Almost Certain Escape from Black Holes in Final State Projection Models

Seth Lloyd*

MIT Mechanical Engineering, Cambridge, Massachusetts 02139, USA

(Received 27 June 2004; published 14 February 2006)

Recent models of the black-hole final state suggest that quantum information can escape from a black hole by a process akin to teleportation. These models rely on a controversial process called final-state projection. This Letter discusses the self-consistency of the final-state projection hypothesis and investigates escape from black holes for arbitrary final states and for generic interactions between matter and Hawking radiation. Quantum information escapes with fidelity $\approx (8/3\pi)^2$: only half a bit of quantum information is lost on average, independent of the number of bits that escape from the hole.

DOI: 10.1103/PhysRevLett.96.061302 PACS numbers: 04.70.Dy, 03.67.Hk

If information escapes as black holes evaporate, then black holes could function as quantum computers [1–3]. Recently, Horowitz and Maldacena proposed a model of black-hole evaporation that imposes a final-state boundary condition at the black-hole singularity [4]. The resulting projection onto a final state gives rise to a nonlinear time evolution for the quantum states in and outside of the black hole, which permits quantum information to escape from the black hole by a process akin to teleportation. The Horowitz-Maldacena model requires a specific final state that is perfectly entangled between the matter that formed the black hole and the incoming Hawking radiation. Whether or not quantum gravity supports such a final state remains to be seen. In addition, even with the proper final state, interactions between the incoming Hawking radiation and the collapsing matter can spoil the unitary nature of the black-hole evaporation [5], destroying some or all of the quantum information inside the hole [6,7].

The purpose of this Letter is to examine the robustness of the escape of quantum information during black-hole evaporation in final-state projection models. Final-state projection is a nonlinear quantum mechanical process [8–16] that has been studied extensively since its introduction in 1964 [17–21]. While controversial, final-state projection is apparently physically self-consistent. If final-state projection actually takes place at final singularities, this Letter shows that for projection onto *any* final state at the singularity (independent of the exact details of quantum gravity) and for *almost all* interactions between the matter and incoming Hawking radiation, properly encoded classical information escapes from the hole with certainty. Of the quantum information that escapes from the hole, only one-half a qubit is lost on average, regardless of the number of bits of quantum information in the hole to begin with. More precisely, the state of the matter that formed the hole is preserved under black-hole evaporation with a fidelity of $f \approx (8/3\pi)^2 \approx 0.85$. Individual quantum bits escape with a fidelity that approaches one as the number of bits in the hole becomes large.

The Horowitz-Maldacena model for black-hole evaporation.—The Horowitz-Maldacena (HM) model is described in [3,4]. Black holes evaporate by absorbing negative-energy "incoming" Hawking radiation and by emitting positive-energy "outgoing" Hawking radiation. Let the dimension of the Hilbert space for the collapsing matter inside the black hole be N. In the ordinary semiclassical treatment of black-hole evaporation, the incoming and outgoing Hawking radiation emitted at some time is in an entangled state that yields a canonical ensemble for the outgoing radiation. Horowitz and Maldacena assume a less familiar "microcanonical" form for the Hawking radiation:

$$|\phi\rangle_{\text{in}\otimes\text{out}} = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} |j\rangle_{\text{in}} \otimes |j\rangle_{\text{out}}, \tag{1}$$

where $\{|j\rangle_{\text{in}}\}$ is an orthonormal basis for the Hilbert space H_{in} of the incoming Hawking radiation and $\{|j\rangle_{\text{out}}\}$ is an orthonormal basis for the Hilbert space H_{out} of the outgoing Hawking radiation. The states $|j\rangle_{\text{in}}$ and $|j\rangle_{\text{out}}$ represent multiparticle states with energy equal to the mass of the hole. Because of the equivalence of the canonical and microcanonical ensembles in statistical mechanics [22], both the canonical and microcanonical forms for the Hawking radiation allow almost all information to escape from the hole.

Let $|\phi\rangle_{\text{matter}\otimes\text{in}} \in H_{\text{matter}} \otimes H_{\text{in}}$ be the final state onto which the collapsing matter together with the incoming Hawking radiation is projected at the singularity. Horowitz and Maldacena postulated a form for this state of

$$|\phi\rangle_{\text{matter}\otimes\text{in}} = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} (S|k\rangle_{\text{matter}}) \otimes |k\rangle_{\text{in}}, \tag{2}$$

where S is a unitary transformation acting on the matter states alone. The usual analysis of quantum teleportation shows that for states of this form, the transformation from the state of the collapsing matter to the state of the outgoing Hawking radiation is

$$T = {}_{\text{matter}\otimes\text{in}}\langle\phi|\phi\rangle_{\text{in}\otimes\text{out}} = S/N. \tag{3}$$

Note that in Eq. (3), only the *in* parts of the two states are contracted with each other. Equation (3) reflects the fact that, when one takes a state entangled between two Hilbert spaces, *in* and *out*, and contracts the *in* part together with the *in* part of a state entangled between the *in* Hilbert space and a third Hilbert space, *matter,* the result is a transformation between the *matter* and *out* spaces. The factor $1/N$ reflects the fact that if this were conventional teleportation, then this particular final state would occur only with probability $1/N^2$. In final-state projection, however, only one final state can occur: accordingly, the final transformation from collapsing matter to outgoing Hawking radiation is renormalized, and the net result is the unitary transformation S. In the HM model, final-state projection leads to a unitary transformation between collapsing matter and the outgoing Hawking radiation.

Final-state projection.—Final-state projection is an intrinsically nonlinear process and shares the virtues and vices of other proposals for nonlinear quantum mechanical processes. Escape of quantum information from black holes via final-state projection is similar to the use of nonlinear quantum mechanics to provide superluminal communication as described (and rejected) in [8–10], to violate the second law of thermodynamics [11], or to solve NP-complete problems [12]. Such nonlinear quantum effects have been investigated experimentally under non-Planckian conditions and ruled out to a high degree of accuracy [13–16] (Refs. [6,7] propose similar tests of such nonlinear quantum effects in a "normal" environment).

Final-state projection was introduced by Aharonov, Bergmann, and Lebowitz to restore time symmetry in quantum mechanics [17]. The process of projecting onto a final state is well formulated in Griffiths' consistent histories approach to quantum mechanics [18]. (See also the work of Schulman on the reversal of the thermodynamic arrow of time in classical and quantum systems [19].) The initial investigations of final-state projection in a cosmological context was performed by Gell-Mann and Hartle [20], using decoherent histories. In this formulation of quantum mechanics, the probability of obtaining a particular set of measurement results given projection both onto an initial state and onto a final state is just the *conditional* probability that those outcomes occur, given a final measurement whose result yields the outcome corresponding to the final-state projection. Because of this relation to conditional probability in conventional quantum mechanics [21], although final-state projection can give rise to superluminal communication, as in the case under consideration here, it cannot give rise to the usual causal paradoxes that arise under superluminal communication (such as the "grandfather paradox").

For the same reason, final-state projection cannot renormalize positive quantities to infinite ones. In particular, let the expectation value of some quantity A in for a conventional quantum measurement be $\bar{A}$, and let p be the probability of the final state. The Markov inequality then implies the following.

Theorem: Final-state projection can increase the expectation value $\bar{A}$ of a positive quantity by at most a factor $1/p$.

So as the usual quantum mechanical probability p of the final state becomes small, the expectation values of operators can be renormalized by a larger and larger amount. But for well-defined final-state projection they do not become infinite. Massar and Parentani [23] showed in the context of black-hole geometries that projecting onto certain final states at infinity yields infinite energy densities at the horizon. Reference [23] and the theorem above suggest that we should restrict our attention to final states that yield finite energy density near the horizon.

Despite its somewhat dubious provenance, nonlinear quantum mechanics including final-state projection might hold sway in extreme Planckian regimes such as the black-hole singularity. Indeed, there seems to be no problem with initial state projection at initial singularities: our universe apparently started out in a highly regular state. Final-state projection at final singularities simply restores time symmetry to this picture [17–21]. In the absence of a full theory of quantum gravity, we are free to postulate such an effect and to investigate its consequences.

Almost certain escape from a black hole.—Even in the presence of final-state projection, without further assurances that go beyond the HM model, the escape of quantum information from a black hole is by no means certain. Gottesman and Preskill [5] point out that if the incoming Hawking radiation interacts with the collapsing matter within the black hole (as is likely), then the HM model no longer preserves quantum information. In particular, let the interaction between incoming Hawking radiation and matter be given by a unitary transformation U. The transformation between the state of the collapsing matter and the state of the outgoing Hawking radiation is then

$$T = {}_{\text{matter}\otimes\text{in}}\langle\phi|U|\phi\rangle_{\text{in}\otimes\text{out}}. \tag{4}$$

Gottesman and Preskill note that if all U's are allowed, T can be *any* matrix satisfying $\sum_{m,n}|\langle m|T|n\rangle|^2 = 1$, including transformations that completely destroy the quantum information in the matter, leading to purely thermal Hawking radiation. In general, if the state $U|\phi\rangle_{\text{matter}\otimes\text{in}}$ is not perfectly entangled, then some quantum information in the matter is lost.

For the purposes of using a black hole as a quantum computer, the key question is how much quantum information is lost on average due to such interactions. I now show that for *any* final state, not just the special HM states, and for *almost any* U, classical information escapes from the hole with certainty, and quantum information escapes from the hole with fidelity $\approx (8/3\pi)^2 \approx 0.85$. (Equivalently, the information escapes for any U and for almost any final state.) Essentially, all but half a qubit of the quantum information escapes. This fidelity holds in the limit $N \gg 1$ and is independent of the exact number of bits escaping from the hole: it is the fidelity of escape for the entire state of the collapsing matter. Individual quantum bits inside the hole escape with higher fidelity. In the limit $N \gg 1$, individual quantum bits escape from the hole with fidelity arbitrarily close to 1.

Let $|\phi\rangle_{\mathrm{matter}\otimes\mathrm{in}}$ be any final state, including a product state, and let U be a random unitary transformation on the matter and incoming Hawking radiation, selected according to the Haar measure. [The Haar measure is the unique measure over $U(n)$ that is invariant with respect to unitary transformation.] In particular, the final state could be the as yet unknown correct final state specified by the as yet unknown correct theory of quantum gravity. Because U is selected according to the Haar measure, the state

$$|\psi\rangle_{\mathrm{matter}\otimes\mathrm{in}} = U|\phi\rangle_{\mathrm{matter}\otimes\mathrm{in}} \tag{5}$$

is a random pure state of the matter and incoming Hawking radiation, i.e., a pure state selected according to the uniform measure on the sphere in N^2 dimensions. That is, it is a random state selected according to the Hilbert-Schmidt measure. The random nature of U implies that the escape of quantum information from a black hole does not depend on details of the final state.

Because $|\psi\rangle_{\mathrm{matter}\otimes\mathrm{in}}$ is random, it is not perfectly entangled. As a result, black-hole evaporation will not preserve all the quantum information in the collapsing matter. But by the same token, because $|\psi\rangle_{\mathrm{matter}\otimes\mathrm{in}}$ is random, it is *almost* perfectly entangled for large N. In particular, a typical random state is within one-half a qubit of maximum entanglement.

More precisely, a random state in $H_{\mathrm{matter}} \otimes H_{\mathrm{in}}$ can be written in Schmidt form as

$$|\psi\rangle_{\mathrm{matter}\otimes\mathrm{in}} = \sum_\ell \lambda_\ell |\ell\rangle'_{\mathrm{matter}} \otimes |\ell\rangle'_{\mathrm{in}}. \tag{6}$$

The distribution of the Schmidt coefficients λ_ℓ for random states is known [24–26]. A random state is almost perfectly entangled [27–30]: the average entropy of entanglement, $-\sum_\ell \lambda_\ell^2 \log_2 \lambda_\ell^2$, is within one-half bit of its maximum possible value, $\log_2 N$. It is the high entanglement of random states that leads to the escape of information from the hole.

We now can calculate the average fidelity with which a state for the collapsing matter fields

$$|\mu\rangle_{\mathrm{matter}} = \sum_\ell \mu_\ell |\ell\rangle'_{\mathrm{matter}} \tag{7}$$

is transferred to the outgoing Hawking radiation.

First, look at what happens to the information inside the hole under final-state projection. Action of U on $|\mu\rangle$ together with the incoming Hawking radiation, followed by projection onto the final state $|\phi\rangle_{\mathrm{matter}\otimes\mathrm{in}}$, yields a transformation from the matter to the outgoing Hawking radiation

$$T = {}_{\mathrm{matter}\otimes\mathrm{in}}\langle\psi|\phi\rangle_{\mathrm{in}\otimes\mathrm{out}}. \tag{8}$$

The (unnormalized) state of the outgoing Hawking radiation is

$$|\phi\rangle_{\mathrm{out}} = \frac{1}{\sqrt{N}} \sum_\ell \lambda_\ell \mu_\ell |\ell\rangle'_{\mathrm{out}}, \tag{9}$$

where $\{|\ell\rangle'_{\mathrm{out}}\}$ is a basis for the Hilbert space of outgoing Hawking radiation, related to the basis $\{|\ell\rangle'_{\mathrm{matter}}\}$ for the Hilbert space for the collapsing matter via a unitary transformation T'. Because the normalization of this state depends in a nonlinear fashion on the μ_ℓ, this is a nonlinear transformation of the input state of the matter.

Now look at how quantum information escapes from the hole. Comparing the (normalized) outgoing state of the Hawking radiation with T' times the state of the collapsing matter, we obtain

$$|{}_{\mathrm{out}}\langle\phi|T'|\mu\rangle_{\mathrm{matter}}|^2 = (\sqrt{N}\sum_\ell \lambda_\ell |\mu_\ell|^2)^2. \tag{10}$$

Since a typical state has $|\mu_\ell|^2 \approx 1/N$, the state of the collapsing matter is transferred to the state of the outgoing Hawking radiation with a fidelity

$$f \approx \left(\frac{1}{\sqrt{N}} \sum_\ell \lambda_\ell\right)^2. \tag{11}$$

This approximate result can be confirmed using standard treatments of teleportation with imperfectly entangled states [29]. The maximum mean teleportation fidelity attainable using imperfectly entangled states with Schmidt coefficients λ_ℓ is

$$\bar{f} = \frac{1}{N+1}\left[1 + \left(\sum_\ell \lambda_\ell\right)^2\right]. \tag{12}$$

This fidelity is attained for the standard teleportation protocols. Because escape from a black hole via final-state projection is equivalent to teleportation with a fixed measurement outcome, this is also the mean fidelity for escape from a black hole.

The techniques of [26] now allow us to estimate the value of $\bar{f}$. For $N \gg 1$, we have

$$\left\langle \sum_\ell \lambda_\ell \right\rangle = \sqrt{N} \frac{\Gamma(2)}{\Gamma(3/2)\Gamma(5/2)}\left[1 + O\left(\frac{1}{N}\right)\right] \approx \frac{8}{3\pi}\sqrt{N}. \tag{13}$$

As a result, for $N \gg 1$, we have

$$\bar{f} \approx \left(\frac{8}{3\pi}\right)^2 \approx 0.85. \tag{14}$$

Quantum information escapes from the hole with fidelity ≈ 0.85. (Note that in this estimate we are approximating $\langle(\sum_\ell \lambda_\ell)^2\rangle$ by $\langle(\sum_\ell \lambda_\ell)\rangle^2$; this approximation is accurate because correlations between λ_ℓ, $\lambda_{\ell'}$ go to zero in the limit that $N \gg 1$.)

This fidelity is the fidelity for escape of the entire state of the collapsing matter. The fidelity of escape of individual quantum bits is higher and approaches 1 asymptotically as N becomes large. Because the escape fidelity lies above the threshold required for quantum error correction, suitably encoded quantum information escapes from the black hole with fidelity arbitrarily close to 1. Given final-state projection, escape from a black hole is almost certain. In the case that the Hawking radiation has the normal canonical distribution instead of the microcanonical form assumed in [4], the fidelity in Eqs. (10)–(14) will be slightly diminished because the Hawking radiation is no longer perfectly entangled, but almost all of the quantum information in the hole still escapes.

The above demonstration of almost certain escape from a black hole via final-state projection relies on a random interaction between the collapsing matter and the incoming Hawking radiation. As only a finite proper time exists for interaction between the matter and the incoming Hawking radiation, this interaction is not truly random. What is important for the escape of the quantum information is not true randomness, however, but entanglement. We have recently demonstrated both theoretically and experimentally that *pseudorandom* states and transformations, implemented by quantum logic circuits with depth of $O(2n)$ gates for $n = \log_2 N$ qubits, exhibit the same Schmidt coefficient statistics as true random states and transformations [30]. In other words, most local interactions between n qubits give rise to states whose entanglement allows almost certain escape from a black hole. In particular, if the final plunge into the singularity is chaotic, the corresponding quantum transformation is typically pseudorandom [30]. Accordingly, we may reasonably hope that the final state, whatever it is, is sufficiently entangled to give high fidelity transfer of the state of the matter within the hole to the state of the outgoing Hawking radiation.

The results of this Letter suggest that if black holes evaporate via final-state projection, they might make good quantum computers. The fidelity of transfer of quantum information is better than what is required for robust quantum computation. Indeed, if all one wants is a yes or no answer from the computation, i.e., a classical bit, then the black hole can deliver the answer with certainty.

Note that a person outside the hole must know the exact interaction that occurred between the collapsing matter and the incoming Hawking radiation in order to reconstruct the information escaping from the hole. Even when that interaction is known, until the hole is almost entirely evaporated, the outgoing Hawking radiation appears essentially random. Final-state projection will have to await experimental and theoretical confirmation before black holes can be used as quantum computers. It would be premature to jump into a black hole just now.

The author thanks A. Hosoya for bringing this issue to his attention.

*Electronic address: slloyd@mit.edu

[1] S. Lloyd, Nature (London) **406**, 1047 (2000).
[2] S. Lloyd, Phys. Rev. Lett. **88**, 237901 (2002).
[3] Y. J. Ng, Phys. Rev. Lett. **86**, 2946 (2001); **88**, 139902(E) (2002).
[4] G. T. Horowitz and J. Maldacena, J. High Energy Phys. 02 (2004) 008; hep-th/0310281.
[5] D. Gottesman and J. Preskill, J. High Energy Phys. 03 (2004) 026; hep-th/0311269.
[6] U. Yurtsever and G. Hockney, quant-ph/0312160.
[7] U. Yurtsever and G. Hockney, quant-ph/0402060.
[8] S. Weinberg, Phys. Rev. Lett. **62**, 485 (1989).
[9] J. Polchinksi, Phys. Rev. Lett. **66**, 397 (1991).
[10] N. Gisin, Phys. Lett. A **143**, 1 (1990).
[11] A. Peres, Phys. Rev. Lett. **63**, 1114 (1989).
[12] D. Abrams and S. Lloyd, Phys. Rev. Lett. **81**, 3992 (1998).
[13] P. K. Majumder *et al.*, Phys. Rev. Lett. **65**, 2931 (1990).
[14] R. L. Walsworth *et al.*, Phys. Rev. Lett. **64**, 2599 (1990).
[15] T. E. Chupp and R. J. Hoare, Phys. Rev. Lett. **64**, 2261 (1990).
[16] J. J. Bollinger, D. J. Heinzen, W. M. Itano, S. L. Gilbert, and D. J. Wineland, Phys. Rev. Lett. **63**, 1031 (1989).
[17] Y. Aharonov, P. Bergmann, and J. Lebovitz, Phys. Rev. **134**, B1410 (1964).
[18] R. B. Griffiths, J. Stat. Phys. **36**, 219 (1984).
[19] L. S. Schulman, Phys. Rev. D **7**, 2868 (1973); J. Stat. Phys. **16**, 217 (1977).
[20] M. Gell-Mann and J. B. Hartle, in *The Physical Origins of Time Asymmetry*, edited by J. Halliwell, J. Pérez-Mercader, and W. H. Zurek (Cambridge University Press, Cambridge, 1994).
[21] A. P. Kent, Phys. Rev. D **59**, 043505 (1999).
[22] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965).
[23] S. Massar and R. Parentani, Phys. Rev. D **54**, 7444 (1996).
[24] E. Lubkin, J. Math. Phys. (N.Y.) **19**, 1028 (1978).
[25] S. Lloyd and H. Pagels, Ann. Phys. (N.Y.) **188**, 186 (1988).
[26] H.-J. Sommers and K. Życzkowski, J. Phys. A **37**, 8457 (2004).
[27] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).
[28] S. Sen, Phys. Rev. Lett. **77**, 1 (1996).
[29] K. Baraszek, Phys. Rev. A **62**, 024301 (2000); quant-ph/0002088.
[30] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. Cory, Science **302**, 2098 (2003).