PHYSICAL REVIEW D **88,** 084010 (2013)

# Mutual information in Hawking radiation

Norihiro Iizuka[1,*] and Daniel Kabat[2,†]

[1]*Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan*

[2]*Department of Physics and Astronomy, Lehman College, City University of New York, Bronx, New York 10468, USA*

(Received 16 August 2013; published 8 October 2013)

We compute the mutual information of two Hawking particles emitted consecutively by an evaporating black hole. Following Page, we find that the mutual information is of order $e^{-S}$ where $S$ is the entropy of the black hole. We speculate on implications for black hole unitarity, in particular on a possible failure of locality at large distances.

DOI: 10.1103/PhysRevD.88.084010 PACS numbers: 04.70.Dy, 04.60.−m, 11.25.Tq

Hawking's discovery that black holes emit thermal radiation [1] is one of the few tangible results in quantum gravity, and the resulting conflict with unitarity [2] has driven much of the research in the field. See [3] for a review. The goal of the present paper is to obtain new insight into this issue, from a computation of the mutual information carried by successive Hawking particles.

Consider two successive Hawking particles emitted by an evaporating black hole, as shown in Fig. 1. Motivated by the AdS/CFT correspondence we assume that a conventional quantum mechanical description of this process is available. In particular we assume there is an underlying Hilbert space with unitary time evolution that describes the microscopic degrees of freedom. In this fine-grained description there is no tension with unitarity: the two Hawking particles are correlated due to their shared history, in which they both interacted with the microscopic black hole degrees of freedom. In this fine-grained description, Hawking radiation from a black hole is no different from the blackbody radiation emitted by any other hot macroscopic object.

However for a black hole we would like to consider a coarse-grained description, in which the black hole is characterized just by its macroscopic thermodynamic properties such as energy and entropy. It seems reasonable that this coarse-graining gives rise to the usual notion of a semiclassical spacetime. That is, only in a coarse-grained description could one hope to describe the black hole using the usual Schwarzschild metric, and could one hope to describe Hawking radiation using effective field theory on the Schwarzschild background. In support of this view, note that the usual black hole metric only captures macroscopic properties such as mass or charge. Also Hawking's calculation, carried out in this context, shows that a black hole emits uncorrelated thermal radiation. This behavior is expected in a coarse-grained description of blackbody radiation, since such radiation is completely characterized by a macroscopic quantity, namely the temperature of the black hole.

In this setting, to understand unitarity, the main challenge is identifying which properties of the coarse-grained description deviate most significantly from the underlying microscopic description. Given some underlying microscopic theory, what modification to the coarse-grained description is most appropriate for restoring unitarity?

To sharpen our discussion we consider the correlation between the two successive Hawking particles $a$ and $b$ shown in Fig. 1. We have in mind two particles that are emitted almost simultaneously from well-separated points on the horizon, so that the separation between $a$ and $b$ is large and spacelike. The correlation can be measured by the mutual information

$$I_{ab} = S_a + S_b - S_{ab} \tag{1}$$

where $S_a$ is the entropy of particle $a$, $S_b$ is the entropy of particle $b$, and $S_{ab}$ is the entropy of both. According to Hawking's calculation $a$ and $b$ are uncorrelated and the mutual information vanishes. Under seemingly reasonable assumptions this will remain true even in the presence of interactions [3]. But if the entire system (including the black hole) is in a random pure state, the true correlation between $a$ and $b$ can be obtained from the fundamental work of Page [4]. Page considers a Hilbert space of dimension $m$ entangled with another Hilbert space of dimension $n \geq m$, and shows that in a random pure state the average entropy is

$$S_{m,n} = \sum_{k=n+1}^{mn} \frac{1}{k} - \frac{m-1}{2n}. \tag{2}$$

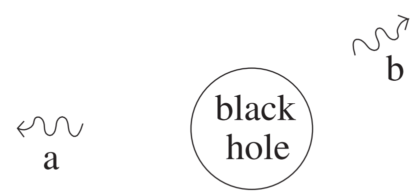


FIG. 1. Successive Hawking particles emitted by a black hole.

*iizuka@yukawa.kyoto-u.ac.jp
†daniel.kabat@lehman.cuny.edu

 

For large $n$ the sum can be estimated using the Euler-Maclaurin formula, which gives

$$S_{m,n} = \log m - \frac{m^2 - 1}{2mn} + \mathcal{O}(1/n^2). \tag{3}$$

To apply this to the situation at hand, let $N_a$ be the dimension of the Hilbert space of particle $a$, let $N_b$ be the dimension of the Hilbert space of particle $b$, and let $N_{bh} = e^S$ be the dimension of the Hilbert space of the black hole. For particle $a$, for example, we have a Hilbert space of dimension $N_a$ entangled with a Hilbert space of dimension $N_b N_{bh}$. Thus

$$S_a = S_{N_a, N_b N_{\rm bh}} \quad S_b = S_{N_b, N_a N_{\rm bh}} \quad S_{ab} = S_{N_a N_b, N_{\rm bh}}. \tag{4}$$

Using (1) and (3) we find that for large $N_{\rm bh}$, the mutual information in the Hawking particles $a$ and $b$ is

$$I_{ab} = \frac{(N_a^2 - 1)(N_b^2 - 1)}{2 N_a N_b N_{\rm bh}} + \mathcal{O}(1/N_{\rm bh}^2). \tag{5}$$

This is our main result. It shows that the mutual information carried by two successive Hawking particles is of order $e^{-S}$. For example if each Hawking particle could carry one bit of information then $N_a = N_b = 2$ and $I_{ab} \approx \frac{9}{8} e^{-S}$, while if each Hawking particle could carry a large amount of information then $I_{ab} \approx \frac{1}{2} N_a N_b e^{-S}$.

As we discussed above, the usual semiclassical picture of gravity must be modified in order to reproduce these correlations. Roughly speaking the possible modifications fall into three categories.

(1) Modify the interior
It could be that microscopic quantum gravity effects become important at or inside the (stretched) horizon of the black hole, invalidating the use of the classical Schwarzschild geometry in this region and generating correlations between outgoing Hawking particles. However outside the horizon semiclassical gravity and effective field theory could be valid. Proposals of this type include fuzzballs [5,6] and firewalls [7,8].

(2) Modify the exterior
It could be that effective field theory is not trustworthy, even at macroscopic distances outside the black hole. For example, it could be that the underlying theory of quantum gravity leads to violations of locality over large distances, in a way that generates correlations and restores unitarity. Some models with nonlocality have been discussed in [9–11].

(3) Modify both
Perhaps both the interior region of the black hole and the rules of effective field theory outside the black hole receive important corrections due to microscopic quantum gravity effects.

Unfortunately, just from considerations of unitarity, there is no clear way to decide between these possibilities. But since most models discussed in the literature take other approaches, let us indulge in a little speculation about the possibility of nonlocality outside the horizon.

A key principle in local field theory is microcausality, that is, the property that field operators commute at spacelike separation. If we are prepared to give up on locality outside the horizon, it could be that spacelike separated field operators no longer commute. We have in mind that the resulting nonlocality extends over macroscopic distances, and would thus fall into the category of modifying the exterior of the black hole. But we must admit that in order to restore unitarity, nonlocality which extends to the stretched horizon could do the job.

In fact, AdS/CFT may provide some motivation for the radical idea of nonlocality over macroscopic distances. Order by order in the $1/N$ expansion of the conformal field theory (CFT) one can construct CFT operators which mimic local field operators in the bulk [12,13]. The algorithm involves starting from a single primary field and adding an infinite tower of higher dimension operators. In the $1/N$ expansion one can show that the resulting CFT operators commute whenever the bulk points are spacelike separated.[1] But at finite $N$ it seems unlikely that the higher dimension operators required for bulk locality could exist. Instead it is more likely that bulk observables will fail to commute at spacelike separation, even over macroscopic distances, by an amount which is nonperturbatively small in the $1/N$ expansion.

As a toy model for this idea, consider a pair of independent harmonic oscillators characterized by

$$[\hat{\alpha}, \hat{\alpha}^\dagger] = [\hat{\beta}, \hat{\beta}^\dagger] = 1 \tag{6}$$

with all other commutators vanishing. We think of these oscillators as representing two independent degrees of freedom in some underlying microscopic description ("the boundary"). Suppose these boundary operators can be mapped to bulk operators which describe the Hawking particles shown in Fig. 1. We assume a boundary-to-bulk map depending on two parameters $\theta$ and $\phi$, explicitly given by

$$\hat{a} = \hat{\alpha} \cosh(\theta + \phi) + \hat{\beta}^\dagger \sinh(\theta + \phi) \tag{7}$$

$$\hat{b} = \hat{\beta} \cosh(\theta - \phi) + \hat{\alpha}^\dagger \sinh(\theta - \phi) \tag{8}$$

$$\hat{a}^\dagger = \hat{\alpha}^\dagger \cosh(\theta + \phi) + \hat{\beta} \sinh(\theta + \phi) \tag{9}$$

$$\hat{b}^\dagger = \hat{\beta}^\dagger \cosh(\theta - \phi) + \hat{\alpha} \sinh(\theta - \phi). \tag{10}$$

[1]In fact bulk microcausality can be taken as a guiding principle for the construction of bulk observables [12]. The procedure is simplest for bulk scalars, but it works for gauge fields as well [14].

This map can be thought of as a Bogoliubov transformation

$$\hat{\alpha}' = \hat{\alpha}\cosh\theta + \hat{\beta}^\dagger \sinh\theta \tag{11}$$

$$\hat{\beta}' = \hat{\beta}\cosh\theta + \hat{\alpha}^\dagger \sinh\theta \tag{12}$$

followed by setting

$$\hat{a} = \hat{\alpha}'\cosh\phi + \hat{\beta}'^\dagger \sinh\phi \tag{13}$$

$$\hat{b} = \hat{\beta}'\cosh\phi - \hat{\alpha}'^\dagger \sinh\phi. \tag{14}$$

The Bogoliubov transformation (11) and (12) preserves the canonical commutation relations. But this is not true of the transformation (13) and (14), due to the relative $-$ sign which appears in (14). Rather the combined map leads to bulk commutators

$$[\hat{a}, \hat{a}^\dagger] = [\hat{b}, \hat{b}^\dagger] = 1 \tag{15}$$

$$[\hat{a}^\dagger, \hat{b}^\dagger] = [\hat{b}, \hat{a}] = \sinh 2\phi \tag{16}$$

with all other commutators vanishing. Note that $\theta$ drops out of the commutation relations. This is expected since $\theta$ in (11) and (12) parametrizes a Bogoliubov transformation between the bulk and boundary degrees of freedom, which by definition is a transformation that preserves the commutators. Having nonzero $\phi$, on the other hand, leads to a nonzero commutator between $\hat{a}$ and $\hat{b}$. We think of this as representing a bulk commutator which is nonzero at spacelike separation. This is a drastic modification to local field theory—a risky game to play—and it is not clear whether a consistent theory can be constructed along these lines. But let us proceed, and explore the consequences of noncommutativity.

One can start from the microscopic vacuum

$$\hat{\alpha}|0,0\rangle = \hat{\beta}|0,0\rangle = 0 \tag{17}$$

and build a Fock space

$$|n_\alpha, n_\beta\rangle = \frac{1}{\sqrt{n_\alpha! n_\beta!}}(\hat{\alpha}^\dagger)^{n_\alpha}(\hat{\beta}^\dagger)^{n_\beta}|0,0\rangle. \tag{18}$$

If one only acts on the vacuum with operators of type $a$ one never notices the noncommutativity (likewise for type $b$). But suppose the Hawking particles shown in Fig. 1 correspond to a two-particle state (which we have not bothered to normalize)

$$|\psi\rangle \sim \hat{a}^\dagger\hat{b}^\dagger|0,0\rangle \sim \sinh(\theta+\phi)|0,0\rangle + \cosh(\theta+\phi)|1,1\rangle. \tag{19}$$

In general this state is entangled. To see this we split the Hilbert space into $\alpha$ and $\beta$ oscillators, $\mathcal{H} = \mathcal{H}_\alpha \times \mathcal{H}_\beta$. The choice of splitting is somewhat arbitrary, and leads to a freedom that we discuss in more detail below. Given the splitting, we construct the density matrix $\hat{\rho} = |\psi\rangle\langle\psi|$ and trace over $\mathcal{H}_\beta$ to obtain the reduced density matrix for particle $a$[2]:

$$\hat{\rho}_a = {}_\beta\langle 0|\hat{\rho}|0\rangle_\beta + {}_\beta\langle 1|\hat{\rho}|1\rangle_\beta. \tag{20}$$

Properly normalized, this procedure gives

$$\hat{\rho}_a = \frac{1}{1+\xi_+^2}(\xi_+^2|0\rangle\langle 0| + |1\rangle\langle 1|) \tag{21}$$

where $\xi_+$ is defined by $\xi_+ = \tanh(\theta+\phi)$. The associated entropy is

$$S_a = -\mathrm{Tr}\hat{\rho}_a \log\hat{\rho}_a = -\frac{\xi_+^2}{1+\xi_+^2}\log\xi_+^2 + \log(1+\xi_+^2). \tag{22}$$

The mutual information between $a$ and $b$ is $I_{ab} = S_a + S_b - S_{ab}$. But $S_b = S_a$, while the combined system is in a pure state with $S_{ab} = 0$. So the mutual information is simply twice the result (22),

$$I_{ab} = -\frac{2\xi_+^2}{1+\xi_+^2}\log\xi_+^2 + 2\log(1+\xi_+^2). \tag{23}$$

Of course this result depends on how we decide to split the Hilbert space. In other words, it depends on what we decide to trace over in constructing $\hat{\rho}_a$. But the freedom to choose a splitting can be absorbed into a shift of the Bogoliubov parameter $\theta$. More precisely $\theta$ parametrizes the freedom to split the Hilbert space into $\mathcal{H}_{\alpha'} \times \mathcal{H}_{\beta'}$, where $\alpha'$ and $\beta'$ are the independent oscillators defined in (11) and (12).[3]

One can use this freedom to set $\theta + \phi = 0$, which makes the mutual information in the state (19) vanish. But even if the mutual information in this particular state vanishes, there will still be other states that carry mutual information. For example the state

$$|\psi\rangle = \hat{b}^\dagger\hat{a}^\dagger|0,0\rangle \tag{24}$$

has mutual information which can be obtained from (23) by replacing $\xi_+ \to \xi_- \equiv \tanh(\theta-\phi)$, namely

$$I_{ab} = -\frac{2\xi_-^2}{1+\xi_-^2}\log\xi_-^2 + 2\log(1+\xi_-^2). \tag{25}$$

In attempting to make both (23) and (25) small the best one can do is set $\theta = 0$. Then (23) and (25) are equal, and at leading order for small $\phi$ the mutual information in either of the states $\hat{a}^\dagger\hat{b}^\dagger|0\rangle$ or $\hat{b}^\dagger\hat{a}^\dagger|0\rangle$ is

[2] A word on notation: the density matrix we are constructing depends on both our choice of state $\hat{a}^\dagger\hat{b}^\dagger|0,0\rangle$ and on our splitting of the Hilbert space $\mathcal{H}_\alpha \times \mathcal{H}_\beta$. The notation $\hat{\rho}_a$ emphasizes the former over the latter.

[3] Note that standard field theory does not have this freedom. In standard field theory operators commute at spacelike separation, so one can unambiguously associate a factor of the Hilbert space with any given spatial region. Then for a given region there is no freedom in deciding what to trace over when computing entropy of entanglement.

$$I_{ab} \approx 2\phi^2(1 - \log \phi^2). \quad (26)$$

This toy model suggests that two operators which have a commutator that is $\mathcal{O}(\phi)$ as in (16) typically produce entangled states with a mutual information that is $\mathcal{O}(\phi^2)$ as in (26). Since we know how much mutual information is present in Hawking radiation, we can estimate how big commutators must be at spacelike separation. Our results suggest that two field operators should have a commutator of order $e^{-S/2}$ in the presence of a black hole, in order to account for the mutual information $I_{ab} \sim e^{-S}$ carried by two successive Hawking particles. (More precisely, we have in mind that matrix elements of the commutator in a typical state of the black hole plus Hawking radiation should be of order $e^{-S/2}$.) In the context of AdS/CFT black hole entropy is $\mathcal{O}(N^2)$, so this effect is nonperturbatively small in the $1/N$ expansion of the CFT.

We are grateful to Gilad Lifschytz for valuable discussions and for comments on the manuscript. D. K. thanks the YITP for hospitality during this work. N. I. is supported in part by JSPS KAKENHI Grant No. 25800143. D. K. is supported by U.S. National Science Foundation Grant No. PHY-1125915 and by grants from PSC-CUNY.

[1] S. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[2] S. Hawking, Phys. Rev. D **14**, 2460 (1976).
[3] S. D. Mathur, Classical Quantum Gravity **26**, 224001 (2009).
[4] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).
[5] S. D. Mathur, Fortschr. Phys. **53**, 793 (2005).
[6] S. D. Mathur, arXiv:0810.4525.
[7] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, J. High Energy Phys. 02 (2013) 062.
[8] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, J. High Energy Phys. 09 (2013) 018.
[9] S. B. Giddings, Phys. Rev. D **85**, 044038 (2012).
[10] S. B. Giddings, Phys. Rev. D **88**, 064023 (2013).
[11] S. B. Giddings, Phys. Rev. D **88**, 024018 (2013).
[12] D. Kabat, G. Lifschytz, and D. A. Lowe, Phys. Rev. D **83**, 106009 (2011).
[13] I. Heemskerk, D. Marolf, J. Polchinski, and J. Sully, J. High Energy Phys. 10 (2012) 165.
[14] D. Kabat and G. Lifschytz, Phys. Rev. D **87**, 086004 (2013).