (1995). "Gravitational phase operator and cosmic strings". To appear in *Phys. Rev. D*.
3. Bell, J. S. (1990). *Physics World* **3**, 33.
4. Bialynicki-Birula, I., and Mycielski, J. (1976). *Ann. Phys. (NY)* **100**, 62.
5. Bohm, D. (1952). *Phys. Rev.* **85**, 166.
6. Bohm, D., and Hiley, B. (1994). *The Undivided Universe* (Routledge, London).
7. Cartan, É. (1923,1924). "Sur les variétés à connexion affine et la théorie de la relativité généralisée", *Ann. École Norm. Sup.* **40**, 325, **41**, 1.
8. Christian, J. (1995). "Definite events in Newton-Cartan quantum gravity" Oxford University preprint.
9. de Broglie, L. (1956). *Tentative d'Interpretation Causale et Nonlineaire de la Mécanique Ondulatoire* (Gauthier-Villars, Paris).
10. DeWitt, B. S., and Graham, R. D. (eds.) (1973). *The Many Worlds Interpretation of Quantum Mechanics* (Princeton University Press, Princeton).
11. Diósi, L. (1989). *Phys. Rev.* **A40**, 1165.
12. Ehlers, J. (1991). In *Classical Mechanics and Relativity: Relationship and Consistency (Int. Conf. in memory of Carlo Cataneo, Elba 1989)*, G. Ferrarese, ed. (Monographs and Textbooks in Physical Science, Lecture Notes 20, Bibliopolis, Naples).
13. Everett, H. (1957). *Rev. Mod. Phys.* **29**, 454.
14. Friedrichs, K. (1927). *Math. Ann.* **98**, 566.
15. Gell-Mann, M., and Hartle, J. B. (1993). *Phys. Rev.* **D47**, 3345.
16. Ghirardi, G. C., Rimini, A. and Weber, T. (1986). *Phys. Rev.* **D34**, 470.
17. Ghirardi, G. C., Grassi, R., and Rimini, A. (1990). *Phys. Rev.* **A42**, 1057.
18. Griffiths, R. (1984). *J. Stat. Phys.* **36**, 219.
19. Haag, R. (1992). *Local Quantum Physics: Fields, Particles, Algebras* (Springer-Verlag, Berlin).
20. Károlyházy, F. (1966). *Nuovo Cimento* **A42**, 390.
21. Károlyházy, F. (1974). *Magyar Fizikai Polyoirat* **12**, 24.
22. Károlyházy, F., Frenkel, A., and Lukács, B. (1986). In *Quantum Concepts in Space and Time*, R. Penrose and C. J. Isham, eds. (Oxford University Press, Oxford), p.109.
23. Kibble, T. W. B. (1981). In *Quantum Gravity 2: a Second Oxford Symposium*, C. J. Isham, R. Penrose and D. W. Sciama, eds. (Oxford University Press, Oxford), p.63.
24. Komar, A. B. (1969). *Int. J. Theor. Phys.* **2**, 157.
25. Omnès, R. (1992). *Rev. Mod. Phys.* **64**, 339.
26. Page, D. N., and Geilker, C. D. (1981). *Phys. Rev. Lett.* **47**, 979.
27. Pearle, P. (1985). In *Quantum Concepts in Space and Time*, R. Penrose and C. J. Isham, eds. (Oxford University Press, Oxford), p.84.
28. Pearle, P. (1989). *Phys. Rev.* **A39**, 2277.
29. Pearle, P., and Squires, E. J. (1995). "Gravity, energy conservation and parameter values in collapse models". Durham University preprint DTP/95/13.
30. Penrose, R. (1981). In *Quantum Gravity 2: A Second Oxford Symposium*, C. J. Isham, R. Penrose and D. W. Sciama, eds. (Oxford University Press, Oxford), p.244.

31. Penrose, R. (1986). In *Quantum Concepts in Space and Time*, R. Penrose and C. J. Isham, eds. (Oxford University Press, Oxford), p.129.
32. Penrose, R. (1987). In *300 Years of Gravity*, S. W. Hawking and W. Israel, eds. (Cambridge University Press, Cambridge), p.17.
33. Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford University Press, Oxford).
34. Penrose, R. (1993). In *General Relativity and Gravitation 13. Part 1: Plenary Lectures*, R. J. Gleiser, C. N. Kozameh and O. M. Moreschi, eds. (IOPP, Bristol/Philadelphia), p.179.
35. Penrose, R. (1994). In *Fundamental Aspects of Quantum Theory*, J. Anandan and J. L. Safko, eds. (World Scientific, Singapore), p.238.
36. Penrose, R. (1994). *Shadows of the Mind; An Approach to the Missing Science of Consciousness* (Oxford University Press, Oxford).
37. Percival, I. C. (1995). *Proc. R. Soc. Lond.* **A451**, 503.
38. Squires, E. (1990). *Phys. Lett.* **A145**, 67.
39. Squires, E. (1992). *Found. Phys. Lett.* **5**, 279.
40. Trautman, A. (1965). In *Lectures on General Relativity (Brandeis 1964 Summer Inst. on Theoretical Physics)*, vol. I, by A. Trautman, F. A. E. Pirani and H. Bondi (Prentice-Hall, Englewood Cliffs, NJ), p.7.
41. Weinberg, S. (1989). *Phys. Rev. Lett.* **62**, 485.
42. Wigner E. P. (1961). In *The Scientist Speculates*, I. J. Good, ed. (Heinemann, London); reprinted in E. Wigner (1967). *Symmetries and Reflections* (Indiana University Press, Bloomington) and in 1983 in *Quantum Theory and Measurement*, J. A. Wheeler and W. H. Zurek, eds. (Princeton University Press, Princeton).
43. Zurek, W. H. (1991). *Physics Today* **44**, 36.