PTEP

Prog. Theor. Exp. Phys. **2015**, 123B04 (20 pages)
DOI: 10.1093/ptep/ptv170

The fall of the black hole firewall: natural nonmaximal entanglement for the Page curve

Masahiro Hotta[1,*] and Ayumu Sugita[2]

[1] *Department of Physics, Faculty of Science, Tohoku University, Sendai 980-8578, Japan*
[2] *Department of Applied Physics, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan*
[*] E-mail: hotta@tuhep.phys.tohoku.ac.jp

Received October 1, 2015; Accepted Novemeber 2, 2015; Published December 21, 2015

The black hole firewall conjecture is based on the Page curve hypothesis, which claims that entanglement between a black hole and its Hawking radiation is almost maximum. Adopting canonical typicality for nondegenerate systems with nonvanishing Hamiltonians, we show the entanglement becomes nonmaximal, and energetic singularities (firewalls) do not emerge for general systems. An evaporating old black hole must evolve in Gibbs states with exponentially small error probability after the Page time as long as the states are typical. This means that the ordinarily used microcanonical states are far from typical. The heat capacity computed from the Gibbs states should be nonnegative in general. However, the black hole heat capacity is actually negative due to the gravitational instability. Consequently the states are not typical until the last burst. This requires inevitable modification of the Page curve, which is based on the typicality argument. For static thermal pure states of a large AdS black hole and its Hawking radiation, the entanglement entropy equals the thermal entropy of the smaller system.

Subject Index B22, B30, E00, E01, E05

1. Introduction

The interesting possibility of black hole firewalls was proposed from the viewpoint of quantum information [1,2] and has attracted much attention. In the firewall conjecture, a black hole horizon is not a smooth region even for free-fall observers who attempt to pass through it. On the horizon the observers see highly energetic quantum walls (firewalls) before they collide against it and burn up.

Essentially the firewall conjecture is based on the Page curve hypothesis of black hole evaporation [3,4], and the hypothesis comes from the Lubkin–Lloyd–Pagels–Page theorem (LLPP theorem) [5–7]. According to the LLPP theorem, quantum entanglement between two macroscopic systems S_I and S_{II} is almost maximum in a typical pure state $|\Psi\rangle_{I,II}$ of the composite Hilbert space $\mathcal{H}_I \otimes \mathcal{H}_{II}$, assuming that dimension N_{II} of $\mathcal{H}_{II}$ is much larger than dimension N_I of $\mathcal{H}_I$. The reduced density operator (quantum state) $\hat{\rho}_I = \mathrm{Tr}_{II}\left[|\Psi\rangle_{I,II}\langle\Psi|_{I,II}\right]$ of S_I almost equals $\hat{I}/N_I$, where $\hat{I}$ is the unit matrix acting on $\mathcal{H}_I$. Inspired by this theorem, Page came up with a fascinating scenario for information leakage from evaporating black holes. He thinks that evaporation of a macroscopic black hole in an initial pure state is modeled by two quantum systems B and R with finite, but time dependent, dimensions N_B and N_R. B represents internal degrees of freedom of the black hole, and R represents the Hawking radiation out of the black hole. It may be possible that the finiteness of N_B and N_R is justified if quantum gravity is taken account of. Such a quantum effect may truncate the degrees of

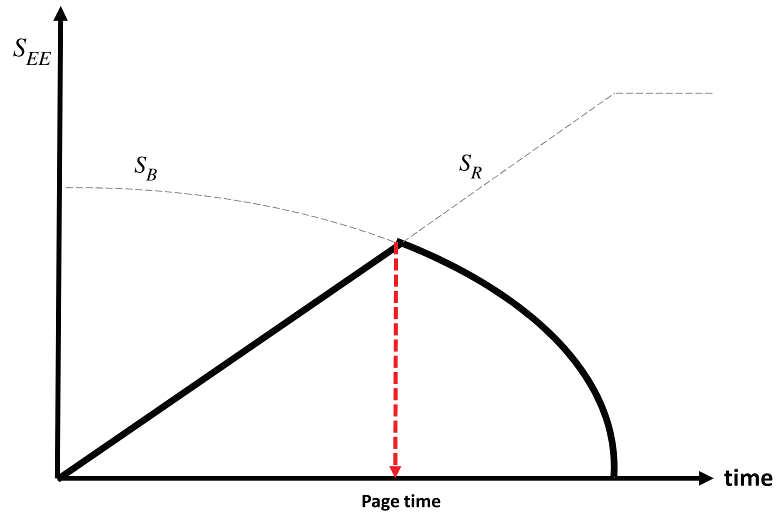

Fig. 1. Schematic figure of Page curve. In the conjecture, entanglement entropy between black hole and Hawking radiation equals thermal entropy of the smaller system and attains almost the maximum value at each time.

freedom in a higher energy scale than Planck energy, like string theory. In condensed matter physics, the total energy E is quite low. Thus high-energy density of states around the cutoff scale of the system becomes irrelevant. So it is enough to treat a finite-dimensional Hilbert space to describe the physics.

The essence of Page's hypothesis is summarized in the following propositions for entanglement between B and R:

(I) When $N_R \gg N_B$ (or $N_R \ll N_B$), B and R in a typical pure state of quantum gravity share almost maximal entanglement. In other words, a typical quantum state of the smaller system among B and R is almost proportional to the unit matrix $\hat{I}$.

(II) The entanglement entropy S_{EE} of the smaller system among B and R is equal to its thermal coarse-grained entropy.

Proposition (I) is clearly motivated by the LLPP theorem. Combining (I) and (II), it is deduced that S_{EE} between B and R takes almost the maximum value and equals the Bekenstein–Hawking entropy of black holes $S_B = \mathcal{A}/(4G)$ after the Page time, at which decreasing S_B equals increasing thermal entropy S_R of the Hawking radiation. The Page time is estimated as about 53% of the lifetime of evaporating black holes, and the mass at the Page time is about 77% of initial mass [3,4]. Thus the black hole remains macroscopic at the Page time, and its semi-classical picture is valid. Black holes after the Page time are referred to as "old." Before the Page time, S_{EE} is equal to S_R, and the black holes are referred to as "young." Since the time evolution of S_B and S_R is computed in an established semi-classical way, this argument provides a prediction for the time curve of S_{EE} during the evaporation. This is the Page curve. Its schematic figure is given in Fig. 1.

The firewall conjecture arises basically from (I). For an old black hole, Hawking radiation R is decomposed into A, which is emitted after the Page time, and C, which is emitted before the Page time. This is depicted in Fig. 2 for the gravitational collapse of a massless shell. The dimensions of sub-Hilbert spaces for A and C are denoted by N_A and N_C. Due to the old age of the black hole, $N_C \gg N_A N_B$ is satisfied. From (I), the AB system is almost maximally entangled with C. Thus a typical quantum state of AB can be approximated as $\hat{\rho}_{AB} \approx \hat{I}_{AB}/(N_A N_B)$. Since the unit matrix $\hat{I}_{AB}$

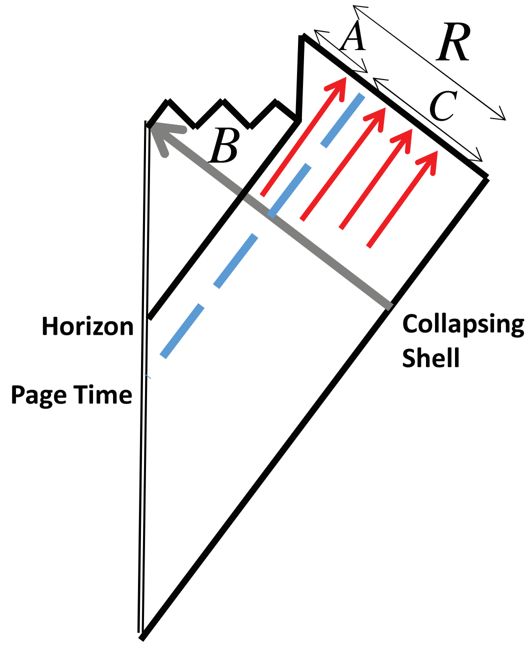

Fig. 2. Early Hawking radiation emitted before the Page time and late radiation emitted out of an old black hole after the Page time.

is written as $\hat{I}_A \otimes \hat{I}_B$, no correlation exists between A and B [8,9]. Consequently, for example, kinetic energy terms of quantum fields for the Hawking radiation diverge on the horizon. Let us denote an inside point x_B near the horizon, and an outside point x_A. Then a kinetic term $\left(\partial_x \hat{\varphi}\right)^2$ of a scalar field $\hat{\varphi}(x)$ is given by $\left(\hat{\varphi}(x_A) - \hat{\varphi}(x_B)\right)^2 / \epsilon^2$, where ϵ is the ultraviolet cutoff (lattice spacing). Apparently, when $\epsilon \to 0$, this diverges like $1/\epsilon^2$ on the horizon for the typical state $\hat{\rho}_{AB} \propto \hat{I}_A \otimes \hat{I}_B$ because of the correlation loss, and a firewall emerges.

Also, the strong subadditivity paradox [1,2,10] is often worried about in the context of the firewall paradox. Let us suppose a strong subadditivity inequality of the von Neumann entropy for an old black hole A, late radiation B, and early radiation C:

$$S_{AB} + S_{AC} \geq S_A + S_{ABC}. \tag{1}$$

Assuming the no-drama conditions $S_{AB} = 0$ and $S_{ABC} = S_C$ in Refs. [1,2,10] yields

$$S_{AC} \geq S_A + S_C.$$

As long as the old black hole continuously emits the stored information after the Page time, the purity of the AC system increases, and S_{AC} decreases in time. Thus $S_A > S_{AC}$ holds. This leads to an apparent contradiction that $S_C \leq 0$. Actually the early radiation has positive thermal entropy $S_C > 0$. This seems to mean a breakdown of the no-drama condition and suggests the existence of firewalls. However it is already known that the paradox is clearly avoided in moving mirror models [11] and long-lived remnant models [12]. In [11], it is pointed out that spacial locality among the subsystems A, B, and C is ill-defined. Consequently the AB system inevitably has nonvanishing entanglement with zero-point fluctuation of the radiation field, and $S_{AB} = 0$ does not hold even if we postulate the no-drama condition for the horizon in a physical sense. This remains true in black hole evaporation. For the long-lived remnant models, S_{AC} does not decrease even after the Page time, but rather increases until the last burst of the evaporating black hole. Thus $S_A > S_{AC}$ is not satisfied and the paradox is evaded. Similarly, any evaporation scenario, in which no information is emitted out of the black hole until the last burst, is free of the strong subadditivity paradox at least. Though the strong subadditivity paradox can be avoided, the long-lived remnant models [13–15] are supposed to have other flaws [16]. The energy of the remnant is of the order of the Planck mass, but in order to store the huge amount of information, the remnants seem to possess almost infinite degeneracy. The tremendous degeneracy may break the past great success of many experiments and observations via loop effects in particle scattering processes and species summation in partition functions for thermal equilibrium in the early universe [15,16]. In this paper, in order to avoid those flaws and firewalls simultaneously, we consider an alternative scenario. In the scenario, all the information comes out at the last burst. Of course, the total energy of the last ray out of the black hole is merely of the order of the Planck mass and very tiny. However, as stressed first by Wilczek [17], an outgoing zero-point fluctuation flow of quantum fields, which is adjacent to the last ray, can share a huge amount of entanglement with the Hawking radiation emitted before. This fluctuation flow in a local vacuum region has zero energy, but transports the information to the future null infinity without any contradiction. At the last burst, quantum gravity critically affects the horizon. Hence the no-drama condition is no longer required. Thus we do not need to care about the strong subadditivity paradox even if the entropy S_{AC} suddenly decreases at the last burst.

In this paper, first of all, it is pointed out that proposition (I) does not hold if nondegeneracy of energy eigenstates of the total system is taken account of. The typical states have to be exponentially close to Gibbs states with finite temperatures. The entanglement between AB and C becomes non-maximal. Therefore, without breaking monogamy of entanglement, A is able to share entanglement with B, and simultaneously with C. The entanglement between A and B yields a correlation that makes the horizon smooth, and no firewall appears. Though it has been proven that such nonmaximal entanglement prevents the emergence of firewalls in moving mirror models [11], more stringent arguments are provided for general systems in this paper. Our result means that the ordinarily used microcanonical states in the arguments of [1–4] are far from typical for quantum entanglement between a black hole and its Hawking radiation. In Sect. 2, we briefly review a general formulation of canonical typicality with nonzero Hamiltonians [18–23]. Our discussion is based on [19,20]. From the rigorous results, it turns out that proposition (I) does not hold for general systems which satisfy natural conditions. Therefore it turns out that, after the Page time, the evaporating old black hole must evolve in Gibbs states with high precision as long as the pure state of the black hole and the Hawking radiation is typical. In general, heat capacity computed from a partition function $Z(1/T)$ of a Gibbs state must be nonnegative, as

$$\frac{d\left\langle \hat{H}\right\rangle}{dT} = \frac{1}{T^2}\left\langle\left(\hat{H} - \left\langle \hat{H}\right\rangle\right)^2\right\rangle \geq 0,$$

where $\hat{H}$ is the Hamiltonian of the system, T is temperature, and $\langle\cdot\rangle = \mathrm{Tr}\left[\cdot \exp\left(-\hat{H}/T\right)\right]/Z(1/T)$. However, the black hole heat capacity is actually negative due to the gravitational instability. For instance, for a Schwarzschild black hole, the energy E is its mass M and equals $(8\pi GT)^{-1}$. The heat capacity is computed as negative: $\frac{dE}{dT} = -\left(8\pi GT^2\right)^{-1} < 0$. Thus the pure state of the system is never typical until the last burst. This leads to inevitable modification of the Page curve. In Sect. 3, we discuss proposition (II). In black hole evaporation, the proposition is implausible. From the viewpoint of semi-classical general relativity, it looks more fascinating to take an alternative for the Page curve. The entanglement entropy continues to increase even after the Page time. At the last burst, it suddenly goes to zero and all the information is retrieved. Finally, it is commented that the typical entanglement entropy of a large AdS black hole and its Hawking radiation equals the thermal entropy of the smaller system. We adopt natural units, $c = \hbar = k_B = 1$.

2. Nonmaximality of entanglement in canonical typicality

In this section, we claim that proposition (I) is not satisfied for general systems with nondegenerate Hamiltonian. When pure states are randomly sampled in a sub-Hilbert space with fixed total energy, a typical state is not maximally entangled. Then the corresponding state of the smaller subsystem is not the completely mixed state which is proportional to the unit matrix, but a Gibbs state. Let us consider two finite quantum systems S_1 and S_2, whose dimensions of Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are denoted by D_1 and D_2, respectively. Let us consider a pure state $|\Psi\rangle_{12}$ in $\mathcal{H}_1 \otimes \mathcal{H}_2$. Its density operator $\hat{\rho}_{12} = |\Psi\rangle_{12}\langle\Psi|_{12}$ is a $D_1 D_2 \times D_1 D_2$ Hermitian matrix. Thus it can be expanded uniquely in terms of a basis of $U(D_1 D_2)$ Hermitian generators $\left\{\hat{I} \otimes \hat{I}, \hat{G}_{n\mu}\right\}$:

$$\hat{\rho}_{12} = \frac{1}{D_1 D_2}\left(\hat{I} \otimes \hat{I} + \sum_{n\mu}\left\langle \hat{G}_{n\mu}\right\rangle \hat{G}_{n\mu}\right),$$

where $\hat{G}_{n\mu}$ are traceless and satisfy $\mathrm{Tr}\left[\hat{G}_{n\mu}\hat{G}_{n'\mu'}\right] = D_1 D_2 \delta_{nn'}$ and

$$\left\langle \hat{G}_{n\mu}\right\rangle = \mathrm{Tr}\left[\hat{\rho}_{12}\hat{G}_{n\mu}\right] = \left\langle \Psi\right|_{12}\hat{G}_{n\mu}\left|\Psi\right\rangle_{12}.$$

The set of $\hat{G}_{n\mu}$ consist of basis generators $\hat{T}_n$ and $\hat{R}_\mu$ for each sub-Hilbert space. $\hat{T}_n$ and $\hat{R}_\mu$ are traceless and Hermitian, and obey the following normalization:

$$\mathop{\mathrm{Tr}}_1\left[\hat{T}_n\hat{T}_{n'}\right] = D_1\delta_{nn'},$$
$$\mathop{\mathrm{Tr}}_2\left[\hat{R}_\mu\hat{R}_{\mu'}\right] = D_2\delta_{\mu\mu'}.$$

For $n = 1 \sim D_1^2 - 1$, $\hat{G}_{n0}$ is defined as

$$\hat{G}_{n0} = \hat{T}_n \otimes \hat{I}.$$

Similarly, for $\mu = 1 \sim D_2^2 - 1$, $\hat{G}_{0\mu}$ is defined as

$$\hat{G}_{0\mu} = \hat{I} \otimes \hat{R}_\mu.$$

The remaining generators are given by

$$\hat{G}_{n\mu} = \hat{G}_{n0}\hat{G}_{0\mu} = \hat{T}_n \otimes \hat{R}_\mu.$$

The reduced quantum state $\hat{\rho}_1$ for S_1 is computed as

$$\hat{\rho}_1 = \mathop{\mathrm{Tr}}_2\left[|\Psi\rangle_{12}\langle\Psi|_{12}\right].$$

It is worth noting that $\hat{\rho}_1$ is uniquely fixed by measuring only $D_1^2 - 1$ expectation values of $\hat{G}_{n0}$ with respect to $|\Psi\rangle_{12}$. This is because

$$\left\langle \hat{G}_{n0}\right\rangle = \left\langle \Psi\right|_{12}\hat{G}_{n0}\left|\Psi\right\rangle_{12} = \left\langle \Psi\right|_{12}\left(\hat{T}_n \otimes \hat{I}\right)\left|\Psi\right\rangle_{12} = \mathrm{Tr}_1\left[\hat{\rho}_1\hat{T}_n\right] = \left\langle \hat{T}_n\right\rangle$$

holds. In fact, $\hat{\rho}_1$ is written as

$$\hat{\rho}_1 = \frac{1}{D_1}\left(\hat{I} + \sum_n \left\langle \hat{T}_n \right\rangle \hat{T}_n\right) = \frac{1}{D_1}\left(\hat{I} + \sum_n \left\langle \hat{G}_{n0} \right\rangle \hat{T}_n\right).$$

In order to analyze the smaller system S_1, expectation values of other components of $\hat{G}_{n\mu}$ are not required.

Next let us consider a nondegenerate Hamiltonian $\hat{H}$ for the composite system. $\hat{H}$ takes its general form of

$$\hat{H} = \hat{H}_1 \otimes \hat{I} + \hat{I} \otimes \hat{H}_2 + \hat{V}_{12},$$

where $\hat{H}_1$ and $\hat{H}_2$ are free Hamiltonians, and $\hat{V}_{12}$ is an interaction term between the subsystems. For simplicity, we ignore $\otimes$ and $\hat{I}$ in later equations such that $\hat{H}_1 \otimes \hat{I}$ is abbreviated as $\hat{H}_1$. The normalized eigenstates of $\hat{H}$ with eigenvalue E_j are denoted by $|E_j\rangle$:

$$\hat{H}\left|E_j\right\rangle = E_j\left|E_j\right\rangle.$$

We define a set $\Delta(E)$ of energy indices for a macroscopically large total energy E and a positive number δ as

$$\Delta(E) = \left\{ j | E_j \in [E-\delta, E] \right\}.$$

Let us introduce a sub-Hilbert space $\mathcal{H}_{\Delta(E)}$, which is spanned by $\left\{ \left|E_j\right\rangle | j \in \Delta(E) \right\}$ and its dimension is denoted by D. A microcanonical energy shell is defined as the set of pure states in $\mathcal{H}_{\Delta(E)}$. It should be stressed that $\mathcal{H}_{\Delta(E)}$ is not a tensor product $\mathcal{H}_B \otimes \mathcal{H}_R$ of any sub-Hilbert spaces $\mathcal{H}_B$ of $\mathcal{H}_1$ and $\mathcal{H}_R$ of $\mathcal{H}_2$. In order to understand this, let us suppose a case in which $\hat{V}_{12}$ is negligibly small. Then $\mathcal{H}_{\Delta(E)}$ is spanned by $\{|E_1\rangle_1 |E-E_1\rangle_2\}$, where $|E_1\rangle_1$ is the eigenstate with eigenvalue E_1 of $\hat{H}_1$, and $|E-E_1\rangle_2$ is the eigenstate with eigenvalue $E-E_1$ of $\hat{H}_2$. However, $|E_1\rangle_1 |E-E_1'\rangle_2$ with $E_1 \neq E_1'$ is not included by $\mathcal{H}_{\Delta(E)}$. This clearly implies $\mathcal{H}_{\Delta(E)} \neq \mathcal{H}_B \otimes \mathcal{H}_R$. Hence, the tensor product structure assumption of the Page curve hypothesis is not appropriate for descriptions of black hole evaporation.

For ordinary physical systems with large volume V, D becomes exponentially large like $\exp(\gamma V)$ with a positive constant γ. Taking a small value of δ gives us a naive picture of the energy shell, which often appears in standard textbooks of statistical mechanics. Note that δ-dependence for the final results of statistical mechanics is irrelevant in general. In fact, δ does not necessarily have to be small in the later discussion, since the density of states $e^{S(E)}$, where S is the entropy, is a very rapidly increasing function and the eigenstates close to the upper bound E give a dominant contribution, as depicted in Fig. 3. Therefore, for simplicity, it is also possible to take an energy shell, say, $[0, E]$ for $\Delta(E)$ instead of $[E-\delta, E]$.

Any pure state in $\mathcal{H}_{\Delta(E)}$ is written as

$$|\Psi\rangle_{12} = \sum_{j\in\Delta(E)} c_j \left|E_j\right\rangle, \tag{2}$$

where c_j satisfy the normalization condition $\sum_{j\in\Delta(E)} \left|c_j\right|^2 = 1$. In order to analyze canonical typicality for $\mathcal{H}_{\Delta(E)}$, let us introduce a uniform probability distribution for c_j as

$$p\left(c_1, \ldots, c_D\right) = \frac{\Gamma(D)}{\pi^D}\delta\left(\sum_{j\in\Delta(E)} \left|c_j\right|^2 - 1\right),$$

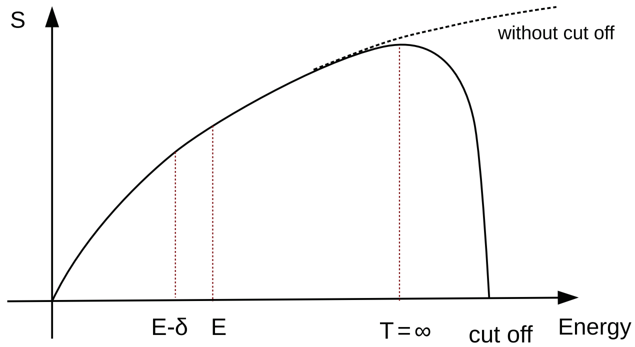

Fig. 3. Behavior of entropy in ordinary systems. When we focus on an energy shell $[E-\delta, E]$, the cutoff dependence becomes irrelevant for the canonical typicality argument.

such that $\int p\,(c_1,\ldots,c_D)\,d^Dc = 1$. The ensemble average value of a function f of c_j with respect to a unit sphere of $\mathcal{H}_{\Delta(E)}$ (microcanonical energy shell) is computed as

$$\overline{f} = \int f\,(c_1,\ldots,c_D)\,p\,(c_1,\ldots,c_D)\,d^Dc.$$

The ensemble average of a quantum expectation value $\left\langle \hat{O} \right\rangle$ of an observable $\hat{O}$ in $|\Psi\rangle_{12}$ in Eq. (2) is denoted by $\overline{\left\langle \hat{O} \right\rangle}$. The statistical deviation from $\overline{\left\langle \hat{O} \right\rangle}$ is given by

$$\delta\left\langle \hat{O} \right\rangle = \left\langle \hat{O} \right\rangle - \overline{\left\langle \hat{O} \right\rangle}.$$

As proven in the Appendix, the ensemble mean square error of $\left\langle \hat{O} \right\rangle$ is bounded above as

$$\overline{\left(\delta\left\langle \hat{O} \right\rangle\right)^2} = \overline{\left\langle \hat{O} \right\rangle^2} - \overline{\left\langle \hat{O} \right\rangle}^2 \leq \frac{\left\| \hat{O}^2 \right\|}{D+1}, \tag{3}$$

where the operator norm $\left\| \hat{O}^2 \right\|$ represents the maximum absolute value of the eigenvalues of $\hat{O}^2$. By taking $\hat{O} = \hat{G}_{n0}$, we have [19,20]

$$\overline{\left(\delta\hat{G}_{n0}\right)^2} \leq \frac{\left\| \hat{G}^2_{n0} \right\|}{D+1}. \tag{4}$$

It should be stressed that $\left\| \hat{G}^2_{n0} \right\| \left(= \left\| \hat{T}^2_n \right\| \right)$ is independent of D_2, though D grows exponentially as $\exp\left(\gamma V_2(D_2)\right)$ with respect to the volume $V_2(D_2)$ of S_2. Hence the right-hand side in Eq. (4) becomes negligibly small as $\exp\left(-\gamma V_2(D_2)\right)$ for large D_2 with D_1 fixed. Because the statistical fluctuation is so small, typical values of $\left\langle \hat{G}_{n0} \right\rangle$ are very close to the central value $\overline{\left\langle \hat{G}_{n0} \right\rangle}$. This implies that $\hat{\rho}_1$ for typical $|\Psi\rangle_{12}$ in $\mathcal{H}_{\Delta(E)}$ coincides with its ensemble average state $\overline{\hat{\rho}_1}$ almost certainly. In fact, we are able to prove that the ensemble deviation of $\hat{\rho}_1$ is estimated [19,20] as

$$\overline{\operatorname*{Tr}_1\left[\left(\hat{\rho}_1 - \overline{\hat{\rho}_1}\right)^2\right]} \leq \frac{1}{D_1\,(D+1)} \sum_n \left\| \hat{G}^2_{n0} \right\|, \tag{5}$$

the right-hand side of which decays rapidly due to D divergence as D_2 becomes large. Note that $\overline{\hat{\rho}_1}$ is given by

$$\overline{\hat{\rho}_1} = \frac{1}{D_1}\left(\hat{I} + \sum_n \overline{\left\langle \hat{G}_{n0} \right\rangle} \hat{T}_n \right).$$

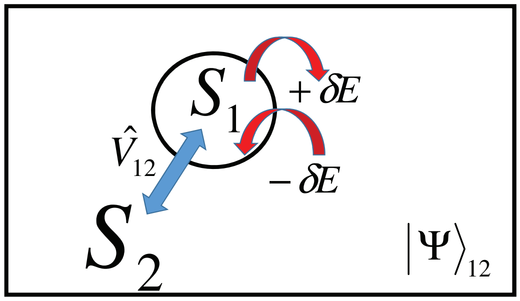

Fig. 4. A canonical typicality setup for two general systems that preserves total energy and allows energy transportation between subsystems bidirectionally.

Due to $\hat{V}_{12}$, energy is exchanged between S_1 and S_2 as shown in Fig. 4. If the contribution of $\hat{V}_{12}$ is negligibly small compared to $\hat{H}_1$ and $\hat{H}_2$, the sum, $\hat{H}_1 + \hat{H}_2$, is approximately conserved. Then, as proven in many textbooks of statistical mechanics, $\overline{\hat{\rho}_1}$ becomes a Gibbs state with a fixed temperature with high precision for $D_2 \gg D_1$:

$$\overline{\hat{\rho}_1} \approx \frac{1}{Z(\beta)} \exp\left(-\beta \hat{H}_1\right), \tag{6}$$

for ordinary physical systems. The difference between the typical state and the Gibbs state must be exponentially small: $\overline{\mathrm{Tr}_1\left[\left(\hat{\rho}_1 - \overline{\hat{\rho}_1}\right)^2\right]} \leq C \exp(-\gamma V)$. The inverse temperature β is determined by the total energy E, which is much less than the cutoff energy scale for black hole evaporation. Contrary to proposition (I), Eq. (6) shows that $\overline{\hat{\rho}_1}$ is not proportional to $\hat{I}_1$ if the temperature is finite, unless $\hat{H}_1 = 0$. Note that a microcanonical state $\hat{\rho}_m$, which is proportional to the projection operator $\hat{I}_E$ onto the microcanonical energy shell, is far from typical, even though its von Neumann entropy I_m is $O(V)$ and the difference from the Gibbs state entropy I_c is merely $O(\ln V)$. For the von Neumann entropy I_t of a typical state, $|I_t - I_c|$ is not $O(\ln V)$, but exponentially small as $O(\exp(-\gamma V))$.[1] The entropy difference $|I_m - I_c| = O(\ln V)$ is too large to regard the microcanonical state as a typical state. Nondegeneracy of $\hat{H}$ provides nonmaximal entanglement between S_1 and S_2 to elude the strong subadditivity paradox. It has been pointed out [11] that such a nonmaximal entanglement appears in moving mirror models and avoids firewalls. The above argument extends the moving mirror result to general ones. This completely removes the reason for black hole firewall emergence in Refs. [1,2].

For evaporating old black holes, Eq. (6) implies that state superposition of different energy black holes emerges in the total pure state of B and R, and generates much more entanglement, compared to a single black hole contribution with a fixed energy in the Page curve hypothesis. It should be noted that when the Gibbs state in Eq. (6) is thermodynamically unstable due to the emergence of negative heat capacity, just like for asymptotically flat black hole spacetimes [25], the typicality argument itself is unable to be applied to black hole evaporation, and never provides any correct insight. Quantum states in the evaporation are nontypical all the time, and proposition (I) loses its reasoning.

The result of canonical typicality with Eq. (6) has already been commented on by Harlow for a weak interaction limit [9]. However, it should be emphasized that Harlow does not give any proof of the typicality. Harlow also pointed out [9] that there remains a subtlety for the firewall removal

[1] Note that $\overline{\hat{\rho}_1}$ is not exactly the same as the Gibbs state $e^{-\beta \hat{H}_1}/Z$, because of a small correction due to the interaction term $\hat{V}_{12}$. Therefore, precisely speaking, I_c should be regarded as the von Neumann entropy of $\overline{\hat{\rho}_1}$.

even if we accept the canonical typicality. Let us recall the setup of gravitational collapse in Fig. 2. Naively, it may be expected that the quantum state $\overline{\hat{\rho}_{AB}}$ of late radiation A and black hole B for a typical state is approximated by

$$\frac{1}{Z(\beta)}\exp\left(-\beta\left(\hat{H}_A+\hat{H}_B\right)\right)=\frac{1}{Z(\beta)}\exp\left(-\beta\hat{H}_A\right)\otimes\exp\left(-\beta\hat{H}_B\right). \tag{7}$$

The above tensor product structure of the state means no correlation between A and B even after taking account of $\hat{H}$ nondegeneracy. Therefore, just as for the original story of firewalls, the expectation value of the kinetic energy density of radiation fields might be divergent. However, we claim that the worry is useless. In order to see no divergence, we go back to the above general formulation. The system Hamiltonian is given by

$$\hat{H}=\hat{H}_A+\hat{H}_B+\hat{H}_C+\hat{V}_{AB}+\hat{V}_{AC}+\hat{V}_{BC}+\hat{V}_{ABC},$$

where $\hat{H}_A$, $\hat{H}_B$, $\hat{H}_C$ are free Hamiltonians for the late radiation, the black hole, and the early radiation, $\hat{V}_{AB}$, $\hat{V}_{AC}$, $\hat{V}_{BC}$ are two-body interactions among A, B, C, and $\hat{V}_{ABC}$ is a three-body interaction (if we have any). The setup is depicted in Fig. 5. Let us consider that $\hat{V}_{AC}+\hat{V}_{BC}+\hat{V}_{ABC}$ are negligibly small as usual in statistical mechanics setups. However, we do not necessarily assume that $\hat{V}_{AB}$ is small. Then $\overline{\hat{\rho}_{AB}}$ does not take the form of Eq. (7), but instead

$$\overline{\hat{\rho}_{AB}}=\frac{\exp\left(-\beta\left(\hat{H}_A+\hat{H}_B+\hat{V}_{AB}\right)\right)}{\mathrm{Tr}_{AB}\left[\exp\left(-\beta\left(\hat{H}_A+\hat{H}_B+\hat{V}_{AB}\right)\right)\right]}. \tag{8}$$

The expression of Eq. (8) is correct irrespective of the interaction strength between A and B. The expectation value of $\hat{V}_{AB}$, which includes the kinetic energy term of radiation fields on the horizon, does not diverge at all:

$$\left|\mathrm{Tr}_{AB}\left[\overline{\hat{\rho}_{AB}}\hat{V}_{AB}\right]\right|<\infty.$$

It can be easily understood if we notice that the decomposition of the AB system into two parts (A and B) is arbitrary. If we choose the boundary of the two systems differently, we have different subsystems A' and B' and different free Hamiltonians $\hat{H}'_{A'}$, $\hat{H}'_{B'}$ and interaction $\hat{V}'_{A'B'}$ between them. But the physics of the composite system does not change at all because

$$\hat{H}_A+\hat{H}_B+\hat{V}_{AB}=\hat{H}'_{A'}+\hat{H}'_{B'}+\hat{V}'_{A'B'}$$

holds in Eq. (8). From the viewpoint of A' and B', $\hat{V}_{AB}$ is an ordinary local operator of A' or B'. There is no cause to make $\mathrm{Tr}_{AB}\left[\overline{\hat{\rho}_{AB}}\hat{V}_{AB}\right]\left(=\mathrm{Tr}_{A'B'}\left[\overline{\hat{\rho}_{A'B'}}\hat{V}_{AB}\right]\right)$ diverge. This remains true even if we take the limit of $\hat{V}_{AB}\to 0$ and recover the expression in Eq. (7). After all, we have no reasoning for firewall emergence on the horizon from the viewpoint of quantum information.

Here it is worth emphasizing that the LLPP theorem can be regarded as a special case of the canonical typicality. When we consider a nondegenerate Hamiltonian with a cutoff, the density of states $e^{S(E)}$ has the maximum value at an energy close to the cutoff (see Fig. 3), and the temperature is infinite at this point since $T^{-1}=\beta=\frac{\partial S}{\partial E}=0$. If we choose a state randomly from the whole Hilbert space without any energy condition, we almost always get a state with energy corresponding to the maximum density of states. Then the subsystem has the Gibbs state with $\beta=0$, which is proportional to the identity operator. This is exactly what the LLPP theorem claims. Consequently, we obtain a very high energy state with infinite temperature. Therefore the firewall argument based on the LLPP theorem actually says that not only the horizon, but also the whole space is on fire.

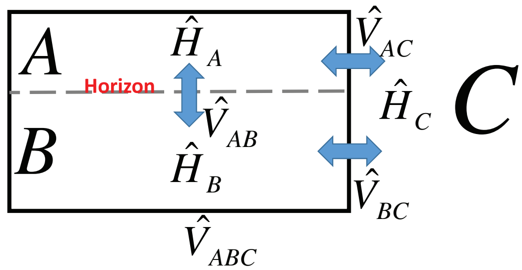

Fig. 5. A canonical typicality setup for an old black hole, early radiation, and late radiation, which allows all energy exchange among them, preserving total energy.

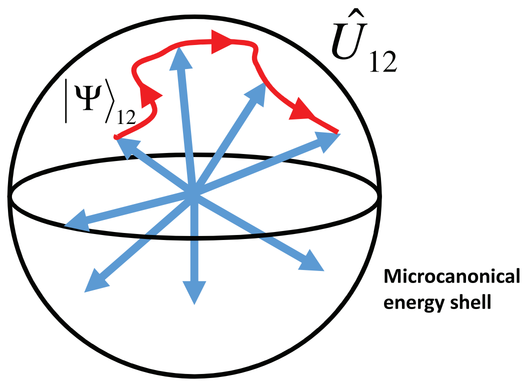

Fig. 6. Fast scrambling into typical states uniformly distributed all over the energy shell.

(Note that the argument in [1,2] does not use any special property of the horizon. Therefore it can be applied to an arbitrary partition of the space.) This wrong conclusion teaches us the importance of introducing a physical Hamiltonian and the energy conservation law to consider this problem.

3. Possible modification of the Page curve

In this section, we revisit Page's proposition (II). In Sect. 2, we explained that the quantum state $\hat{\rho}_1$ of the smaller system S_1 in a typical state $|\Psi\rangle_{12}$ of $\mathcal{H}_{\Delta(E)}$ equals a Gibbs thermal state. Because entanglement entropy is defined as $S_{EE} = -\mathrm{Tr}_1\left[\hat{\rho}_1 \ln \hat{\rho}_1\right]$, S_{EE} is actually the same as thermal entropy when S_1 and S_2 exchange energy via the boundary, and interaction $\hat{V}_{12}$ is negligibly small compared to the free Hamiltonians. This condition of small $\hat{V}_{12}$ really holds in nearest-neighbor interaction cases, because the free Hamiltonians are proportional to the volumes of S_1 and S_2, and $\hat{V}_{12}$ is merely proportional to the boundary area. Thus proposition (II) might sound convincing. However, we should not forget a crucial condition which makes $\hat{\rho}_1$ a typical state. Energy has to be transported not only from S_1 to S_2, but also from S_2 to S_1. In such a situation, all the energy eigenstates in $\mathcal{H}_{\Delta(E)}$ are able to contribute on the same footing with each other, as depicted in Fig. 6.

$\hat{U}_{12}$ is expected to generate very complicated time evolution, and may yield fast scrambling of the system in $\mathcal{H}_{\Delta(E)}$. Here, "scrambling" means relaxation of nontypical initial states with zero entanglement into typical states with high entanglement. After the relaxation, it is very unlikely to find the system in a nontypical state again for ordinary systems. This makes the canonical typicality method in Sect. 2 promising for late time. Finding typical $\hat{\rho}_1$ makes sense after the relaxation. How about cases in which energy is transferred only from S_1 to S_2? In these cases, the energy transportation does not happen from S_2 to S_1, as depicted in Fig. 7.

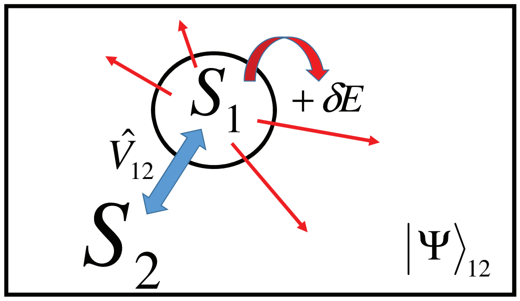

Fig. 7. An actual setup for black hole evaporation, in which energy is transferred only from the black hole to radiation. The energy transportation from outgoing Hawking radiation to the black hole does not take place.

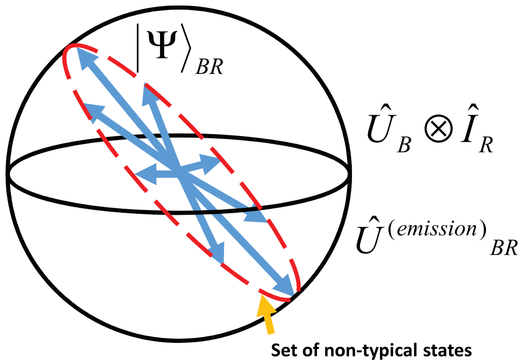

Fig. 8. A slow relaxation into typical states, which is generated by fast-scrambling inside the black hole and semi-classical radiation emission.

This actually arises in black hole evaporation, because the outgoing Hawking radiation emitted by black holes does not come back. The radiation is not able to give any amount of energy to the black holes without putting a mirror outside the horizon, or assuming thermal equilibrium. As well as the negative heat capacity of evaporating black holes, the one-way energy transportation is caused by the gravitational instability. Such a one-way dynamics of energy transportation makes the system remain in nontypical states before the last burst of the black hole. Generation of the Hawking radiation takes place in an outside region (a few times the black hole radius away from the horizon) with very small spacetime curvature. In an ordinary sense, the semi-classical treatment of the generation is justified, and the process is not random at all. Time evolution of B and R is described essentially by fast scrambling $\hat{U}_B \otimes \hat{I}_R$ of black holes and the nonrandom process $\hat{U}^{(\text{emission})}_{BR}$ of the Hawking radiation emission. $\hat{U}_B \otimes \hat{I}_R$ does not change the entanglement between B and R due to its locality. No fast scrambling for R happens. This nonchaotic setup does not ensure that the relaxation of the BR system finishes before the last burst. Therefore, in realistic evaporations, the validity of the typical state postulate becomes dubious. The system may always evolve among nontypical states as depicted in Fig. 8. Therefore the quantum state of the old black hole is able to be far from Gibbs states after the Page time. Hence S_{EE} can be totally different from thermal entropy, as opposed to Page's proposition (II).

Taking account of the possibility without (II), it is important to discuss modification of the Page curve. Moving mirror models, which mimic gravitational collapse and Hawking radiation emission out of black holes, may be a good device to see the possibilities. Let us consider a massless scalar field in $1+1$ dimensions. Adopting light cone coordinates $x^{\pm} = t \pm x$, the mirror trajectory is expressed as

$$x^+ = f\left(x^-\right),$$

where f is a monotonically increasing function of x^-. Even if the vacuum state is set as the initial state of the field, the mirror excites the field and emits radiation, whose expectation value of the outgoing energy flux is computed [26] as

$$\left\langle \hat{T}_{--}(x^-)\right\rangle = -\frac{1}{24\pi}\left[\frac{\partial^3_{x^-} f(x^-)}{\partial_{x^-} f(x^-)} - \frac{3}{2}\left(\frac{\partial^2_{x^-} f(x^-)}{\partial_{x^-} f(x^-)}\right)^2\right]. \tag{9}$$

A trajectory given by

$$f_o(x^-) = -\ln\left(1 + e^{-\kappa x^-}\right) \tag{10}$$

describes a mirror which is at rest in the past, and accelerates with constant acceleration κ in the future. This trajectory is related to a realistic $1+3$-dimensional gravitational collapse, which makes an eternal black hole without the back reaction of radiation emission [17]. Eventually the mirror emits constant thermal flux with temperature $T = \kappa/(2\pi)$. In fact, substitution of Eq. (10) into Eq. (9) yields the correct thermal flux,

$$\left\langle \hat{T}_{--}(x^-)\right\rangle = \frac{\pi}{12}T^2$$

for $x^- \gg 1/\kappa$. Now let us consider mirror trajectories which may approximately describe black hole evaporation with its back reaction. The first candidate is the following:

$$f_\kappa(x^-) = -\ln\left(\frac{1 + e^{-\kappa x^-}}{1 + e^{\kappa(x^- - h)}}\right), \tag{11}$$

where h is a very large real constant, and controls the lifetime of the corresponding black hole. The trajectory is depicted in Fig. 9. Due to the trajectory deformation, the mirror stops in the future. The time evolution of radiation emission is given by the plot in Fig. 10 for $\kappa = 1$ and $h = 500$. During the evaporation, almost constant flux is emitted, though real black holes increase the temperature and flux of radiation. It is interesting to compute the entanglement entropy S_{EE} between the field degrees of freedom inside $\left[x_1^-, x_2^-\right]$ and those outside $\left[x_1^-, x_2^-\right]$. There exists an ultraviolet divergence in S_{EE} due to the infinite number of degrees of freedom of the quantum field [27]. To remove the divergence, a renormalized entanglement entropy ΔS_{EE} is introduced by substituting the vacuum contribution [28]. ΔS_{EE} is given by

$$\Delta S_{EE} = \frac{1}{12}\ln\left(\frac{\left(f(x_2^-) - f(x_1^-)\right)^2}{\left(x_2^- - x_1^-\right)^2 \partial_{x^-} f(x_2^-)\partial_{x^-} f(x_1^-)}\right).$$

In Fig. 11, a plot of ΔS_{EE} as a function of x_2^- is provided for $\kappa = 1$, $h = 500$, and $x_1^- = -2$. The curve is almost symmetric and looks like the Page curve.[2] This aspect comes from the fact that the order of the mirror deceleration, which is expected due to back reaction of the radiation

[2] We thank Daniel Harlow for pointing out that the trajectory in Eq. (11) yields a Page-like curve for ΔS_{EE}.

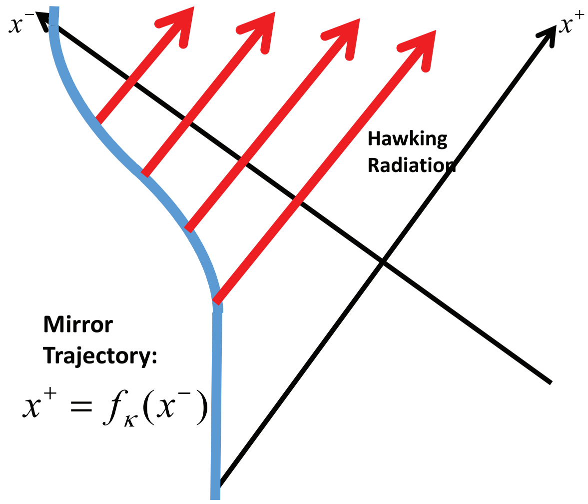

Fig. 9. Schematic figure of mirror trajectory described by Eq. (11).

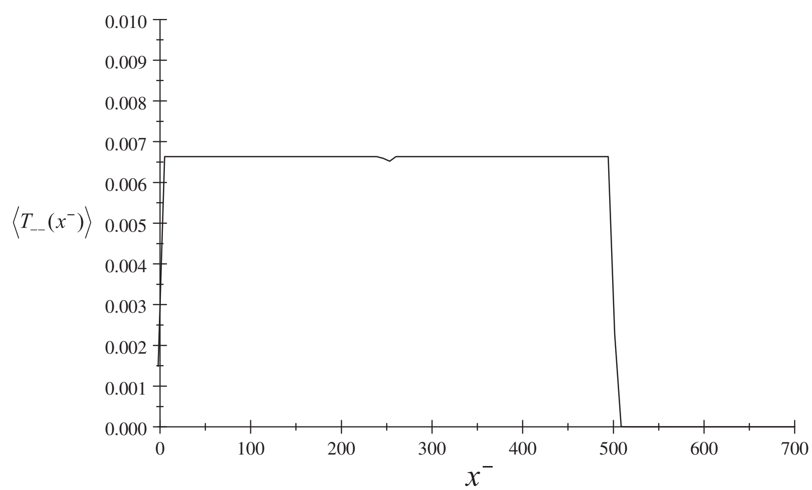

Fig. 10. Energy flux of Hawking radiation out of a mirror for the trajectory of Eq. (11) with $\kappa = 1$ and $h = 500$.

emission, is equal to κ of the acceleration in Eq. (11). However, this means locality breaking of dynamics for the evaporation and seems unlikely. κ corresponds to the scale of surface gravity of the $1+3$-dimensional black hole, and is considered as of the order of M_{pl}^2/M_{BH}, where M_{pl} is the Planck mass and M_{BH} is the black hole mass. Thus κ is very small compared to M_{pl} and the inverse of κ provides a cosmologically long time scale. The macroscopic black hole, whose evolution is described by Eq. (11), needs to estimate by itself its destiny, how much time remains before its death, and when the deceleration must start. At the half of its lifetime ($x^- \sim 250$), the black hole decides to emit its quantum information, which is stored inside the horizon, so as to finish leaking all the information before the last burst. In order to achieve this, the black hole has to slightly change its evolution at its Page time, which is long before its death, in a different way from those of other black holes with the same mass. For example, let us consider two black holes with the same mass M_{BH}, as depicted in Fig. 12. The left black hole in the figure is just born and very young. It has not begun the emission of Hawking radiation yet and is almost in a pure state. The right black hole in Fig. 12 was born with mass $1.3M_{BH}$ ($\sim M_{BH}/0.77$), and has decreased the mass to M_{BH} via radiation emission. Thus it is an old black hole around the Page time. It is worth stressing that the classical geometries

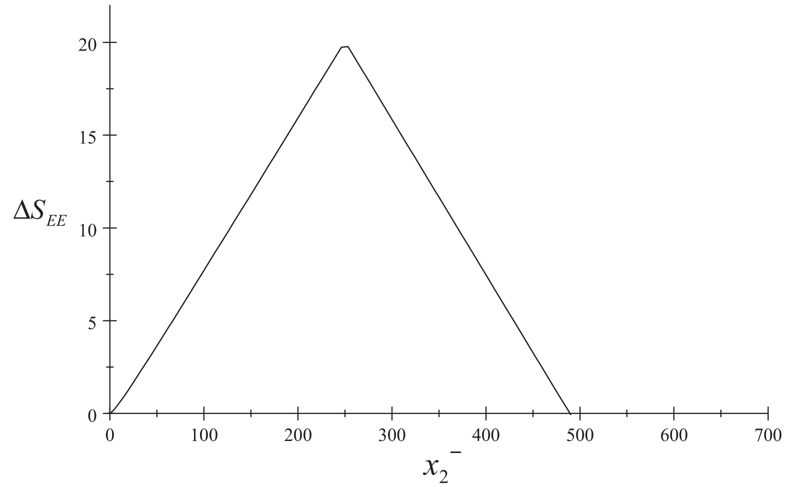

Fig. 11. Time curve of renormalized entanglement entropy between degrees of freedom in $[-2, x_2]$ and outside ones for the trajectory of Eq. (11) with $\kappa = 1$ and $h = 500$.

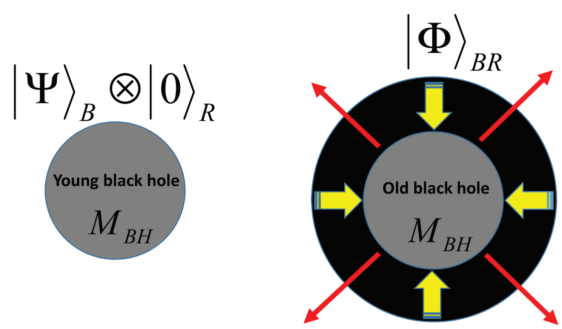

Fig. 12. Young and old black holes with the same mass. The entanglement evolutions are quite different from each other in the Page curve hypothesis.

of the two macroscopic black holes are the same. Nevertheless, only the old black hole begins to change its evolution at $x^- \sim 250$. The young black hole will change in a similar way further in the future. To perform such a cooperative motion, all quantum microscopic ingredients of the black hole must watch each other carefully and preserve the long-term memory. Each part must estimate the black hole age by use of the memory. This dynamics is non-Markovian and non-local, at least in time. If their dynamics can be approximated by the semi-classical general relativity, which is Markovian, such non-local evolution does not emerge. One might expect that the small back reaction of the Hawking radiation to the black hole geometry becomes a trigger to leak the information for the old black hole. However, if the black hole has such a sensitivity, the geometrical perturbation induced by a small amount of in-falling matter must also drastically change the time schedule of information leakage and completely modify the Page curve. This implies that the Page curve is unstable and useless for realistic situations.

Of course, we are able to consider a more natural mirror trajectory, which is consistent with semi-classical general relativity. One example is given by

$$f_{\kappa,\lambda}(x^-) = -\ln\left(\frac{1+e^{-\kappa x^-}}{1+e^{\lambda(x^- - h)}}\right). \tag{12}$$

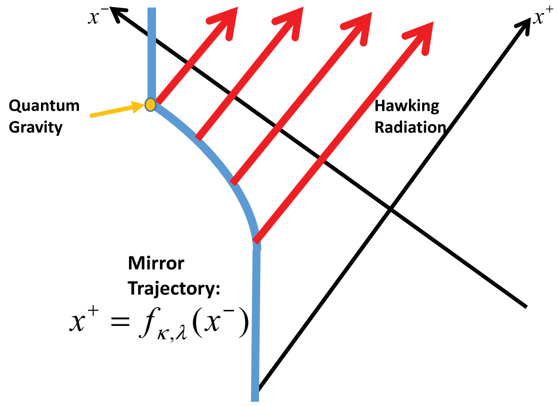

Fig. 13. Schematic figure of the mirror trajectory described by Eq. (12).

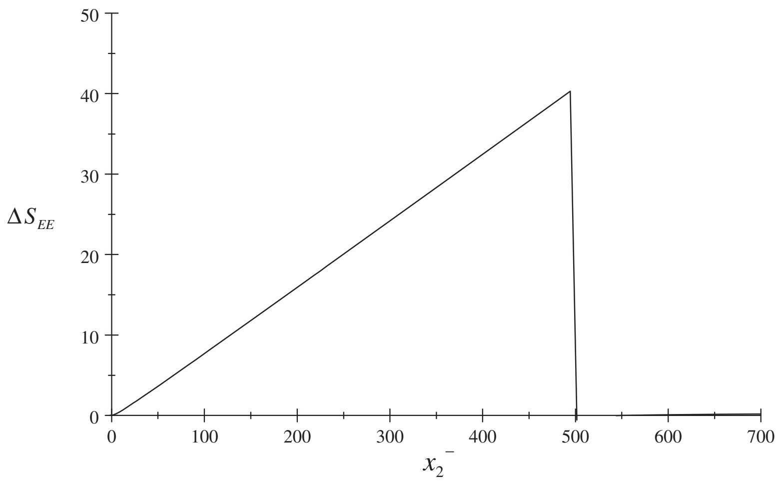

Fig. 14. Time curve of renormalized entanglement entropy between degrees of freedom in $[-2, x_2]$ and outside ones for the trajectory of Eq. (12) with $\kappa = 1$, $\lambda = 100$, and $h = 500$.

Here, two different scale parameters κ, λ are introduced. κ is the acceleration parameter for the emission of Hawking radiation. λ is the deceleration parameter, which is of Planck scale order M_{pl} and describes the sudden stop of the mirror due to the last burst of the black hole. Thus $\lambda \gg \kappa$ holds. The schematic behavior is given in Fig. 13. Clearly, the last burst region should be described by quantum gravity. In Fig. 14, ΔS_{EE} is plotted as a function of x_2^- for $\kappa = 1$, $\lambda = 100$, $h = 500$, and $x_1^- = -2$. Note that, until just before the last burst, ΔS_{EE} is estimated by the semi-classical results of the outside Hawking radiation without use of knowledge about quantum black holes. This aspect circumvents the uncertainty of quantum gravity, and strengthens the plausibility of this scenario. All the information is retrieved by the last burst. It is often commented that a huge amount of energy is required for such information leakage. However, it may be possible to attain it by use of entanglement with zero-point fluctuation of quantum fields [11,17,24], as already emphasized above. Hence energy of Planck mass order is enough to leak the information. In order to see the possibility, let us consider an entangled particle pair in the vacuum state [24], as depicted in Fig. 15. One of the particles becomes a Hawking particle with positive energy after scattering by the mirror. However, the partner particle is scattered by the mirror at rest and has no energy even after the scattering. Thus an enormous amount of quantum information, which makes the system state pure, is shared among

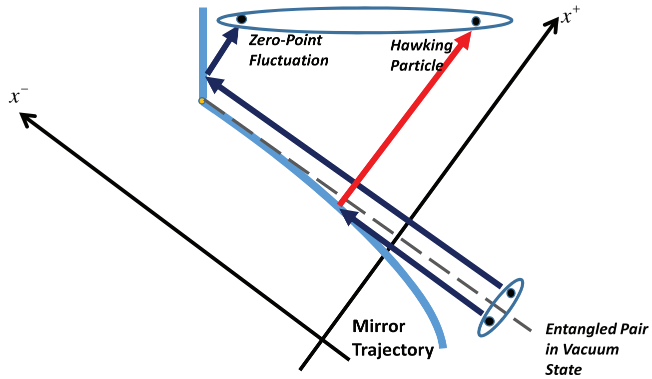

Fig. 15. Entangled particles scattered by mirror.

the Hawking radiation and the outgoing zero-point fluctuation flow of quantum fields in the future null infinity. When this scenario is applied to realistic black hole evaporation, we would expect that quantum gravity generates higher-derivative corrections to the Einstein equation and the smeared would-be singularity inside the black hole horizon becomes timelike and preserves the information of falling matter.

Here we add some comments. The famous qualitative bound for the information leakage time of Carlitz and Willey [14] is unable to be applied to the sudden stop case in Fig. 15, because their bound is derived by assuming a slow change of the acceleration of the mirror and ignoring emission of negative energy flux, which is generated with the informational zero-point fluctuation. Recently, Bianchi and Smerlak found an interesting identity between the energy flux emitted from moving mirrors and the entanglement entropy of the radiation [29]. They also argue that when applied to two-dimensional models of black hole evaporation, this identity implies that unitarity is incompatible with monotonic mass loss. However, it should be stressed that they assume the Page curve as a unitary model of black hole evaporation for the argument. Thus the outgoing zero-point fluctuation flow scenario, as well as the long-lived remnant scenario, is consistent with their claim. It seems natural that the singular sudden stop of the mirror trajectory in Fig. 15 may be smoothed around the Planck length scale by quantum gravity, as depicted in Fig. 16. Then a gravitational shock wave induced by the last burst ray may trap the entangled partner particle with zero energy for a while, just like in a high-energy gravitational scattering in a Minkowski vacuum. Even after re-emission of the particle from the shock wave, the energy of the partner particle may remain zero [30]. Thus the possibility is still alive that the information leakage by the last burst does not need a huge amount of energy in quantum gravity. This totally differs from the Page curve hypothesis, but is one interesting possibility.

Before closing this section, we add an interesting comment about a thermal equilibrium for the BR system. Instead of black hole evaporation, let us suppose a pure state which describes a static and stable thermal equilibrium of the composite system in a coarse-grained meaning [31,32]. In asymptotically anti-de Sitter spacetimes, there exist thermal equilibriums for a large black hole B and its Hawking radiation R. They exchange energy bidirectionally. By adiabatically slowly changing system parameters, including external forces and the position and pressure of a mirror surrounding R (if we have a mirror), various sizes of the black hole may appear. From the general results in Sect. 2, it turns out that each equilibrium state is typical, and the reduced state for the smaller subsystem among B and R is a Gibbs state with finite temperature, though no firewall emerges. Plotting entanglement entropy as a function of the inverse of black hole size generates a Page-like curve, in which

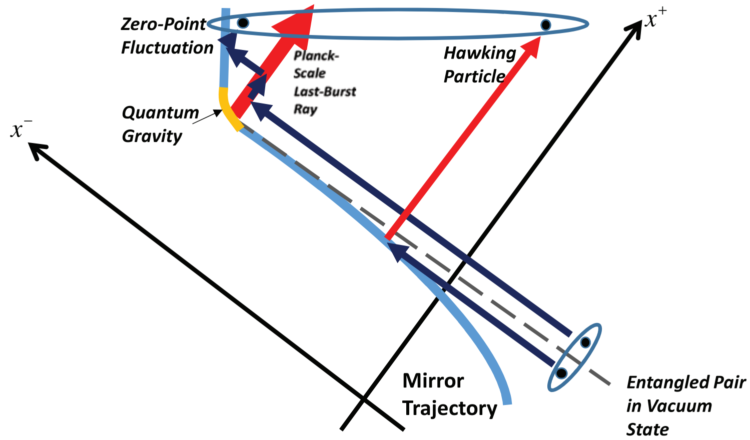

Fig. 16. Gravitational shock wave induced by the last burst ray and spacetime shift of zero-point fluctuation.

entanglement entropy equals thermal entropy for the smaller subsystem. This may become relevant in the future research of quantum black holes, though it is merely a side story for the original information loss problem.

4. Summary

In this paper, we revisit the Page curve hypothesis. Adopting a general formulation of canonical typicality with nondegenerate Hamiltonian, it is proven that Page's proposition (I) is not actually satisfied for ordinary systems. The typical states are exponentially close to Gibbs states with finite temperatures. The entanglement between subsystems becomes nonmaximal and removes firewalls. The microcanonical state, which is proportional to $\hat{I}_E$, is far from typical for the entanglement between a black hole and its Hawking radiation. In the dynamical situation of black hole evaporation, proposition (II) is also unlikely. We have no strong reason to expect that the entanglement entropy equals the thermal entropy for the smaller subsystem in the evaporation. Taking account of semi-classical general relativity, a conservative scenario becomes more fascinating in which all the information inside the black hole is emitted by the last burst of the black hole. Finally, it is pointed out, using the general results in Sect. 2, that for static thermal pure states of the BR system in the sense of canonical typicality, entanglement entropy between B and R certainly coincides with the thermal entropy of the smaller system. This holds for large AdS black holes.

Acknowledgements

We acknowledge support from the workshops, "Relativistic Quantum Information North 2014" at the University of Seoul and "Molecule-Type Workshop on Black Hole Information Loss Paradox" at YITP in Kyoto. We would like to thank Bill Unruh, Don Page, Daniel Harlow, Yasusada Nambu, Misao Sasaki, Pijin Chen, Anzhong Wang, Yen Chin Ong, Dong-hang Yeom, Shinji Mukohyama, and Ted Jacobson for related discussions. In particular, we appreciate a long and sincere discussion with Page and private communication with Harlow.

Appendix A. Proof of Eqs. (3) and (5)

Let us take a uniform distribution in the microcanonical energy shell $\mathcal{H}_{\Delta(E)}$ as

$$p\left(c_1, \ldots, c_D\right) = \frac{\Gamma\left(D - \frac{1}{2}\right)}{2\pi^{D-1/2}} \delta\left(\sum_{j\in\Delta(E)} \left|c_j\right|^2 - 1\right)$$

and introduce the ensemble average of a function of c_j as

$$\overline{f\left(c_1, \ldots, c_D\right)} = \int f\left(c_1, \ldots, c_D\right) p\left(c_1, \ldots, c_D\right) d^D c.$$

Using invariant tensors of $U(D)$, it is easy to show that

$$\overline{c_j^* c_{j'}} = \frac{1}{D} \delta_{jj'}, \tag{A1}$$

$$\overline{c_j^* c_{j'} c_k^* c_{k'}} = \frac{1}{D(D+1)} \left(\delta_{jj'}\delta_{kk'} + \delta_{jk'}\delta_{kj'}\right). \tag{A2}$$

A pure state of $\mathcal{H}_{\Delta(E)}$ is given by

$$\hat{\rho}\left(c_1, \ldots, c_D\right) = |\Psi\rangle\langle\Psi| = \sum_{j,j' \in \Delta(E)} c_j c_{j'}^* |E_j\rangle\langle E_{j'}|.$$

Using Eq. (A1), the ensemble average of the expectation value of observable $\hat{O}$ is computed as

$$\overline{\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle} = \frac{1}{D} \sum_{j \in \Delta(E)} \langle E_j | \hat{O} | E_j \rangle.$$

For the ensemble deviation,

$$\delta\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle = \left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle - \overline{\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle},$$

the mean square error is given by

$$\overline{\left(\delta\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle\right)^2} = \overline{\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle^2} - \left(\overline{\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle}\right)^2.$$

Using Eq. (A2), we have

$$\begin{aligned}
\overline{\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle^2} &= \sum_{j,j' \in \Delta(E)} \sum_{k,k' \in \Delta(E)} \overline{c_j^* c_{j'} c_k^* c_{k'}} \langle E_j | \hat{O} | E_{j'} \rangle \langle E_k | \hat{O} | E_{k'} \rangle \\
&= \frac{1}{D(D+1)} \sum_{j,k \in \Delta(E)} \left(\langle E_j | \hat{O} | E_j \rangle \langle E_k | \hat{O} | E_k \rangle + \langle E_j | \hat{O} | E_k \rangle \langle E_k | \hat{O} | E_j \rangle \right) \\
&= \frac{D}{D+1} \overline{\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle}^2 + \frac{1}{D(D+1)} \sum_{j,k \in \Delta(E)} \left| \langle E_j | \hat{O} | E_k \rangle \right|^2.
\end{aligned}$$

Thus the mean square error is estimated as

$$\overline{\left(\delta\left\langle \hat{O}\left(c_1, \ldots, c_D\right)\right\rangle\right)^2} = \frac{1}{D(D+1)} \sum_{j,k \in \Delta(E)} \left| \langle E_j | \hat{O} | E_k \rangle \right|^2 - \frac{1}{D^2(D+1)} \left(\sum_{j \in \Delta(E)} \langle E_j | \hat{O} | E_j \rangle \right)^2.$$

Because $\left(\sum_{j\in\Delta(E)}\langle E_j|\hat{O}|E_j\rangle\right)^2$ and $\sum_{k\notin\Delta(E)}\left|\langle E_j|\hat{O}|E_k\rangle\right|^2$ are nonnegative, we get

$$\overline{\left(\delta\left\langle\hat{O}\left(c_1,\ldots,c_D\right)\right\rangle\right)^2} \le \frac{1}{D(D+1)}\sum_{j\in\Delta(E)}\sum_{k\in\Delta(E)}\left|\langle E_j|\hat{O}|E_k\rangle\right|^2$$
$$\le \frac{1}{D(D+1)}\sum_{j\in\Delta(E)}\sum_{k=1}^{D_1D_2}\left|\langle E_j|\hat{O}|E_k\rangle\right|^2 \tag{A3}$$
$$= \frac{1}{D(D+1)}\sum_{j\in\Delta(E)}\langle E_j|\hat{O}^2|E_j\rangle$$
$$= \frac{1}{D+1}\overline{\left\langle\hat{O}\left(c_1,\ldots,c_D\right)^2\right\rangle}. \tag{A4}$$

Here we have used

$$\sum_{k=1}^{D_1D_2}\left|\langle E_j|\hat{O}|E_k\rangle\right|^2 = \langle E_j|\hat{O}\left(\sum_{k=1}^{D_1D_2}|E_k\rangle\langle E_k|\right)\hat{O}|E_j\rangle = \langle E_j|\hat{O}^2|E_j\rangle.$$

Because an expectation value does not exceed its operator norm,

$$\overline{\left\langle\hat{O}\left(c_1,\ldots,c_D\right)^2\right\rangle} \le \left\|\hat{O}^2\right\| \tag{A5}$$

is always satisfied. Combining Eqs. (A4) and (A5) yields Eq. (3).

The ensemble average state for S_1 is given by

$$\overline{\hat{\rho}_1} = \frac{1}{D_1}\left(\hat{I} + \sum_n \overline{\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle}\hat{T}_n\right).$$

The mean square error is computed as

$$\operatorname*{Tr}_1\left[\left(\hat{\rho}_1\left(c_1,\ldots,c_D\right) - \overline{\hat{\rho}_1}\right)^2\right]$$
$$= \frac{1}{D_1^2}\operatorname*{Tr}_1\left[\left(\sum_n\left(\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle - \overline{\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle}\right)\hat{T}_n\right)^2\right]$$
$$= \frac{1}{D_1}\sum_n\left(\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle - \overline{\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle}\right)^2.$$

Therefore we can manipulate it as follows:

$$\overline{\operatorname*{Tr}_1\left[\left(\hat{\rho}_1\left(c_1,\ldots,c_D\right) - \overline{\hat{\rho}_1\left(c_1,\ldots,c_D\right)}\right)^2\right]}$$
$$= \frac{1}{D_1}\sum_n\left(\overline{\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle^2} - \overline{\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle}^2\right)$$
$$\le \frac{1}{D_1}\sum_n\overline{\left(\delta\left\langle\hat{G}_{n0}\left(c_1,\ldots,c_D\right)\right\rangle\right)^2}$$
$$= \frac{1}{D_1\left(D+1\right)}\sum_n\left\|\hat{G}_{n0}^2\right\|.$$

In the last step, we used Eq. (3). Thus, Eq. (5) is proven.

References

[1] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, J. High Energy Phys. **02**, 062 (2013).
[2] S. L. Braunstein, Phys. Rev. Lett. **110**, 101301 (2013) [arXiv:0907.1190 [quant-ph]] [Search INSPIRE].
[3] D. N. Page, Proc. 5th Canadian Conf. General Relativity and Relativistic Astrophysics, p. 1 (1994).
[4] D. N. Page, Phys. Rev. Lett. **71**, 3743 (1993).
[5] E. Lubkin, J. Math. Phys. **19**, 1028 (1978).
[6] S. Lloyd and H. Pagels, Ann. Phys. (NY) **188**, 186 (1988).
[7] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).
[8] D. Harlow and P. Hayden, J. High Energy Phys. **1306**, 085 (2013).
[9] D. Harlow, [arXiv:1409.1231 [hep-th]] [Search INSPIRE].
[10] S. D. Mathur, Class. Quant. Grav. **26**, 224001 (2009).
[11] M. Hotta, J. Matsumoto, and K. Funo, Phys. Rev. D **89**, 124023 (2014).
[12] A. Almheiri and J. Sully, J. High Energy Phys. **108**, 1402 (2014).
[13] Y. Aharonov, A. Casher, and S. Nussinov, Phys. Lett. B **191**, 51 (1987).
[14] R. D. Carlitz and R. S. Willey, Phys. Rev. D **36**, 2336 (1987).
[15] S. B. Giddings, Phys. Rev. D **46**, 1347 (1992).
[16] J. Preskill, in *Black Holes, Membranes, Wormholes and Superstrings*, ed. S. Kalara and D. Nanopoulos (World Scientific, Singapore, 1992), p. 22.
[17] F. Wilczek, in *Black Holes, Membranes, Wormholes and Superstrings*, ed. S. Kalara and D. Nanopoulos (World Scientific, Singapore, 1992), p. 1.
[18] J. von Neumann, Z. Phys. **57**, 30 (1929).
[19] A. Sugita, RIMS Kokyuroku (Kyoto) **1507**, 147 (2006).
[20] A. Sugita, Nonlinear Phenom. Complex Syst. **10**, 192 (2007).
[21] S. Goldstein, J. L. Lebowitz, T. Tumulka, and N. Zanghi, Phys. Rev. Lett. **96**, 050403 (2006).
[22] S. Popescu, A. J. Short, and A. Winter, Nature Physics **2**, 754 (2006).
[23] P. Reimann, Phys. Rev. Lett. **101**, 190403 (2008).
[24] M. Hotta, R. Schüzhold, and W. G. Unruh, Phys. Rev. D **91**, 124060 (2015).
[25] S. W. Hawking and D. N. Page, Commun. Math. Phys. **87**, 577 (1983).
[26] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, 1982).
[27] M. Srednicki, Phys. Rev. Lett. **71**, 666 (1993).
[28] C. Holzhey, F. Larsen, and F. Wilczek, Nucl. Phys. B **424**, 443 (1994).
[29] E. Bianchi and M. Smerlak, Phys. Rev. D **90**, 041904(R) (2014).
[30] 't Hooft, Phys. Lett. B **198**, 61 (1987).
[31] S. Sugiura and A. Shimizu, Phys. Rev. Lett. **108**, 240401 (2012).
[32] S. Sugiura and A. Shimizu, Phys. Rev. Lett. **111**, 010401 (2013).