

Black holes, information, and decoherenceStephen D. H. Hsu^{*} and David Reeb[†]*Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403, USA*

(Received 21 March 2009; published 24 June 2009)

We investigate the experimental capabilities required to test whether black holes destroy information. We show that an experiment capable of illuminating the information puzzle must necessarily be able to detect or manipulate macroscopic superpositions (i.e., Everett branches). Hence, it could also address the fundamental question of decoherence versus wave function collapse.

DOI: [10.1103/PhysRevD.79.124037](https://doi.org/10.1103/PhysRevD.79.124037)

PACS numbers: 04.70.Dy, 03.65.Ta, 03.65.Ud, 03.65.Yz

I. INTRODUCTION: BLACK HOLE INFORMATION

In 1976 Hawking proposed that black holes destroy information: pure states which collapse to form black holes evaporate into mixed states described by density matrices [1]. The argument in favor of information destruction can be pared to a few essential components; for reviews, see [2–6]. Hawking radiation, into which the hole evaporates, originates from outside the horizon and is causally disconnected from the interior: a spacelike slice can be constructed which intersects both the infalling matter and the outgoing Hawking radiation. The no-cloning theorem [7] in quantum mechanics prevents information from residing in two places on the same slice, so the outgoing radiation must be independent of the initial state.

It is safe to say that, over 30 years after Hawking's paper, theoreticians remain divided as to whether Hawking was originally correct, or whether some locality-violating mechanism somehow allows the information to escape.

In this paper we investigate the following questions: Is the black hole information puzzle simply philosophy, or is it subject to experimental test? If the latter, then what capabilities are required? We find connections to a different question, from the foundations of quantum mechanics: do wave functions collapse, or is quantum evolution strictly unitary, leading inevitably to macroscopic superposition states? An experiment which sheds light on the black hole information puzzle would also be capable of addressing fundamental issues in quantum mechanics. See Zeh [8] and also [9] for related discussion.

To highlight the importance of macroscopic superpositions, we emphasize that the very formulation of the information puzzle relies on the use of the semiclassical black hole spacetime—e.g., in the construction of the spacelike slice used in the no-cloning argument, or in the original Hawking calculation. But, in any fully quantum mechanical treatment of the black hole formation process there exist branches of the wave function, possibly of very small amplitude, in which no apparent horizon is formed and the initial state particles all escape to infinity. In other words,

in which the spacetime is radically different. This is most easily seen if one considers black holes formed in the collision of two particles [10]: there is always a nonzero amplitude for no scattering—i.e., no black hole creation—in which the particles simply pass each other. This remains the case even as the number of particles in the initial state becomes large, although for certain semiclassical initial data one can make the no-formation amplitude arbitrarily small. Nevertheless, the information puzzle cannot avoid the issue of macroscopic superpositions: could such small amplitude branches restore purity or unitarity? [11]

II. DECOHERENCE: PURE TO MIXED EVOLUTION

Quantum mechanics as conventionally formulated (the Copenhagen interpretation) allows for two kinds of time evolution: the usual Schrödinger evolution, which is unitary, and measurement collapse, which is nonunitary and leads to von Neumann projection onto a particular eigenstate of the operator associated with the measurement. It is appealing to think that wave function collapse might only be an apparent phenomenon, which results from unitary evolution. This idea dates to Everett [12], but has been developed substantially in recent decades as the theory of decoherence [13].

Consider a system prepared in a superposition state

$$|\psi\rangle = c_1|1\rangle + c_2|2\rangle. \quad (1)$$

Suppose that, due to interactions between the system and its environment (or, equivalently, a measuring apparatus), the two evolve into an entangled state

$$|\psi\rangle \otimes |E\rangle \rightarrow |\Psi\rangle = c_1|1\rangle \otimes |E_1\rangle + c_2|2\rangle \otimes |E_2\rangle. \quad (2)$$

The states $E_{1,2}$ are referred to as pointer states of the environment or measuring device. These pointer states are determined by the dynamics—that is, the 1, 2 bases along which the device makes measurements is determined by its specific properties. A Stern-Gerlach machine measures spin along a particular axis given by its magnetic field; it will not decohere the system into spin states along any other directions. After the measurement interaction, the state of the system is encoded redundantly in the

^{*}hsu@uoregon.edu

[†]dreeb@uoregon.edu

environment states $E_{1,2}$ and can be read out by making simple measurements on subsets of the degrees of freedom—i.e., did the red light flash (look for red photons), or did the green light flash?

The environment is assumed to have a large number of degrees of freedom, so that after a relatively short time (determined by the specific dynamics) the states E_1 and E_2 are nearly orthogonal. The dimensionality of a Hilbert space describing N degrees of freedom is exponential in N : for qubits, $d = 2^N$. Two randomly chosen vectors from this space will have overlap $\langle E_1|E_2 \rangle \sim 1/d$, which is tremendously small for any macroscopic environment or measuring device (e.g., $N \sim$ Avogadro’s number). Consequently, interference phenomena between the two “branches” of system plus environment are highly suppressed (see below).

Absent the capability to measure the environmental degrees of freedom, it is appropriate to trace over the degrees of freedom of E :

$$\hat{\rho} = \text{Tr}_E |\Psi\rangle\langle\Psi| = |c_1|^2 |1\rangle\langle 1| + |c_2|^2 |2\rangle\langle 2|; \quad (3)$$

this density matrix, from which the outcome of all subsequent measurements on the system alone can be predicted just as well as from the knowledge of the complete state $|\Psi\rangle$, is diagonal if one neglects the exponentially small overlap $\langle E_1|E_2 \rangle$. This process then has the appearance of measurement with fundamental wave function collapse and probabilistic outcomes, despite purely unitary Schrödinger evolution. For All Practical Purposes—FAPP, as formulated by Bell [14]—decoherence leads to the usual Copenhagen interpretation.

But a nagging issue remains—the presence of the other branches in the pure state $|\Psi\rangle$. Are they real? Can they ever be detected experimentally? The off-diagonal elements of the reduced density matrix are suppressed by the small overlap of typical environmental states E_1, E_2 , but can their effects be measured?

Omnès [15] made detailed estimates of the capabilities required to distinguish between decoherence and fundamental collapse, which we will examine in the following section. He found that for macroscopic objects, e.g., containing Avogadro’s number of degrees of freedom, any beyond-FAPP device would have to be larger than the visible universe. Hence, Omnès asserted, any distinction between the Copenhagen interpretation and unitary wave function evolution (leading to Everett branches) for macroscopic objects is untestable, and beyond the realm of scientific inquiry.

The foregoing discussion, in particular, Omnès’ estimate, assumes big environments (e.g., $N \sim$ Avogadro’s number). However, decoherence mechanisms are also at work if the number of degrees of freedom N of the environment is smaller, or if the interaction between system and environment is weaker; in these cases, however, decoherence effects are either not as strong (the off-

diagonal elements in $\hat{\rho}$ cannot be completely neglected as in (3), i.e., FAPP does not hold in this case), or happen only over time scales longer than in the strongly-interacting case, respectively. In recent years, this gradual onset of decoherence in controlled environments (also dubbed the “Quantum-to-Classical transition”) has been the focus of quite a number of laboratory experiments. For example, in [16] the gradual loss of spatial coherence (interference pattern) of fullerene molecules in a slit experiment was observed with increasing pressure of the gas, i.e., with increasing interaction strength (cross section) between fullerenes and gas molecules (environment). Other experiments, e.g. [17], have verified the gradual loss of coherence in a system (superposition of two states of a Rydberg atom) with increasing number of degrees of freedom of the interacting environment ($N \sim 10$ photons in a cavity). On the other hand, one of the big challenges in achieving useful quantum computing [18] is to build and control large and scalable quantum systems ($N \sim 100$ or more) in which coherence is maintained (possibly via quantum error correction) over the time of the computation.

In the decoherence approach to measurement an initially pure state is later described by a reduced density matrix which, FAPP, represents a mixed state. The black hole information puzzle is often described in similar terms: infalling matter in a pure state is somehow transformed into a mixed state of Hawking radiation. Or, equivalently, if the quantum information in the black hole precursor is not to be found in the outgoing radiation, the radiation is surmised to be in a mixed state. It has been claimed that pure to mixed evolution implies, necessarily, catastrophic consequences, such as violation of energy conservation [19] (see [20] for additional arguments, for and against this point of view). Potential resolutions of the puzzle in which the information ends up somewhere outside our universe (e.g., involving baby universes or spacetime topology change [21]) must still have an effective description in our universe in terms of pure to mixed state evolution. Decoherence provides an example of *effective* pure to mixed evolution without catastrophic consequences. Tracing over the environmental states, one loses track of the total energy of the system, but without any resulting catastrophe.

III. BLACK HOLE INFORMATION EXPERIMENTS

Below we describe three categories of experiments which test different aspects of theoretical ideas about black hole information. Despite their differences, all require the ability to detect or manipulate macroscopic superpositions.

A. Test unitarity

The strictest experiment one can imagine is a check of purity, linearity (superposition) and unitarity. That is,

- (i) Are pure initial states $|a\rangle$ mapped to pure final states $|b\rangle$?
- (ii) Is linearity preserved: $\alpha|a\rangle + |a'\rangle \rightarrow \alpha|b\rangle + |b'\rangle$?
- (iii) Is the mapping unitary: $|b\rangle = U|a\rangle$ with $UU^\dagger = \mathbb{1}$?

To test all of these properties the experimenter must be able to create and measure all initial and final states $|a\rangle$ and $|b\rangle$, including macroscopic superposition states. To test (i), one must, at minimum, be able to distinguish pure from mixed states. This is explored further in point 2 below. To test (ii), linearity for all states, one has to build and detect macroscopic superpositions. Finally, to test (iii) one has to measure all matrix elements of U , including those connecting macroscopic superpositions: even though unitarity is a stronger requirement than linearity, measuring the matrix elements of U when already assuming linearity is not quite as difficult as verifying linearity for all states in the first place, although (for most dynamics) one at least has to either prepare initial or measure final macroscopic superpositions in this case also.

In [22] it is proposed that a sufficiently precise energy measurement, combined with measurements of (many) other operators that do not commute with the energy, would determine the state of the black hole if one can ignore degeneracies (i.e., the energy eigenvalue uniquely determines the eigenstate). The energy precision would have to be finer than the level spacing, which is exponentially small in the number of degrees of freedom. In principle, such measurements might be used to check (i)–(iii), although there are obvious challenges, both theoretical and experimental. Clearly, though, such capabilities could also be used on systems other than black holes to detect macroscopic superpositions of decoherent branches by their tiny energy shifts.

B. Test purity vs decohered mixed state

A basic goal would be to differentiate between pure and mixed states of the type produced by decoherence. Without this minimal capability one can hardly investigate whether pure states evolve to mixed states, as proposed originally by Hawking. (Was the initial state pure? Is the final state pure or mixed?)

In the black hole context, one could imagine forming the hole from an initial state with at least some degrees of freedom in a superposition (e.g., two spin states). The remaining degrees of freedom can be considered the environment E from our earlier discussion, assuming the dynamics are such that the environment evolves into two different pointer states corresponding to the superposition. This would be the case if, for example, the two spin states had slightly different energy due to a magnetic field provided by the other degrees of freedom.

If, for each $i = 1, 2$, the initial state $|i\rangle$ of the system leads to the final state $|\Psi_i\rangle = |i\rangle \otimes |E_i\rangle$ for system plus environment, then, starting from the initial state $c_1|1\rangle +$

$c_2|2\rangle$, two candidates for the final state would be: on the one hand a pure state superposition

$$|\Psi\rangle = c_1|\Psi_1\rangle + c_2|\Psi_2\rangle \quad (4)$$

as predicted by unitary evolution, corresponding to the pure density matrix

$$\hat{\rho}_P = |\Psi\rangle\langle\Psi|, \quad (5)$$

and, on the other hand, a mixed state density matrix

$$\hat{\rho}_M = |c_1|^2|\Psi_1\rangle\langle\Psi_1| + |c_2|^2|\Psi_2\rangle\langle\Psi_2|, \quad (6)$$

predicted by wave function collapse. What is required to differentiate between these alternatives? Consider a measurement operator M , which, without loss of generality, we might take to be a projector onto some subspace of the Hilbert space. The pure and mixed states can be distinguished if we can detect the difference

$$\text{Tr}[\hat{\rho}_P M] - \text{Tr}[\hat{\rho}_M M] = \langle\Psi_1|M|\Psi_2\rangle + \langle\Psi_2|M|\Psi_1\rangle. \quad (7)$$

Let us consider two possibilities. If M is a generic operator—for example, only couples to a limited subset of the degrees of freedom—then the matrix elements in (7) will be exponentially small. In particular,

$$\langle\Psi_1|M|\Psi_2\rangle \sim \frac{1}{\dim\mathcal{H}_{S\otimes E}} \sim \exp(-N), \quad (8)$$

where N is the number of degrees of freedom of system plus environment. For a black hole, N is of order its entropy, or area in Planck units. Omnès [15] has argued that when N is of order Avogadro's number (e.g., for 1 g of ordinary matter), a measurement of this accuracy is impossible *in principle*. Based on this, Omnès concludes that questions about decoherent branches other than the one observed as an outcome are not scientific. In effect, his calculations purport to extend Bell's FAPP to a statement of principle. Roughly speaking, he argues that the sensitivity of a classical device with N' degrees of freedom only improves as a power of N' (not exponentially with N'). Since the precision needed to detect a decohered macroscopic superposition involving N degrees is $\sim \exp(-N)$, as in (8), the required N' grows exponentially with N . To detect a macroscopic superposition with $N \sim 10^{24}$, Omnès concludes, would require N' larger than the number of particles in the visible universe.

A concrete proposal [23] has been made of how to experimentally decide between unitary (pure to pure) and nonunitary (pure to mixed) evaporation of a black hole (in AdS space), namely, by measuring whether a certain correlation function drops, over time, either to zero or to the finite value $\exp(-N)$. Interestingly, the required measurement accuracy is similar to that necessary to detect Everett branches of a system with a similar number N of degrees of freedom, as described in the previous paragraph.

On the other hand, for a carefully engineered operator M , the amplitude $\langle \Psi_1 | M | \Psi_2 \rangle$ in (7) can be of order $1/2$, even if $\langle \Psi_1 | \Psi_2 \rangle = 0$. Indeed, the maximum value is obtained when M is a projector onto a macroscopic superposition of the type $\sim |\Psi_1\rangle + |\Psi_2\rangle$, like $|\Psi\rangle$ itself. It is a formidable challenge to perform a measurement of such an operator M , presumably even harder than preparing $|\Psi\rangle$ itself, and we give a concrete, albeit idealized, example below in the context of the Coleman-Hepp model of measurement [24] to illustrate the difficulties arising even in simple cases.

Coleman and Hepp proposed an explicit model of a quantum measurement which results from the interaction of spin states. In their model the interaction between the system (itself a spin) and measuring device causes evolution as in (2), with

$$\begin{aligned} |\Psi_1\rangle &= |1\rangle \otimes |E_1\rangle = |+\rangle \otimes |+\dots+\rangle \\ |\Psi_2\rangle &= |2\rangle \otimes |E_2\rangle = |-\rangle \otimes |-\dots-\rangle, \end{aligned} \quad (9)$$

where \pm are spin-up and spin-down states along the z axis, and the measuring device has N degrees of freedom (qubits). That is, the interaction between system and measuring device leads to a correlation between the initial spin state and the (macroscopic, when N is large) state of the device. The state of the spin can be read out by measuring some subset of the N degrees of freedom in the device.

In this context it is straightforward to design an operator M for which $\langle \Psi_1 | M | \Psi_2 \rangle$ is large: we simply take the tensor product operator

$$M = \bigotimes_i \sigma_x^i, \quad (10)$$

where $\sigma_x^i |\pm\rangle = |\mp\rangle$ measures the spin of environmental qubit i in the x -direction. An experimental realization of M would be able to distinguish the two considered post-measurement states of the system plus device: pure $\hat{\rho}_P$ vs mixed $\hat{\rho}_M$. Note, though, that the simple M considered in this toy example is not a projection (in particular, not the projector into the one-dimensional subspace $|\Psi\rangle$) and so cannot, e.g., distinguish the considered macroscopic superposition pure state (4) from the (seemingly simpler) pure state $|+\rangle \otimes |+\dots+\rangle$ in which all spins are aligned in the $+x$ -direction.

One might object that in this simple example the pointer states $E_{1,2}$ of the device in (9) are mutually orthogonal, but not thermalized. In this sense the model does not represent a realistic measurement (the “environment” is highly constrained). This could easily be remedied by allowing some interactions between the spins in the device, which leads to some (presumably) ergodic but unitary evolution. If one keeps the spins isolated from the rest of the universe, these interactions evolve $\Psi_{1,2}$ into something more random, at least in appearance:

$$|\Psi\rangle \rightarrow |\Psi'\rangle = U(c_1|\Psi_1\rangle + c_2|\Psi_2\rangle) \quad (11)$$

or, for the proposed mixed state (6) after the measurement,

$$\hat{\rho}_M \rightarrow \hat{\rho}'_M = U \hat{\rho}_M U^\dagger. \quad (12)$$

It would still be the case that the operator $M' = U M U^\dagger$ can distinguish between pure and mixed states of the post-measurement device. However, if the N spins are separated in space (e.g., correspond to isolated qubits in a quantum device), then the operator M' would itself have to be realized out of macroscopic superpositions of spatially separated objects, unlike the original M which acted independently on each of the spins.

To summarize, a measurement which can differentiate between pure and mixed states either has to rely on extreme precision to detect very small matrix elements as in (8), or on the ability to prepare a very special, typically nonlocal, operator like M' .

C. Test Hawking mixed state vs typical pure state

In this scenario we compare Hawking radiation in a mixed state ρ_* to radiation in a pure state ψ .

It is widely believed that large scale violation of locality at the black hole horizon is necessary for unitary evolution. This might lead to significant deviations from Hawking’s results describing what is emitted from the hole. If such deviations were observed, they would undermine the usual semiclassical reasoning which leads to the information puzzle, although deviations of the radiation from Hawking’s mixed state description do not by themselves imply that evolution is unitary or purity-conserving.

Perhaps a more plausible scenario, assumed in what follows, is that the radiation, although described by a pure state ψ , only deviates in subtle ways from the Hawking mixed state. That is, information is encoded in correlations (phases or superpositions) between particle states in the radiation, but the overall distribution appears to be thermal and the temperature evolution is as predicted by Hawking.

It is extremely difficult to differentiate between a pure state ψ of this type and the Hawking mixed state ρ_* . Local measurements on the radiation (i.e., over length scales much smaller than its full spatial extent) can only exclude tiny subsets of the ψ Hilbert space. Moreover, it can be shown that for almost all states ψ these local measurements are governed by the same (thermal) probability distribution as the one obtained from ρ_* .

A simple way to understand this is to recall that maximizing the entropy of a system subject to an energy constraint leads to a thermal distribution. Pure states which conform, at least macroscopically, to the Hawking predictions are constrained to describe the same total energy emission over any particular interval of time. To be specific, consider a time interval Δt_i over which the Hawking temperature is close to constant, $T = T_i$, but during which many particles are emitted. Let the Hawking calculation predict that a total amount of energy \mathcal{E}_i be emitted during

the interval. Ordinary statistical mechanics tells us that the overwhelming majority of states (of the system in the volume corresponding to Δt_i) with energies close to \mathcal{E}_i will be approximately thermal—i.e., maximizing the entropy leads to a Boltzmann distribution for energy occupation numbers, with temperature equal to the Hawking temperature T_i . The probability distribution governing measurements of the energy distribution of individual emitted quanta over the interval Δt_i will then coincide with that given by ρ_* , except for an exponentially rare subset of states satisfying the constraint (i.e., configurations with much lower entropies than the Boltzmannian, or thermal, ones).

A more explicit computation follows. Consider the subset of pure states ψ which conform to the Hawking predictions governing the amount of energy radiated in each time interval Δt_i . Specifically, require that, for every i , the reduction (by taking the partial trace) of ψ to the degrees of freedom emitted in the interval Δt_i be a mixture of superpositions of the energy eigenstates (of the theory in that volume) with eigenvalues close to \mathcal{E}_i . That is, consider region i and only the degrees of freedom within it (neglecting boundary effects, which are negligible for large regions). The reduction of ψ to these degrees of freedom, when expanded in an energy eigenstate basis for the region Δt_i , must have support only on states with energies close to \mathcal{E}_i . The superposition of two pure states ψ_1 and ψ_2 satisfying this condition will also satisfy the condition as superposition does not extend the region of support; therefore the condition defines a subspace of the larger Hilbert space. Denote by \mathcal{H}_R this restricted (“energetically conforming”) subspace of the overall radiation Hilbert space \mathcal{H} .

Further, divide the radiation into a subsystem S , to be measured, and the remaining degrees of freedom which constitute an environment E , so $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$ and

$$\rho_S \equiv \rho_S(\psi) = \text{Tr}_E |\psi\rangle\langle\psi| \quad (13)$$

is the density matrix which governs measurements on S for a given pure state ψ . Note the assumption that these measurements are local to S , hence the trace over E .

Then, a recent theorem [25] on entangled states, which exploits properties of Hilbert spaces of very high dimension, shows that for almost all $\psi \in \mathcal{H}_R$, $\rho_S(\psi) \approx \text{Tr}_E(\rho_*)$ as long as $d_E \gg d_S$, where $d_{E,S}$ are the dimensionalities of the \mathcal{H}_E and \mathcal{H}_S Hilbert spaces. In the theorem, $\rho_* = \mathbb{1}_R/d_R$ is the equiprobable mixed state on the restricted Hilbert space \mathcal{H}_R ($\mathbb{1}_R$ is the identity projection on \mathcal{H}_R and d_R the dimensionality of \mathcal{H}_R), so $\text{Tr}_E(\rho_*)$ is the corresponding canonical state of the subsystem S . In other words, ρ_* describes a perfectly thermalized radiation system with temperature profile equal to that of Hawking radiation from an evaporating black hole.

To state the theorem in [25] more precisely, the (measurement-theoretic) notion of the *trace-norm* is re-

quired, which can be used to characterize the distance between two mixed states ρ_S and Ω_S :

$$\|\rho_S - \Omega_S\|_1 \equiv \text{Tr} \sqrt{(\rho_S - \Omega_S)^2}. \quad (14)$$

This sensibly quantifies how easily the two states can be distinguished by measurements, according to the identity

$$\|\rho_S - \Omega_S\|_1 = \sup_{\|O\| \leq 1} \text{Tr}(\rho_S O - \Omega_S O), \quad (15)$$

where the supremum runs over all observables O with operator norm $\|O\|$ smaller than 1 (projectors $P = O$ are in some sense the best observables, all other observables can be composed out of them, and they have $\|P\| = 1$). Note that the trace on the right-hand side of (15) is the difference of the observable averages $\langle O \rangle$ evaluated on the two states ρ_S and Ω_S , and therefore specifies the experimental accuracy necessary to distinguish these states in measurements of O . A special form of the theorem then states that the probability that

$$\|\rho_S(\psi) - \text{Tr}_E(\rho_*)\|_1 \geq d_R^{-1/3} + \sqrt{\frac{d_S^2}{d_R}} \quad (16)$$

is less than $2 \exp(-d_R^{1/3}/18\pi^3)$. In words: let ψ be chosen randomly (according to the natural Hilbert space measure) out of the space of allowed states \mathcal{H}_R ; the probability that a measurement on the subsystem S only, with measurement accuracy of $d_R^{-1/3} + \sqrt{d_S^2/d_R}$, will be able to tell the pure state ψ (of the entire system) apart from the mixed state ρ_* is exponentially small ($\sim \exp(-d_R^{1/3})$) in the dimension of the space \mathcal{H}_R of allowed states. That is, the overwhelming majority of energetically conforming pure Hawking evaporation states $\psi \in \mathcal{H}_R$ cannot be distinguished from Hawking’s predicted density matrix ρ_* by measurements on a small subsystem S of the radiation, even if the experimental error in measurements of projectors is only $d_R^{-1/3} + \sqrt{d_S^2/d_R}$. This is an incredible precision, considering the estimates $d_R \sim \exp S_{\text{BH}} \sim \exp M^2$ and $d_S \sim \exp S_S$, where the entropy $S_S \sim VT^3 \ll S_{\text{BH}}$ of the (energetically conforming) subsystem S can be computed from its volume V and the Hawking temperature T of the particle excitation distribution inside.

Thus, as long as individual measurements are localized in spacetime, so that S is small relative to E , one cannot distinguish a typical state $\psi \in \mathcal{H}_R$ from ρ_* without exponential sensitivity in the measurement on S . This is true even if one performs many measurements on distinct subsystems S_i , in particular, even if the union of S_i covers all of the radiation. This is because, even if each subsystem were in a pure state that could in principle be measured exactly (as opposed to merely measuring all the single-particle excitations in it separately), at the very least the phase relations between the states of the different subsystems S_i are lost. Only complete measurements (including

phase relations) on very big subsystems S (\sim half of the degrees of freedom of E) have a noninfinitesimal chance of distinguishing ψ from ρ_* .

In analogy to what we learned in the previous cases, measurements that can distinguish generic $\psi \in \mathcal{H}_R$ from ρ_* with reasonable probability, or without exquisite sensitivity, must be highly nonlocal, covering a spacetime region which includes most of the radiation emitted by the black hole. They must, in a sense, measure it all “at once”. But measurements of this sophistication could, again, also differentiate between macroscopic superpositions and mixed states.

IV. CONCLUSIONS

Black hole information experiments are at least as hard as experiments which test decoherence vs fundamental collapse. One has to create a semiclassical black hole (in fact, many of them in *identical* states) and then make very challenging measurements on the relativistic decay products, which include gravitons. (Note, it appears difficult to determine the quantum state of gravitons with physically realizable detectors that obey the positive energy condition [26].) In particular, the experiment must be sensitive to the phase relations in coherent superpositions of Fock states rather than simply counting occupation numbers as ordinary particle detectors do. By comparison, the most accessible tests of decoherence would be in the context of an artificial toy system like the Coleman-Hepp model, or other controlled quantum computing environment.

To be more precise, let us define three classes of experiments as follows, with N the number of degrees of freedom.

B: experiments on black holes of the type 1, 2 or 3 described in Sec. III above. Note the black hole needs to be semiclassical in order to satisfy the assumptions of the black hole puzzle, i.e. $S \sim N \gg 1$. On the other hand, if the black hole is too large, its lifetime will be much greater than the age of the universe. Some mechanism for creating small black holes is required [10].

Q1: experiments which test whether ordinary macroscopic systems with $N \sim 10^{24}$ (e.g., a gram of dust) undergo fundamental wave function collapse, or decoherence.

Q2: experiments of type Q1, but on small, controlled systems such as quantum computers or related devices. Here $N \gg 1$ but presumably $N \ll 10^{24}$.

One might argue that the difficulty of B is greater than that of Q1 or Q2 on the grounds that a black hole will evaporate into relativistic degrees of freedom (including gravitons!) that are hard to control, and that one has to make the black hole in the first place. On the other hand one can address B with $N \ll 10^{24}$ (e.g. with a semiclassical black hole $N \sim 1000$), so in an alternate sense Q1 is more difficult:

$$B \cong Q1.$$

Whereas, B is certainly more difficult than Q2:

$$B > Q2.$$

Therefore, if fundamental questions about measurement, decoherence and wave function collapse are philosophy rather than science—i.e., cannot be tested by experiments—then so is the black hole information puzzle.

Our results can be summarized rather simply. Hawking suggested black holes cause pure states to evolve to mixed states. But, for all practical purposes (FAPP), decoherence does the same thing (or at least appears to). In order to test Hawking’s proposal one therefore has to go beyond FAPP and beyond decoherence. Such capability allows fundamental tests of quantum measurement.

ACKNOWLEDGMENTS

The authors thank S. Hossenfelder, C. Kiefer, J. Polchinski, L. Smolin and H. D. Zeh for useful comments. This work was supported by the Department of Energy under DE-FG02-96ER40969.

-
- [1] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).
 - [2] A. Strominger, arXiv:hep-th/9501071.
 - [3] T. Banks, Nucl. Phys. B, Proc. Suppl. **41**, 21 (1995).
 - [4] S. B. Giddings, arXiv:hep-th/9412138.
 - [5] T. Jacobson, arXiv:gr-qc/9908031.
 - [6] L. Susskind and J. Lindesay, *An Introduction to Black Holes, Information and the String Theory Revolution: The Holographic Universe* (World Scientific, Hackensack, USA, 2005).
 - [7] D. Dieks, Phys. Lett. A **92**, 271 (1982); W. K. Wootters and W. H. Zurek, Nature (London) **299**, 802 (1982); H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. Lett. **76**, 2818 (1996).
 - [8] H. D. Zeh, Phys. Lett. A **347**, 1 (2005).
 - [9] C. Kiefer, Classical Quantum Gravity **18**, L151 (2001); Lect. Notes Phys. **633**, 84 (2003); H. Nikolic, arXiv:0805.2555.
 - [10] D. M. Eardley and S. B. Giddings, Phys. Rev. D **66**, 044011 (2002); S. D. H. Hsu, Phys. Lett. B **555**, 92 (2003).
 - [11] In analogy with ordinary scattering theory we could write the quantum state at future infinity as $\psi' = \psi_0 + \psi_1$ where ψ_0 is the future-evolved unscattered state (i.e., no

- black hole formed; initial state particles can propagate to infinity) and ψ_1 the Hawking radiation state (black hole was formed and evaporated). For suitable initial data (e.g., a large ball of dust) one can ensure that the magnitude of ψ_0 is arbitrarily small (a black hole is almost certainly formed) and that ψ_1 is largely independent of the internal state of the hole (the no-cloning argument applies to the semiclassical black hole spacetime). This suggests a unitarity problem: if varying the internal state does not change ψ_1 , then many distinct states are mapped to almost the same final state $\psi' = \psi_0 + \psi_1 \approx \psi_1$. This argument does not, however, require the final state to be a mixture.
- [12] H. Everett, III, *Rev. Mod. Phys.* **29**, 454 (1957); “The Theory of the Universal Wave Function,” (Ph.D. thesis), reprinted in *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton Series in Physics, edited by B. S. De Witt and R.N. Graham (Princeton University Press, Princeton, USA, 1973).
- [13] H.D. Zeh, *Found. Phys.* **1**, 69 (1970); R. Omnès, *Rev. Mod. Phys.* **64**, 339 (1992); W. H. Zurek, *Rev. Mod. Phys.* **75**, 715 (2003); E. Joos, H. D. Zeh, C. Kiefer, D. Giulini, J. Kupsch, and I.-O. Stamatescu, *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, New York, USA, 2003), 2nd ed..
- [14] J. S. Bell, *Helv. Phys. Acta* **48**, 93 (1975); *Phys. World* **3**, 33 (1990).
- [15] R. Omnès, *The Interpretation of Quantum Mechanics*, Princeton Series in Physics (Princeton University Press, Princeton, USA, 1994).
- [16] K. Hornberger, S. Uttenthaler, B. Brezger, L. Hackermüller, M. Arndt, and A. Zeilinger, *Phys. Rev. Lett.* **90**, 160401 (2003).
- [17] M. Brune, E. Hagley, J. Dreyer, X. Maitre, A. Maali, C. Wunderlich, J.M. Raimond, and S. Haroche, *Phys. Rev. Lett.* **77**, 4887 (1996); P. Bertet, S. Osnaghi, A. Rauschenbeutel, G. Nogues, A. Auffeves, M. Brune, J.M. Raimond, and S. Haroche, *Nature (London)* **411**, 166 (2001).
- [18] R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467 (1982); D.P. DiVincenzo, *Fortschr. Phys.* **48**, 771 (2000).
- [19] T. Banks, M.E. Peskin, and L. Susskind, *Nucl. Phys.* **B244**, 125 (1984).
- [20] M. Srednicki, *Nucl. Phys.* **B410**, 143 (1993); S.B. Giddings, *Phys. Rev. D* **49**, 4078 (1994); W.G. Unruh and R.M. Wald, *Phys. Rev. D* **52**, 2176 (1995); J. Oppenheim and B. Reznik, arXiv:0902.2361.
- [21] S.D.H. Hsu, *Phys. Lett. B* **644**, 67 (2007); S. Hossenfelder and L. Smolin, arXiv:0901.3156.
- [22] V. Balasubramanian, D. Marolf, and M. Rozali, *Gen. Relativ. Gravit.* **38**, 1529 (2006); *Int. J. Mod. Phys. D* **15**, 2285 (2006).
- [23] J. M. Maldacena, *J. High Energy Phys.* 04 (2003) 021.
- [24] K. Hepp, *Helv. Phys. Acta* **45**, 237 (1972).
- [25] S. Popescu, A. J. Short, and A. Winter, *Nature Phys.* **2**, 754 (2006); arXiv:quant-ph/0511225.
- [26] L. Smolin, *Gen. Relativ. Gravit.* **17**, 417 (1985); **16**, 205 (1984).