Radiation from a Collapsing Object is Manifestly Unitary

Anshul Saini and Dejan Stojkovic
HEPCOS, Department of Physics, SUNY at Buffalo, Buffalo, New York 14260-1500, USA
(Received 8 November 2014; revised manuscript received 25 January 2015; published 17 March 2015)

The process of gravitational collapse excites the fields propagating in the background geometry and gives rise to thermal radiation. We demonstrate by explicit calculations that the density matrix corresponding to such radiation actually describes a pure state. While Hawking's leading order density matrix contains only the diagonal terms, we calculate the off-diagonal correlation terms. These correlations start very small, but then grow in time. The cumulative effect is that the correlations become comparable to the leading order terms and significantly modify the density matrix. While the trace of the Hawking's density matrix squared goes from unity to zero during the evolution, the trace of the total density matrix squared remains unity at all times and all frequencies. This implies that the process of radiation from a collapsing object is unitary.

DOI: 10.1103/PhysRevLett.114.111301 PACS numbers: 04.70.Dy, 03.65.-w, 03.70.+k

Introduction.—One of the most pressing problems in modern physics is the information loss paradox in black hole physics. Since Hawking radiation is purely thermal [1], it is possible to convert a pure state into a mixed state, which is forbidden in unitary quantum mechanics [2]. It was often argued that subtle correlations between the emitted Hawking quanta which are usually neglected could be enough to recover information about the initial state and convert an apparently maximally mixed thermal state into a pure state [3,4]. This point of view was also often criticized by noticing that small corrections to the leading order Hawking terms are not enough to recover unitarity. The purpose of this Letter is to perform explicit calculations which may clarify this issue. We find indeed that the process of gravitational collapse and subsequent evaporation is manifestly unitary as seen by an asymptotic observer.

We used the functional Schrödinger formalism, which is especially convenient for this question since it gives us the time evolution of the system rather than radiation from a preexisting black hole [5–17]. We start with a massive shell that is collapsing under its own gravitational pull. This process induces a nontrivial time-dependent metric which then excites the field quanta. The process of the gravitational collapse takes infinite time for an outside observer; however, radiation is pretty close to thermal when the collapsing shell approaches its own Schwarzschild radius. Our formalism gives us an explicit form of the wave function of the emitted radiation, which contains complete information not only about the diagonal Hawking terms, but also about the nondiagonal correlations terms. Correlations between the Hawking quanta are at first indeed negligible with respect to the diagonal terms. However, time evolution creates progressively more off-diagonal terms than the diagonal ones. Moreover, time evolution is such that these cross terms become of the same order of magnitude as the Hawking terms. As a result, the density matrix for the emitted radiation is significantly modified, in particular, it is not purely diagonal. We calculate the time evolution of the complete density matrix as a function of time and frequency. The relevant quantity that we want to obtain is the trace of the density matrix squared [$\mathrm{Tr}(\hat{\rho}^2)$], which tells us whether the system is in a pure or mixed state. We find that if we take only diagonal terms in density matrix then $\mathrm{Tr}(\hat{\rho}_h{}^2)$ goes from unity to zero, which means that the state goes from pure to mixed. This is the standard Hawking's result that implies information loss. However, if we include the off-diagonal terms then $\mathrm{Tr}(\hat{\rho}^2)$ remains unity at all frequencies and all times during the evolution. This means that the initial state stays pure during the evolution. This is the main result of our analysis.

The formalism.—We consider a thin shell of matter which collapses under its own gravity. We use Schwarzschild coordinates because we are interested in the point of view of an observer at infinity. The metric outside the shell can be written as

$$ds^2 = -\left(1 - \frac{R_s}{r}\right)dt^2 + \left(1 - \frac{R_s}{r}\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (1)$$

The interior of the domain wall is a flat spacetime due to the Birkhoff theorem

$$ds^2 = -dT^2 + dr^2 + r^2 d\Omega^2. \quad (2)$$

The time coordinates of the two regions are related with the proper time inside the shell as

$$\frac{dT}{d\tau} = \sqrt{1 + R_\tau{}^2}, \qquad \frac{dt}{d\tau} = \frac{\sqrt{B + R_\tau{}^2}}{B}, \quad (3)$$

where $B = 1 - R_s/R$ and $R_\tau = (dR/d\tau)$. From here we get

$$\frac{dT}{dt} = \sqrt{B - \left(\frac{1-B}{B}\right) R_t{}^2}. \tag{4}$$

An action of the massless scalar field propagating in the background of the collapsing shell can be written as

$$S = \int d^4x \sqrt{-g} \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi, \tag{5}$$

where ϕ is a scalar field, which we can expand in terms of the modes as

$$\phi = \sum_\lambda a_\lambda(t) f_\lambda(r). \tag{6}$$

In the interior of shell, the action takes the form

$$S_{\rm in} = 2\pi \int dt \int_0^{R(t)} drr^2 \left[-\frac{(\partial_t \phi)^2}{T_t} + T_t (\partial_t \phi)^2 \right]. \tag{7}$$

Similarly, outside of the shell it becomes

$$S_{\rm out} = 2\pi \int dt \int_{R(t)}^\infty drr^2 \left[-\frac{(\partial_t \phi)^2}{1 - \frac{R_s}{r}} + \left(1 - \frac{R_s}{r}\right) (\partial_t \phi)^2 \right]. \tag{8}$$

The classical equation of motion for this collapsing shell near the horizon can be written as [5]

$$R_t = \pm B \sqrt{1 - \frac{BR^4}{h^2}}, \tag{9}$$

where h is a constant. Using Eq. (4) and Eq. (9), we get

$$T_t = B \sqrt{1 + (1 - B) \frac{R^4}{h^2}}. \tag{10}$$

When the shell is approaching its own Schwarzschild radius, $R \to R_s$, then $T_t \to 0$; hence, the total action becomes

$$S \sim 2\pi \int dt \left(-\frac{1}{B} \int_0^{R_s} drr^2 (\partial_t \phi)^2 + \int_{R_s}^\infty drr^2 \left(1 - \frac{R_s}{r}\right) (\partial_t \phi)^2 \right), \tag{11}$$

which in terms of the modes gives

$$S = \int dt \left(-\frac{1}{2B} \frac{da_k}{dt} A_{kk'} \frac{da_{k'}}{dt} + \frac{1}{2} a_k B_{kk'} a_{k'} \right) \tag{12}$$

with

$$A_{kk'} = 4\pi \int_0^{R_s} drr^2 f_k(r) f_{k'}(r), \tag{13}$$

$$B_{kk'} = 4\pi \int_{R_s}^\infty drr^2 \left(1 - \frac{R_s}{r}\right) f'_k(r) f'_{k'}(r). \tag{14}$$

Matrices A and B are independent of $R(t)$. From the action (12), we can find the corresponding Hamiltonian and write down the Schrödinger equation $H\psi = i\partial\psi/\partial t$ as

$$\left[\left(1 - \frac{R_s}{R}\right) \frac{1}{2} \Pi_k (A^{-1})_{kk'} \Pi_{k'} + \frac{1}{2} a_k B_{kk'} a_{k'} \right] \psi = i \frac{\partial \psi}{\partial t}, \tag{15}$$

where the momentum is defined as

$$\Pi_k = -i \frac{\partial}{\partial a_k}. \tag{16}$$

Since matrices A and B are symmetric and real, both can be diagonalized simultaneously with respective to eigenvalues α and β. One can then write the Schrödinger equation in terms of eigenmodes y (which are linear combinations of the original modes a) as

$$\left[-\left(1 - \frac{R_s}{R}\right) \frac{1}{2\alpha} \frac{\partial^2}{\partial y^2} + \frac{1}{2} \beta y^2 \right] \psi(y,t) = i \frac{\partial \psi(y,t)}{\partial t}. \tag{17}$$

Defining

$$\eta = \int_0^t dt \left(1 - \frac{R_s}{R}\right) \tag{18}$$

one can rewrite Eq. (17) in a form similar to the harmonic oscillator equation as

$$\left[-\frac{1}{2\alpha} \frac{\partial^2}{\partial y^2} + \frac{\alpha}{2} \omega^2(\eta) y^2 \right] \psi(y,\eta) = i \frac{\partial \psi(y,\eta)}{\partial \eta}, \tag{19}$$

where

$$\omega^2(\eta) = \left(\frac{\beta}{\alpha}\right) \frac{1}{B} \equiv \frac{\omega_0{}^2}{B}. \tag{20}$$

The exact solution to this equation is

$$\psi(y,\eta) = e^{i\delta(\eta)} \left[\frac{\alpha}{\pi \theta^2} \right]^{(1/4)} \exp \left[i \left(\frac{\theta_\eta}{\theta} + \frac{i}{\theta^2} \right) \frac{\alpha y^2}{2} \right], \tag{21}$$

where θ is the solution of the differential equation

$$\theta_{\eta\eta} + \omega^2(\eta) \theta = \frac{1}{\theta^3} \tag{22}$$

with initial conditions

$$\theta(0) = \frac{1}{\sqrt{\omega_0}}, \qquad \theta_\eta(0) = 0. \tag{23}$$

Since the background spacetime is time dependent, we make a distinction between the initial frequency ω_0 at which the mode is created from the vacuum, and the final frequency at some later time t defined as

$$\bar{\omega} = \omega_0 e^{t/2R_s}. \tag{24}$$

The wave function $\psi(y,t)$ contains information about the modes or particles excited in the spacetime in terms of their frequencies at the final moment t. We want to construct a density matrix of the system so we need to expand the wave function in terms of a complete basis. We will use the simple harmonic oscillator (SHM) basis $\zeta_n(y)$.

$$\psi(y,t) = \sum_n c_n(t)\zeta_n(y). \tag{25}$$

The number of states in this basis is infinite so the size of the density matrix will be infinite too. However, one can see that the probability of exciting higher n states decreases rapidly as n increases. Therefore, one can easily identify trends even by considering finite (but large enough) n. The coefficients $c_n(t)$ can be written as

$$c_n(t) = \int dy \zeta_n{}^*(y)\psi(y,t). \tag{26}$$

The probability of finding a particle in a particular state n is given by $|c_n(t)|^2$. The coefficients c_n can be explicitly found as (see Supplemental Material [18])

$$c_n(t) = \frac{(-1)^{n/2} e^{i\alpha}}{(\bar{\omega}\rho^2)^{1/4}} \sqrt{\frac{2}{P}} \left(1 - \frac{2}{P}\right)^{n/2} \frac{(n-1)!!}{\sqrt{n!}}, \tag{27}$$

where P is given by

$$P = 1 - \frac{i}{\bar{\omega}} \left(\frac{\theta_\eta}{\theta} + \frac{i}{\theta^2} \right). \tag{28}$$

In order to find c_n we need to solve for θ. The simplest analytic method is given in [19]. θ and θ_η can be found in terms of η and ξ as

$$\theta = \frac{1}{\sqrt{\omega_0}} \sqrt{\xi^2 + \chi^2}. \tag{29}$$

$$\theta_\eta = \frac{1}{\omega_0 \rho} (\xi \xi_\eta + \chi \chi_\eta). \tag{30}$$

η and ξ and their derivatives can be written in terms of Bessel's function as

$$\xi = \frac{\pi u}{2} [Y_0(2\omega_0) J_1(u) - J_0(2\omega_0) Y_1(u)], \tag{31}$$

$$\chi = \frac{\pi u}{2} [Y_1(2\omega_0) J_1(u) - J_1(2\omega_0) Y_1(u)], \tag{32}$$

$$\xi_\eta = -\pi\omega_0^2 [Y_0(2\omega_0) J_0(u) - J_0(2\omega_0) Y_0(u)], \tag{33}$$

$$\chi_\eta = -\pi\omega_0^2 [Y_1(2\omega_0) J_0(u) - J_1(2\omega_0) Y_0(u)], \tag{34}$$

where $u \equiv 2\omega_0 \sqrt{1-\eta}$.

The occupation number at eigenfrequency $\bar{\omega}$ is given by the expectation value

$$N(t, \bar{\omega}) = \sum_n n |c_n|^2. \tag{35}$$

The process of the gravitational collapse takes infinite time for an outside observer; however, radiation is pretty close to Planckian when the collapsing shell approaches its own Schwarzschild radius. Since we are already working in a near-horizon approximation, if we plot $N(t, \bar{\omega})$ for some fixed late t, the spectrum will resemble the thermal Hawking distribution [5]. However, we are here interested in correlations between the emitted quanta, which is contained not in the diagonal spectrum, but actually in the total density matrix for the system.

Density matrix.—Knowing the expansion coefficients c_n explicitly, we can construct the density matrix. The density matrix is defined as

$$\hat{\rho} = \sum |\psi\rangle\langle\psi|. \tag{36}$$

In our basis it can be rewritten as

$$\hat{\rho} = \sum_{mn} c_{mn} |\zeta_m\rangle\langle\zeta_n|, \tag{37}$$

where $c_{mn} \equiv c_m c_n$. The original Hawking radiation density matrix ρ_h contains only the diagonal elements c_{nn}, while the cross terms c_{mn} for $m \neq n$ are absent. The off-diagonal terms represent interactions and correlations between the states. The rationale behind neglecting the cross terms is that these correlations are usually higher order effects and will not affect the Hawking's result in the first order. However, as argued recently in [20] (see also [21]), during the process of Hawking radiation, the correlations may start off very small, but gradually grow as the process continues. It may happen at the end that these off-diagonal terms can modify the Hawking density matrix significantly enough to yield a pure sate. The time-dependent functional Schrödinger formalism is especially convenient to test this proposal since it gives us the time evolution of the system. In Fig. 1, we plot some terms (both diagonal and off-diagonal) in the density matrix. We plot their time evolution with the fixed frequency $\bar{\omega}$. We took absolute values of the

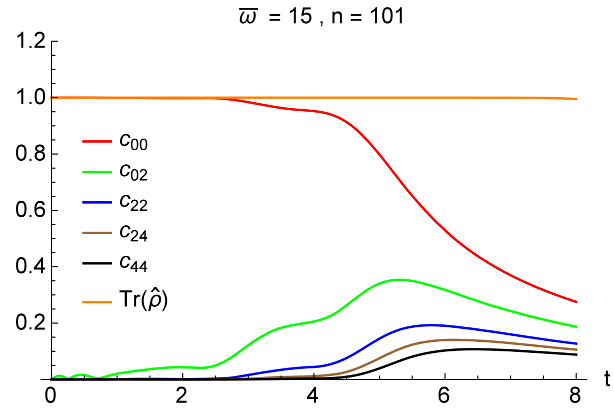

FIG. 1 (color online). Elements of the density matrix c_{mn} and $\mathrm{Tr}(\hat{\rho})$ as a function of time at $\bar{\omega} = 15$, where an index n labeling the modes goes up to $n = 101$. As time increases, the magnitude of c_{00} decreases, $\mathrm{Tr}(\hat{\rho})$ remains unity, and all other c_{mn} increase, reach the maximum values, then decrease again. This implies that small correlations between the modes become as important as the diagonal terms.

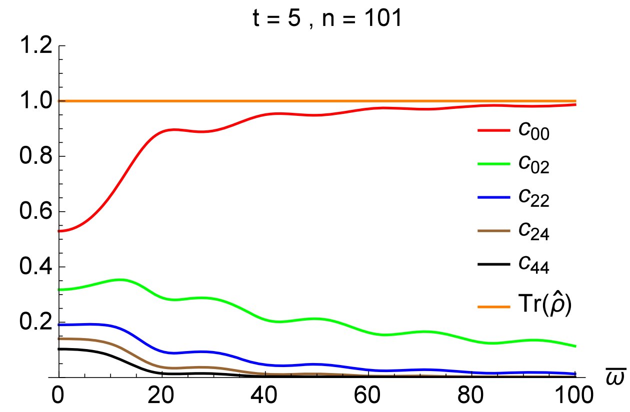

FIG. 2 (color online). Cross terms c_{nm} and $\mathrm{Tr}(\hat{\rho})$ as a function of $\bar{\omega}$ at fixed time $t = 5$. As $\bar{\omega}$ increases all c_{mn} decrease, but c_{00} increases.

off-diagonal c_{mn} because they can be imaginary. All the units are dimensionless. Dimensionless frequency is given as $\bar{\omega}R_S$, while dimensionless time is given as t/R_S. From the plot one can see that the coefficient c_{00} is initially almost exactly 1, but then it decreases with time. The higher terms start small but then they increase with time, reach their maximum value and then they decrease. This is expected because the system starts in the ground state. As time progresses more modes are excited and higher order terms increase in magnitude. This increase of higher order terms cannot proceed indefinitely if unitarity is preserved; i.e., any increase must be accompanied by a decrease somewhere else. On the same plot we show the trace of the density matrix $\mathrm{Tr}(\hat{\rho})$ as a check. The trace must remain unity at all times to preserve probabilities. However, we can numerically take into account only a finite number of modes. Therefore, at some late time, the trace will start decreasing on the graph since higher modes which have not been included in numerics will become important. The more modes we include, the longer the trace will remain unity. In the Supplemental Material [18] we proved that if one takes $n \to \infty$, then $\mathrm{Tr}(\hat{\rho})$ always remains unity. Hence we plotted the graph only up to the time when $\mathrm{Tr}(\hat{\rho})$ remains one.

What is more important is that the magnitudes of the off-diagonal terms also increase with time. This implies that correlations among the created particles increase with time up to the point when even higher order terms start increasing. Since there are progressively more cross terms than the diagonal terms, their cumulative contribution to the total density matrix simply cannot be neglected. In Fig. 2, we plotted c_{mn} and $\mathrm{Tr}(\hat{\rho})$ as a function of $\bar{\omega}$ at a constant time. $\mathrm{Tr}(\hat{\rho})$ remains 1 for all frequencies. The lowest term c_{00} increases with $\bar{\omega}$, but all other terms decrease. This means that the lowest diagonal term dominates and correlations are not that important at high frequencies. Information content in the system is usually given in terms of a trace of the squared density matrix. If the trace of the squared density matrix is 1, then the state is pure, while the zero trace corresponds to a mixed state. In Fig. 3, we plot the traces of squares of two density matrices as functions of time for a fixed frequency. One is the Hawking density matrix $\hat{\rho}_h$, which contains only the diagonal terms c_{nn} and neglects correlations. The other one is the total density matrix $\hat{\rho}$ defined in Eq. (37), which contains all the elements, including the off-diagonal correlations. As expected, $\mathrm{Tr}(\hat{\rho}_h^2)$ goes to zero as time progresses, which means that the system is going from a pure state to a maximally mixed thermal state. This is

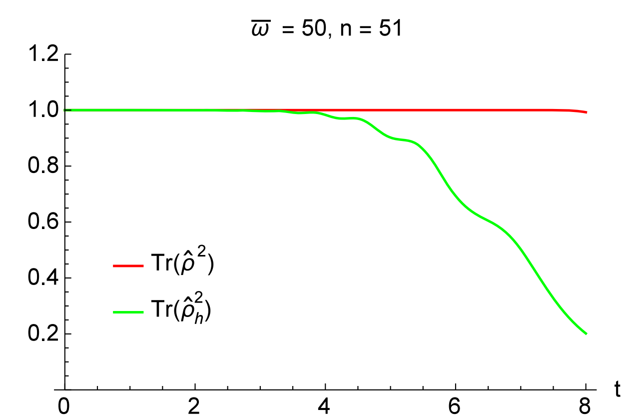

FIG. 3 (color online). $\hat{\rho}_h$ is the diagonal Hawking density matrix, ρ is the total density matrix as in Eq. (37). We plot $\mathrm{Tr}(\hat{\rho}^2)$ and $\mathrm{Tr}(\hat{\rho}_h^2)$ as functions of time at a fixed frequency $\bar{\omega} = 50$. The magnitude of $\mathrm{Tr}(\hat{\rho}_h^2)$ decreases with time, meaning that the system is losing information by going from a pure to a mixed state. However, $\mathrm{Tr}(\hat{\rho}^2)$ remains unity at all times, which means that the state remains pure. This implies that the information of the system is conserved if cross-correlations are accounted for.

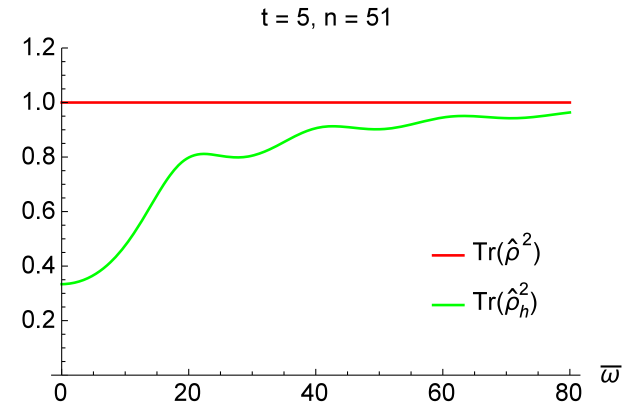

FIG. 4 (color online). $\mathrm{Tr}(\hat{\rho}^2)$ and $\mathrm{Tr}(\hat{\rho}_h^2)$ as a function of $\bar{\omega}$ at $t = 5$. Again $\mathrm{Tr}(\hat{\rho}^2)$ remains one at all frequencies, but $\mathrm{Tr}(\hat{\rho}_h^2)$ differs from unity at low frequencies which means it does not account for the full information of the system at low frequencies.

often labeled as the information loss in the process of Hawking radiation. However, if we plot the total $\mathrm{Tr}(\hat{\rho}^2)$ we see that it always remains unity, which means that the state always remain pure during the evolution and information does not get lost. This clearly tells us that correlations between the excited modes are very important, and if one takes them into account, the information in the system remains intact. In Fig. 4, we plot $\mathrm{Tr}(\hat{\rho}^2)$ and $\mathrm{Tr}(\hat{\rho}_h^2)$ as a function of $\bar{\omega}$ at a fixed time. As expected, $\mathrm{Tr}(\hat{\rho}^2)$ remains 1 at all frequencies, but $\mathrm{Tr}(\hat{\rho}_h^2)$ differs from unity at low frequencies. This implies that ρ_h gives a good description of the system at high frequencies, but it fails to do so at low frequencies.

Conclusions.—In conclusion, we showed by explicit calculations that radiation coming from a collapsing object is manifestly unitary. Hawking's thermal density matrix is diagonal and inevitably leads to information loss. However, when we take the off-diagonal correlation terms into account, the density matrix describes a pure state at all times. This result agrees well with [22], where it was shown at that at late enough time all the information in the system is contained in correlations between the small subsystems (in this case emitted particles). Our analysis was done for a static outside observer; however, it will be very important to learn what an infalling observer would see during the collapse in order to settle the question of information loss.

This work was also partially supported by the U.S. National Science Foundation, under Grants No. PHY-1066278 and No. PHY-1417317.

[1] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975); **46**, 206 (1976).
[2] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).
[3] D. N. Page, Phys. Rev. Lett. **44**, 301 (1980).
[4] P. Kraus and F. Wilczek, Nucl. Phys. **B433**, 403 (1995).
[5] T. Vachaspati, D. Stojkovic, and L. M. Krauss, Phys. Rev. D **76**, 024005 (2007).
[6] T. Vachaspati and D. Stojkovic, Phys. Lett. B **663**, 107 (2008).
[7] E. Greenwood and D. Stojkovic, J. High Energy Phys. 09 (2009) 058.
[8] E. Greenwood, D. C. Dai, and D. Stojkovic, Phys. Lett. B **692**, 226 (2010).
[9] E. Greenwood and D. Stojkovic, J. High Energy Phys. 06 (2008) 042.
[10] J. E. Wang, E. Greenwood, and D. Stojkovic, Phys. Rev. D **80**, 124027 (2009).
[11] E. Halstead, J. Cosmol. Astropart. Phys. 08 (2013) 043.
[12] E. Greenwood, arXiv:1002.2433.
[13] E. Greenwood, J. Cosmol. Astropart. Phys. 01 (2010) 002.
[14] E. Greenwood, E. Halstead, and P. Hao, J. High Energy Phys. 02 (2010) 044.
[15] E. Greenwood, D. I. Podolsky, and G. D. Starkman, J. Cosmol. Astropart. Phys. 11 (2011) 024.
[16] T. Vachaspati, Classical Quantum Gravity **26**, 215007 (2009).
[17] E. Greenwood, arXiv:1402.4557.
[18] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.111301 for details about technical derivation of the analytic formula for the coefficient c_n which is often used in our calculations. We prove that the trace of the density matrix is always one as required. We show that the occupation number of produced particles is approximately Planckian.
[19] M. Kolopanis and T. Vachaspati, Phys. Rev. D **87**, 085041 (2013).
[20] J. Hutchinson and D. Stojkovic, arXiv:1307.5861.
[21] K. Papadodimas and S. Raju, J. High Energy Phys. 10 (2013) 212.
[22] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).