Eigenstate fluctuation theorem in the short- and long-time regimes

Eiki Iyoda,[1] Kazuya Kaneko,[2] and Takahiro Sagawa[2,3]

[1]*Department of Physics, Tokai University, 4-1-1 Kitakaname, Hiratsuka-shi, Kanagawa 259-1292, Japan*
[2]*Department of Applied Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[3]*Quantum-Phase Electronics Center (QPEC), The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

The canonical ensemble plays a crucial role in statistical mechanics in and out of equilibrium. For example, the standard derivation of the fluctuation theorem relies on the assumption that the initial state of the heat bath is the canonical ensemble. On the other hand, the recent progress in the foundation of statistical mechanics has revealed that a thermal equilibrium state is not necessarily described by the canonical ensemble but can be a quantum pure state or even a single energy eigenstate, as formulated by the eigenstate thermalization hypothesis (ETH). Then a question raised is how these two pictures, the canonical ensemble and a single energy eigenstate as a thermal equilibrium state, are compatible in the fluctuation theorem. In this paper, we theoretically and numerically show that the fluctuation theorem holds in both of the long- and short-time regimes, even when the initial state of the bath is a single energy eigenstate of a many-body system. Our proof of the fluctuation theorem in the long-time regime is based on the ETH, while it was previously shown in the short-time regime on the basis of the Lieb-Robinson bound and the ETH [Phys. Rev. Lett. **119**, 100601 (2017)]. The proofs for these time regimes are theoretically independent and complementary, implying the fluctuation theorem in the entire time domain. We also perform a systematic numerical simulation of hard-core bosons by exact diagonalization and verify the fluctuation theorem in both of the time regimes by focusing on the finite-size scaling. Our results contribute to the understanding of the mechanism that the fluctuation theorem emerges from unitary dynamics of quantum many-body systems and can be tested by experiments with, e.g., ultracold atoms.

I. INTRODUCTION

Conventional statistical mechanics relies on the concept of ensembles, such as the microcanonical ensemble and the canonical ensemble. These are characterized as the maximum entropy states under certain energy constraints [1] and thus given by statistical mixtures of enormous energy eigenstates. The canonical ensemble plays a crucial role even beyond equilibrium situations. For example, we can prove the second law of thermodynamics and the fluctuation theorem [2–13] out of equilibrium of the system, by relying on the assumption that the initial state of the heat bath is in the canonical ensemble. Here the fluctuation theorem is a universal relation that incorporates the role of fluctuations of the entropy production, represented as $\langle e^{-\sigma} \rangle = 1$ where σ is the stochastic entropy production. The fluctuation theorem implies the second law at the average level, stating that the average entropy production is nonnegative:

$$\langle \sigma \rangle := \Delta S_{\mathrm{S}} - \beta Q \geqslant 0, \tag{1}$$

where ΔS_{S} is the change of the von Neumann entropy of the system, β is the inverse temperature of the heat bath, and Q is the average heat emitted from the bath. We note that the fluctuation theorem reproduces the second law through the Jensen inequality $\langle e^{-\sigma} \rangle \geqslant e^{-\langle \sigma \rangle}$.

On the other hand, in recent years, it has been established that most states, not limited to those ensembles (maximum entropy states), can represent thermal equilibrium states [14–26]. The extreme case opposite to the canonical ensemble is zero entropy states such as a single energy eigenstate; the eigenstate thermalization hypothesis (ETH) [27–43] states that a single energy eigenstate can represent a thermal equilibrium state. The ETH is a sufficient condition for thermalization in isolated quantum many-body systems [27–29] and numerically shown to be valid in various nonintegrable quantum many-body systems [29–43]. Furthermore, in recent experiments such as cold atoms [19–24], trapped ions [25], and superconducting qubits [26], it has been observed that thermal equilibrium states are not necessarily canonical ensembles but can be pure states.

A critical question here is how universally the thermodynamic properties appear without canonical ensembles, even in nonequilibrium situations. In particular, it is interesting to investigate the validity of the fluctuation theorem: The question focuses on whether it is possible to understand how the fluctuation theorem and the second law emerge from quantum mechanics without assuming the conventional statistical ensembles.

A partial understanding of the above question has been addressed in recent papers. Specifically, Ref. [44] theoretically showed the fluctuation theorem (and the second law) in the short-time regime, which is based on the ETH and the Lieb-Robinson bound [45–47]. We note that in the numerical simulation of Ref. [44], the initial state of the system was

chosen to be a pure state, and the fluctuation theorem was apparently broken in the long-time regime. This apparent breakdown originates from so-called absolute irreversibility [48,49] induced by the pure initial state, and as will be shown in this paper, the fluctuation theorem still holds in the long-time regime if absolute irreversibility is properly treated. Moreover, Ref. [50] theoretically showed the second law at the average level in the long-time regime on the basis of the ETH. Also in previous papers [51,52], it has been numerically suggested that the fluctuation theorem holds even in the long-time regime. Given these researches, it is desirable to make a comprehensive understanding of the validity of the fluctuation theorem in the entire time domain.

The main result of this paper is to show, analytically and numerically, that the fluctuation theorem holds with the non-canonical bath in both of the short- and long-time regimes. Specifically, the initial state of the bath is supposed to be a single energy eigenstate sampled from an energy shell. This establishes that the universal property of nonequilibrium fluctuations in entropy production emerges even without the canonical ensemble. In particular, since we prove the fluctuation theorem for the extreme case with a single energy eigenstate, the fluctuation theorem holds when the initial state of the bath is any mixture of energy eigenstates in the energy shell.

We discuss the entire time domain by considering the long- and short-time regimes. We theoretically show that the long-time average of the deviation from the fluctuation theorem vanishes in the thermodynamic limit of the heat bath. We also perform numerical simulations and show that the deviation decreases with the bath size N. On the other hand, Ref. [44] theoretically showed the fluctuation theorem for the short-time regime in the thermodynamic limit on the basis of the Lieb-Robinson bound and the ETH. While the theory of Ref. [44] is valid in the thermodynamic limit, the fluctuation theorem in the short-time regime for the numerically-accessible system size was not fully established in Ref. [44] because of the large finite-size effect. In the present paper, we perform systematic numerical calculations to clarify whether the theoretical scenario is indeed relevant for the numerically accessible system size. In particular, we numerically investigate the bath-size dependence of the error term of the fluctuation theorem, which eventually excludes other scenarios than the theoretical scenario of Ref. [44].

The rest of this paper is organized as follows. In Sec. II we introduce the setup of this study. In Sec. III we overview the main results of this paper without going into details. In Sec. IV we theoretically derive the fluctuation theorem in the long-time regime and show our numerical results to support the theory. In Sec. V we discuss the fluctuation theorem in the short-time regime and show the corresponding numerical results. In Sec. VI we summarize the results and make some remarks. In Appendix A we introduce the concept of absolute irreversibility [48,49]. In Appendix B we examine the details of the interaction-induced error of the fluctuation theorem. In Appendix C we show that the interaction-induced error vanishes under the rotating wave approximation. In Appendix D we discuss another naive approach to show the fluctuation theorem. In Appendix E we discuss the initial time dependence of the errors of the fluctuation theorem. In Appendix F we

explain the details of the proof of the fluctuation theorem in the long-time regime. In Appendix G we show the additional calculations for the interaction-induced error. In Appendix H we show the supplementary numerical results.

II. SETUP

In this section we introduce the setup of the study. In Sec. II A we explain the conventional setup for the fluctuation theorem with the canonical bath and briefly overview the fluctuation theorem [2–13]. In Sec. II B we explain the setup of the present study with the energy eigenstate bath. In Sec. II C we explain the setup of our numerical simulation.

A. Fluctuation theorem for the canonical bath

As a preliminary, we consider the conventional setup of the fluctuation theorem. The total system is composed of the system S and the heat bath B. The Hamiltonian of the total system is written as

$$H := H_S + H_I + H_B, \qquad (2)$$

where H_S, H_B are the Hamiltonian of system S and bath B, respectively, and $H_I(\neq 0)$ describes the interaction between S and B. The initial state of the total system SB is assumed to be a product state:

$$\rho(0) := \rho_S(0) \otimes \rho_B(0). \qquad (3)$$

We also define the canonical ensembles of system S and bath B as $\rho_S^{can} := e^{-\beta H_S}/Z_S$ with $Z_S := \mathrm{tr}_S[e^{-\beta H_S}]$ and $\rho_B^{can} := e^{-\beta H_B}/Z_B$ with $Z_B := \mathrm{tr}_B[e^{-\beta H_B}]$, respectively.

The time evolution of the total system is given by $\rho(t) = U\rho_S(0) \otimes \rho_B(0)U^\dagger$, where $U(t) := e^{-iHt/\hbar}$ is the unitary time evolution operator. We write the reduced density operator of system S and bath B at time t as $\rho_S(t) := \mathrm{tr}_B[\rho(t)]$, $\rho_B(t) := \mathrm{tr}_S[\rho(t)]$, respectively.

In this setup, we explain the concept of the stochastic entropy production σ [7]. To introduce it, let us consider an operator $\sigma(t) := -\ln \rho_S(t) + \beta H_B$ (see [9,53,54] for the same approach to the stochastic entropy production). The first term on the right-hand side is the informational contribution, whose average is the von Neumann entropy of the system $S_S(t) := -\mathrm{tr}_S[\rho_S(t)\ln\rho_S(t)]$. We note that such an informational contribution has been experimentally investigated in quantum systems [55,56]. The second term of $\sigma(t)$ is the thermal contribution that gives the heat term.

While $\sigma(t)$ is not an ordinary physical observable because it depends on the state $\rho_S(t)$, the above definition of $\sigma(t)$ enables us to treat various physically relevant situations in a unified manner. One of typical situations is that the initial state of system S is the canonical ensemble, in which the measurement of $\sigma(t)$ is equivalent to the measurement of energy, because $-\ln\rho_S^{can} = \beta H_S + \ln Z_S$. If we do not have any prior knowledge about the state of the system, we can measure $-\ln\rho_S(t)$ by performing quantum state tomography of $\rho_S(t)$ beforehand [57]. In fact, by quantum state tomography, we obtain the information about the eigenbasis of $\rho_S(t)$ and its eigenvalues. Then we can perform the projection measurement of the eigenbasis and associate an eigenvalue with the outcome.

We consider the projection measurement of $\sigma(t)$ at time 0 and t. Let the measurement outcomes be σ_i at $t = 0$ and σ_f at t. We then define the stochastic entropy production σ as $\sigma := \sigma_f - \sigma_i$, whose average $\langle \sigma \rangle = \Delta S_S - \beta Q$ is the average entropy production of the total system. Here Q is the heat emitted from bath B to system S, defined as $Q := -\mathrm{tr}_B[H_B(\rho_B(t) - \rho_B(0))]$.

For the conventional fluctuation theorem, the initial state of bath B is assumed to be the canonical ensemble $\rho_B(0) = \rho_B^{\mathrm{can}}$. The fluctuation theorem states that

$$\langle e^{-\sigma} \rangle = 1. \tag{4}$$

It is straightforward to show [7] that $\langle e^{-\sigma} \rangle$ can be rewritten as

$$\langle e^{-\sigma} \rangle = \mathrm{tr}\big[e^{-\beta H_B} U e^{\beta H_B} \rho_B^{\mathrm{can}} U^\dagger \rho_S^{\mathrm{can}}(t) \big]. \tag{5}$$

Substituting $\rho_B^{\mathrm{can}} = e^{-\beta H_B}/Z_B$, we obtain

$$\langle e^{-\sigma} \rangle = \mathrm{tr}\bigg[e^{-\beta H_B} U e^{\beta H_B} \frac{e^{-\beta H_B}}{Z_B} U^\dagger \rho_S^{\mathrm{can}}(t) \bigg] \tag{6}$$

$$= \mathrm{tr}\bigg[\frac{e^{-\beta H_B}}{Z_B} U U^\dagger \rho_S^{\mathrm{can}}(t) \bigg] \tag{7}$$

$$= \mathrm{tr}\big[\rho_S^{\mathrm{can}}(t) \otimes \rho_B^{\mathrm{can}} \big] \tag{8}$$

$$= 1. \tag{9}$$

We note that if absolute irreversibility [48,49] occurs, the fluctuation theorem is modified (see details in Appendix A). Absolute irreversibility is an apparent violation of the fluctuation theorem that occurs when $\rho_S(0)$ does not have the full rank [e.g., $\rho_S(0)$ is a pure state]. The modification term appears in the form of

$$\langle e^{-\sigma} \rangle = 1 - \lambda(t). \tag{10}$$

See Appendix A for details. Absolute irreversibility can be understood as a kind of singularity about the initial state of system S. We emphasize that this is not a purely mathematical problem, but can be observed experimentally [55].

B. Fluctuation theorem for the energy eigenstate bath

In this subsection, we introduce the setup of the fluctuation theorem for the energy eigenstate bath, which will be shown as the main result of the present paper.

The total system is composed of system S and bath B in the same manner as the setup for the conventional fluctuation theorem, while in the present setup, bath B is defined on a lattice as shown in Fig. 1. We assume that the Hamiltonian of bath B is local and system S interacts with a local region of B by the interaction H_I. Let N be the number of sites in bath B, and $d = 1, 2, 3, \ldots$ be the spatial dimension of it. The size of system S and the support of H_I are fixed and do not depend on N. We refer to the dimensions of the Hilbert space of S, B, and the total system SB as D_S, D_B, and D_{SB}, respectively. We denote an eigenstate of H with eigenenergy E_a as $|E_a\rangle$. In the same manner, we denote eigenstates of H_S, H_B as $|E_i^S\rangle$, $|E_\alpha^B\rangle$, respectively. We denote a matrix element of an operator with respect to the eigenstates of H as $(\cdots)_{ab} := \langle E_a | \cdots | E_b \rangle$. We also define an operator of system S as $q_S^{ij} := |E_i^S\rangle\langle E_j^S|$.

In the present paper, we consider the thermodynamic limit that means the large-bath limit ($N \to \infty$) without changing
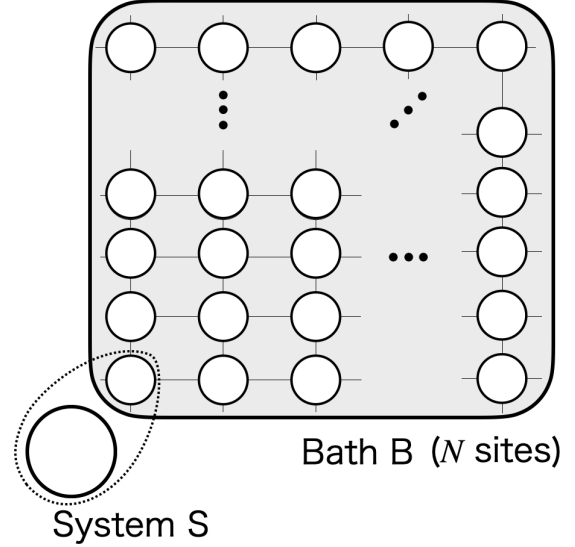

FIG. 1. The setup of our study. The total system consists of the system S and the heat bath B. The system S interacts with a local (bounded) region of B, and the interaction in B is also local. The bath B is defined on a d-dimensional lattice. We denote the number of sites of B by N and the dimension of the Hilbert space of B by D_B. We also consider an energy shell of B, whose Hilbert-space dimension is D. See the inset of Fig. 2 for the specific lattice used in our numerical simulation.

system S. To describe asymptotic behaviors of this limit, we use the following asymptotic notations:

$$f(D) = \Theta(g(D)) \Leftrightarrow 0 < \lim_{D\to\infty} \left| \frac{f(D)}{g(D)} \right| < \infty, \tag{11}$$

$$f(D) = O(g(D)) \Leftrightarrow \lim_{D\to\infty} \left| \frac{f(D)}{g(D)} \right| < \infty, \tag{12}$$

$$f(D) = o(g(D)) \Leftrightarrow \lim_{D\to\infty} \left| \frac{f(D)}{g(D)} \right| = 0. \tag{13}$$

We define the operator norm $\|X\|$ as the largest singular value of X and the trace norm $\|X\|_1$ as the sum of the singular values of X. For the sake of simplicity, we use $A \simeq B$ for operators A and B to represent that $\|A - B\| = o(1)$ or $\|A - B\|_1 = o(1)$ depending on the context.

The initial state of SB is assumed to be a product state as in (3). We also assume that the initial state of S is diagonal with respect to H_S, because it simplifies the discussion of the time dependence in Secs. III and V.

Now, a crucial assumption is that the initial state of bath B is a single energy eigenstate, which is sampled from the energy shell $[E - \Delta E, E]$. The energy E is given by $E := \mathrm{tr}_B[\rho_B^{\mathrm{can}} H_B]$ with the canonical ensemble at inverse temperature β. The width ΔE can be chosen as $\Theta(1) \leqslant \Delta E \leqslant \Theta(N)$, which gives the normal thermodynamic scaling $D = \exp(sN)$ with the entropy density s [58].

We next summarize the assumptions regarding the ETH, as summarized in Table I. For the proof of the fluctuation theorem in the long-time regime (Sec. IV), we assume that H satisfies the ETH for all the eigenstates in the energy shell (the "strong" ETH) with respect to all observables of system S and H_I. Here we can introduce the energy shell of H, which

TABLE I. The assumptions regarding the ETH.

Time regime (section)	Long-time regime (Sec. IV)	Short-time regime (Sec. V)
Which ETH	The diagonal and off-diagonal ETH of H	The diagonal ETH of H_B
For which eigenstates	All the eigenstates	Only for the given initial state of B

is determined by the energy and energy width of the initial state (3) (see details in Appendix F). Specifically, we assume the (strong) diagonal and off-diagonal ETH [18,27–29] for all observables of system S and H_I, written as O_{SI}, stating that the following relations hold for all the eigenstates in the energy shell:

$$\langle E_a | O_{SI} | E_a \rangle = \langle O_{SI} \rangle_{MC} + O(D'^{-1/2}), \quad (14)$$

$$|\langle E_a | O_{SI} | E_b \rangle| = O(D'^{-1/2}) \quad (a \neq b), \quad (15)$$

where $\langle \cdot \rangle_{MC}$ is the microcanonical average of the energy shell of H. Since the size of system S is $\Theta(1)$, the dimension of the energy shell $D' \, (< D_{SB})$ satisfies $D' = \Theta(D)$.

In Sec. V we discuss the fluctuation theorem in the short-time regime, which has been theoretically proved in Ref. [44]. Here we assume the diagonal ETH of H_B:

$$\langle E_a^B | L_{B_0} | E_a^B \rangle = \langle L_{B_0} \rangle_{MC,B} + O(D^{-1/2}), \quad (16)$$

where $\langle \cdot \rangle_{MC,B}$ is the microcanonical average of the energy shell $[E - \Delta E, E]$, $D \, (< D_B)$ is the dimension of the Hilbert space of the energy shell of H_B, and L_{B_0} is a quasilocal operator defined in Sec. V. We note that Eq. (16) is assumed only for the *given* initial state $|E_a^B\rangle$ of B.

We remark on the validity of the ETH. It has been numerically shown that the strong ETH holds in various nonintegrable quantum many-body systems [29–43], while it does not hold in the presence of quantum many-body scar [59] and in integrable systems [32]. The assumptions (14) and (15) rely on the strong ETH of the local operator O_{SI}. On the other hand, the ETH for a given single energy eigenstate can be valid even in integrable systems [30,44]. In fact, it has been shown that almost all energy eigenstates in the energy shell satisfy the ETH even in integrable systems [44,60,61]. This is true even for quasilocal operators such as L_{B_0} in Eq. (16) [44]. Specifically, if the support size of a quasilocal operator increases as N^α $(0 < \alpha < 1/2)$, it satisfies the ETH for almost all energy eigenstates as proven in Ref. [44]. This theorem applies to integrable systems but requires translation invariance. We note that a consistent observation has been made in terms of the entanglement properties of integrable systems [62,63]. We also note that the ETH for typical energy eigenstates does not hold in many-body localized (MBL) systems that break translation invariance [64,65].

Meanwhile, in the following sections, we sometimes consider the following condition for the interaction Hamiltonian:

$$[H_S + H_B, H_I] = 0, \quad (17)$$

which simplifies the derivation of the fluctuation theorem. Under this condition, the sum of the energies of system S and bath B does not change, and we can rewrite the heat Q by the energy change in system S. We note that the interaction itself (i.e., $\|H_I\|$) is not necessarily small for the condition (17) to

hold. While the condition (17) is satisfied in simple models such as the Jaynes-Cummings model at the resonant condition [66], it is not necessarily satisfied in generic quantum many-body systems.

When the condition (17) is not satisfied, an error term induced by the interaction can appear in the fluctuation theorem. It is noteworthy, however, that Eq. (17) is approximately satisfied under the rotating wave approximation [66], which holds well even for many-body systems in a long-time regime where high-frequency oscillations can be neglected. In Appendix C we show that

$$\|[H_S + H_B, H_I]\| = o(1) \quad (18)$$

holds by using the rotating wave approximation and the off-diagonal ETH. Intuitively, the rotating wave approximation ensures the energy conservation without including the interaction energy, i.e., it prohibits transitions between eigenstates of $H_S + H_B$, implying that H_I becomes commutable with $H_S + H_B$.

C. Hamiltonian for numerical simulation

In this subsection, we explain the setup of our numerical simulation discussed in Secs. III, IV, and V.

We perform numerical calculations of hard-core bosons with nearest-neighbor repulsion using numerically exact diagonalization. System S is a single site, and bath B is on a two-dimensional lattice. The Hamiltonian is given by

$$H_S := \omega n_0, \quad (19)$$

$$H_I := -\gamma' \sum_{\langle 0, j \rangle} (c_0^\dagger c_j + c_j^\dagger c_0), \quad (20)$$

$$H_B := \omega \sum_{i=1}^N n_i - \gamma \sum_{\langle i,j \rangle} (c_i^\dagger c_j + c_j^\dagger c_i) + g \sum_{\langle i,j \rangle} n_i n_j, \quad (21)$$

where ω is the onsite potential, $-\gamma$ is the hopping in bath B, $-\gamma'$ is the hopping between system S and bath B, and $g > 0$ represents repulsion between hard core bosons. We note that $\langle i, j \rangle$ means the sum over the nearest-neighbor sites. Hard core bosons cannot exist simultaneously on a single site, and their annihilation (creation) operator c_i $(c_i^\dagger)$ satisfies the commutation relations $[c_i, c_j] = [c_i^\dagger, c_j^\dagger] = [c_i, c_j^\dagger] = 0$ for $i \neq j$, $\{c_i, c_i\} = \{c_i^\dagger, c_i^\dagger\} = 0$, and $\{c_i, c_i^\dagger\} = 1$. The occupation number operator is defined as $n_i := c_i^\dagger c_i$. The site of $i = 0$ corresponds to system S. We set bath B as a two-dimensional square lattice with the open boundary condition, and the bath size N is written as $N = L_x \times L_y$. The operator of the particle number in bath B is written as $n_B := \sum_{i=1}^N n_i$. The Hamiltonian H is nonintegrable when $g \neq 0$, $\gamma \neq 0$, and $\gamma' \neq 0$.

We write the initial state of system S as

$$\rho_S(0) := p|1\rangle\langle 1| + (1-p)|0\rangle\langle 0|, \qquad (22)$$

where $|n_S\rangle$ is the eigenstate of n_0 and satisfies $n_0|n_S\rangle = n_S|n_S\rangle$ ($n_S = 1, 0$). In order to avoid absolute irreversibility, we choose $\rho_S(0)$ to be a mixed state by setting $p = 0.99$ in our numerical simulation. See Appendixes A and F 3 for the case with absolute irreversibility. We define the change of the particle number in S as

$$\delta n_S := \mathrm{tr}_S[n_0(\rho_S(t) - \rho_S(0))]. \qquad (23)$$

The initial state of bath B is an energy eigenstate of H_B with a particle number N_P, which samples from the energy shell $[E - \Delta E, E]$. We write the width of the energy shell as $\Delta E = N\delta_E$ with δ_E be a positive constant and set $\delta_E = 0.02$ in our numerical simulation. We perform the calculation of the fluctuation theorem for each energy eigenstate in the energy shell and investigate the dependence of the error of the fluctuation theorem on the bath size and the initial state of bath B. We define the inverse temperature of $|E_\alpha^B\rangle$ as β_α satisfying $E_\alpha^B = \mathrm{tr}_B[\rho_B^{\mathrm{can}}(\beta_\alpha)H_B]$, where we explicitly write the β dependence of the canonical ensemble. The onsite potential ω is determined such that the canonical expectation of the particle number of the bath equals N_P: $\mathrm{tr}_B[n_B\rho_B^{\mathrm{can}}] = N_P$. We set ω as above, because the ETH is satisfied only within each particle number sector, and N_P should be close to the canonical expectation number. To investigate the bath-size dependence, we set $L_x = 3$, $L_y = 3, 4, 5$ and $N_P = N/3$.

III. OVERVIEW OF THE RESULTS

In this section, we give an overview of the main results of this paper: The fluctuation theorem holds in both the long- and short-time regimes (as defined below) in the setting of the previous section. The proof of the fluctuation theorem in the long-time regime is based on the ETH of H (Sec. IV). Also, the fluctuation theorem in the short-time regime has been shown on the basis of the ETH of H_B and the Lieb-Robinson bound [44] (Sec. V). The proofs of the fluctuation theorem in these time regimes are theoretically independent of each other and play complementary roles. Our systematic numerical calculations about the bath-size dependence support the theories in both of the long- and short-time regimes.

This section is organized as follows. In Sec. III A we show the numerical result of the real-time dynamics and define the errors of the fluctuation theorem. In Secs. III B and III C we explain the results on the fluctuation theorem in the long- and short-time regimes, respectively. In Sec. III D we comment on the two time regimes.

A. Errors of the fluctuation theorem

We first demonstrate whether the fluctuation theorem (4) holds or not by numerically investigating the real-time dynamics. Figure 2 shows the time dependence of $\langle e^{-\sigma}\rangle$, which implies that $\langle e^{-\sigma}\rangle \simeq 1$ holds in the entire time domain. The right inset of Fig. 2 shows the time dependence of the error of the fluctuation theorem $|\langle e^{-\sigma}\rangle - 1|$. The error increases until the relaxation time of the system S ($t \sim 1$) and then saturates. Qualitatively the same behavior is seen with other interacting
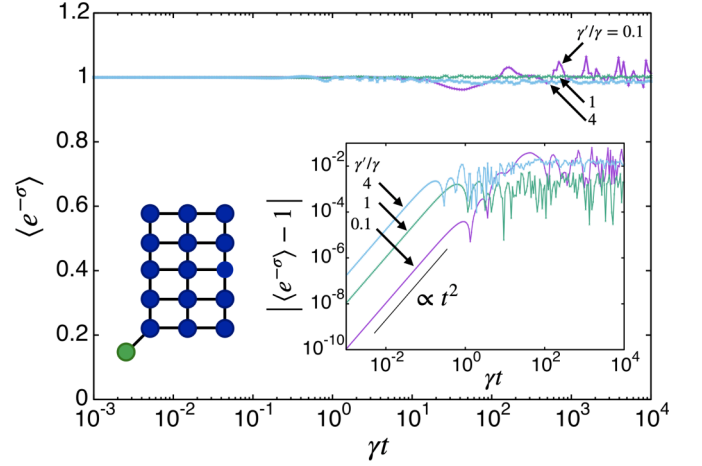

FIG. 2. The time dependence of $\langle e^{-\sigma}\rangle$. The right inset shows the time dependence of the error of the fluctuation theorem $|\langle e^{-\sigma}\rangle - 1|$, whose initial rise is proportionate to t^2. The left inset is a schematic of the total system used in numerical calculations. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\gamma' = 0.1\gamma$ (purple), γ (green), 4γ (blue). The initial state of the bath is the energy eigenstates of H_B, whose energy is maximum in the energy shell at $\beta = 0.1$. The onsite potential ω is determined by $\mathrm{tr}_B[n_B\rho_B^{\mathrm{can}}] = N_P$.

parameters (see Appendix H), and Fig. 3 is a schematic of the typical time dependence of $|\langle e^{-\sigma}\rangle - 1|$.

We denote $\langle e^{-\sigma}\rangle$ when the initial state of bath B is $\rho_B(0)$ (resp. ρ_B^{can}) by G (resp. G^{can}). We define G and G^{can} as

$$G := \mathrm{tr}[e^{-\beta H_B}Ue^{\beta H_B}\rho_B(0)U^\dagger\rho_S(t)], \qquad (24)$$

$$G^{\mathrm{can}} := \mathrm{tr}[e^{-\beta H_B}Ue^{\beta H_B}\rho_B^{\mathrm{can}}U^\dagger\rho_S^{\mathrm{can}}(t)] = 1, \qquad (25)$$

where $\rho_S^{\mathrm{can}}(t)$ is the reduced density operator of system S defined as $\rho_S^{\mathrm{can}}(t) := \mathrm{tr}_B[U\rho_S(0)\otimes\rho_B^{\mathrm{can}}U^\dagger]$. Below we focus on the error of the fluctuation theorem $G - G^{\mathrm{can}}$.

In the case where Eq. (17) holds, H_B in G and G^{can} can be replaced by $-H_S$, and correspondingly we define

$$G_S := \mathrm{tr}[e^{\beta H_S}Ue^{-\beta H_S}\rho_B(0)U^\dagger\rho_S(t)], \qquad (26)$$

$$G_S^{\mathrm{can}} := \mathrm{tr}[e^{\beta H_S}Ue^{-\beta H_S}\rho_B^{\mathrm{can}}U^\dagger\rho_S^{\mathrm{can}}(t)]. \qquad (27)$$
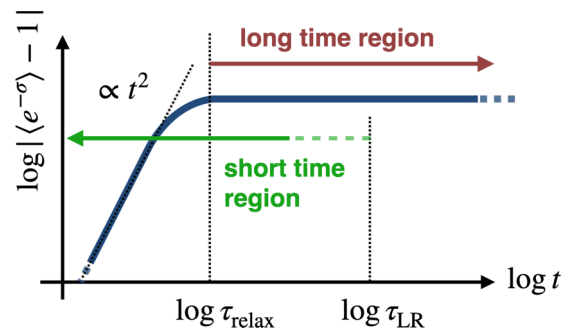

FIG. 3. Sketch of a typical time dependence of $|\langle e^{-\sigma}\rangle - 1|$, which initially rises in proportionate to t^2 and relaxes to the long-time average in $t \gtrsim \tau_{\mathrm{relax}}$. We refer to the time regime after the relaxation as the long-time regime. We also call the sufficiently shorter time regime than the Lieb-Robinson time τ_{LR} (38) the short-time regime ($t \ll \tau_{\mathrm{LR}}$).

Using the Baker-Campbell-Hausdorff formula for $e^{-\beta(H_S+H_B)}Ue^{\beta(H_S+H_B)}$, we can show $G = G_S$ and $G^{can} = G_S^{can}$ under Eq. (17), and the error of the fluctuation theorem is written as

$$G - G^{can} = G_S - G_S^{can}. \tag{28}$$

With the above argument, we define the error of the fluctuation theorem in the general case as

$$G - G^{can} = \delta G_S + \delta G_I, \tag{29}$$

where we define

$$\delta G_S := G_S - G_S^{can}, \tag{30}$$

$$\delta G_I := G - G^{can} - \delta G_S \tag{31}$$

$$= G - G_S + G_S^{can} - G^{can} \tag{32}$$

$$= \delta G_I^{(1)} + \delta G_I^{(2)}, \tag{33}$$

$$\delta G_I^{(1)} := G - G_S, \tag{34}$$

$$\delta G_I^{(2)} := G_S^{can} - G^{can}. \tag{35}$$

We refer to δG_I as the interaction-induced error.

As argued above, $\delta G_I = 0$ holds under Eq. (17). Even if Eq. (17) does not hold, the interaction-induced error δG_I satisfies

$$|\delta G_I| = o(1) \tag{36}$$

under the rotating wave approximation and the off-diagonal ETH (see Appendix C). We also note that $\delta G_I = 0$ holds without the rotating wave approximation when $\rho_B(0) = \rho_B^{can}$, even if Eq. (17) does not hold. See also Appendix B for the form of δG_I.

We have decomposed the error of the fluctuation theorem into δG_S and δG_I. If we try to show the error without the decomposition, we cannot show that the error vanishes in the thermodynamic limit (see Appendix D).

B. Long-time regime

We consider the fluctuation theorem in the long-time regime. In this paper, we prove that $\overline{\langle e^{-\sigma}\rangle}$, the long-time average of $\langle e^{-\sigma}\rangle$, nearly equals 1 in the large-bath limit. Note that we denote the long-time average of any quantity $O(t)$ as

$$\overline{O(t)} := \lim_{T\to\infty}\frac{1}{T}\int_0^T O(t)\,dt. \tag{37}$$

In the special case that the condition (17) is satisfied, we can prove that $|\overline{\langle e^{-\sigma}\rangle} - 1| = o(1)$ holds, where the right-hand side represents the asymptotic behavior with respect to the bath size N.

If Eq. (17) is not strictly satisfied, the interaction-induced error δG_I defined in Sec. III A can generally appear in the fluctuation theorem. However, the interaction-induced error vanishes in the thermodynamic limit under the rotating wave

approximation, which holds well in the long-time regime (see Appendix C for details).

In the absence of localization [64,65] or persistent oscillations [59], $\langle e^{-\sigma}\rangle$ relaxes to the long-time average $\overline{\langle e^{-\sigma}\rangle}$ after the finite relaxation time τ_{relax}. We refer to the regime of t satisfying $\tau_{\text{relax}} \lesssim t$ as the long-time regime. We argue that the relaxation time satisfies $\tau_{\text{relax}} = \Theta(1)$ with respect to the bath size N for the following reason. The relaxation of $\langle e^{-\sigma}\rangle$ can be associated with observables of system S, because under the condition (17), the heat Q is replaced by the energy change of system S. Under physically reasonable conditions, the relaxation time of observables of system S is $\Theta(1)$ when the initial state of bath B is the microcanonical ensemble [67]. We note that even when the condition (17) does not strictly hold, we can replace the heat Q under the rotating wave approximation, which again implies the relaxation time of $\Theta(1)$. We note that Ref. [68] showed that the relaxation time is independent of the size of the total system if the Hamiltonian is random, and that numerical results consistent with the random Hamiltonians have been obtained with realistic models [69,70]. In summary, we argue that $\tau_{\text{relax}} = \Theta(1)$ holds in our setup, which is consistent with our own numerics as well.

We briefly remark on absolute irreversibility (see Appendixes A and F 3 for details). As mentioned in the introduction, the numerical calculations in Ref. [44] showed that $\langle e^{-\sigma}\rangle$ is smaller than 1 in the long-time regime, which was argued to be a violation of the fluctuation theorem. As a matter of fact, this is due to the numerical setup of Ref. [44] that the initial state of system S is a pure state, causing absolute irreversibility. That is, the reason why the fluctuation theorem in the long-time regime appeared to be broken in the numerical calculations of Ref. [44] is that the correction term λ in Eq. (10) was not taken into account, while the numerical calculation itself is correct. In the present paper, we prove and numerically show that the fluctuation theorem holds in the long-time regime with absolute irreversibility and the energy eigenstate bath if the correction term λ is taken into account. Finally, we remark that there is an alternative approach to make the fluctuation theorem hold by regularizing the pure initial state [71].

C. Short-time regime

We next consider the short-time regime. To define the short-time regime in line with Ref. [44], we divide bath B into B_0 and B_1, where B_0 is near S and B_1 is far from S (see Fig. 4). We refer to the boundary between B_0 and B_1 as ∂B, and the size of B_0 is set to be $\Theta(N^{\mu d})$ [$0 < \mu < 1/(2d)$]. Then the short-time regime is defined as $t \ll \tau_{\text{LR}}$. Here τ_{LR} is the Lieb-Robinson (LR) time introduced by the Lieb-Robinson bound [45–47], which, in the present case, represents a timescale that the information of S reaches B_1. Specifically, under the above choice of the size of B_0, the LR time is given by

$$\tau_{\text{LR}} = \Theta(N^{\mu}). \tag{38}$$

Again in the special case that the condition (17) is satisfied, $|\langle e^{-\sigma}\rangle - 1| = o(1)$ holds in the short-time regime, as shown in Ref. [44] on the basis of the Lieb-Robinson bound and
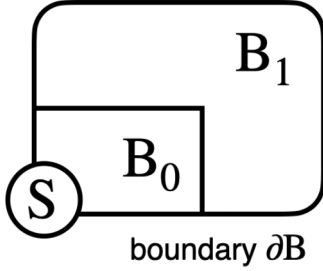
FIG. 4. Division of bath B used in the discussion of the short-time regime in Sec. V. We divide bath B into the near part (B_0) and the far part (B_1). The size of B_0 depends on N as $\Theta(N^{\mu d})$ [$0 < \mu < 1/(2d)$]. System S interacts with a part of B_0. From the Lieb-Robinson bound, we introduce the Lieb-Robinson time $\tau_{LR} = \Theta(N^{\mu})$, at which the information of system S reaches B_1. We note that the velocity of the information propagation does not depend on N.

the ETH. Even when the condition (17) does not hold, we show that the interaction-induced error vanishes within the perturbation theory at high temperature if the initial state of bath B satisfies the ETH (see Appendix G 2).

In the present paper, to show the validity of the fluctuation theorem in the short-time regime, we perform systematic numerical calculations of the bath-size dependence of the error of the fluctuation theorem. Our results in Sec. V show that the error decreases as N increases, which supports the validity of the fluctuation theorem.

We note that the error of the fluctuation theorem is 0 at $t = 0$ and initially increases in proportion to t^2 as shown in Fig. 3. We theoretically show this time dependence in Appendix E. In Sec. V, assuming that the error of the fluctuation theorem initially grows in the form of at^2 with a being a t-independent coefficient, we numerically show that this coefficient a decreases as the bath size increases, implying $a = o(1)$.

We note, as a side remark, that the time dependence of the change of the particle number in S (23) also initially takes the form of $a_n t^2$ with a_n being a t-independent coefficient. In this case, we observe that the bath-size dependence of a_n is just $a_n = \Theta(1)$. The bath-size independence of a_n contrasts with the error of the fluctuation theorem $a = o(1)$, implying that the ETH plays a crucial role only for the latter.

D. On the two time regimes

We emphasize that the short- and long-time regimes are defined and analyzed theoretically independently. The short-time regime is defined by the Lieb-Robinson time, which increases at most linearly with the bath size, while the long-time regime is defined by coincidence with the long-time average, which covers a very long timescale characterized by the quantum recurrence time that increases doubly exponentially with the bath size [72–74]. From the argument in Secs. III B and III C, the short- and long-time regimes have an overlap and thus cover the entire time domain as shown in Fig. 3, when the bath size is sufficiently large (see also Sec. V A).

Suppose that the time evolution of the error is given as in Fig. 3, i.e., the error of the fluctuation theorem initially

increases in t^2 and relaxes monotonically to the long-time average after the N-independent relaxation time. Then the fluctuation theorem in the long-time regime implies that in the short-time regime, and vice versa. In general, however, the fluctuation theorem in these time regimes is theoretically shown independently, and the respective theorems together show the fluctuation theorem in the entire time domain. In fact, only from the fluctuation theorem in the long-time regime, we cannot exclude the situation that the error is not monotonic like Fig. 3, but overshoots to a value larger than $o(1)$ after a time evolution of t^2 and then relaxes to the long-time average. This possibility can be excluded from the fluctuation theorem in the short-time regime.

Let us discuss the timescales of real experimental setups. For example, in Ref. [75], real-time dynamics of ultracold atoms has been experimentally studied over about 300τ with τ being the tunneling time. On the other hand, the Lieb-Robinson time τ_{LR} is estimated as $\tau_{LR} \sim \ell\tau$, where $\ell = N^{1/4}$ is the distance between S and B_1. As the experiment is performed in the two-dimensional system where the site number is $N \sim 250$, the Lieb-Robinson time is evaluated as 4τ. Thus, the Lieb-Robinson time is reasonably longer than the experimental time resolution, while it is much shorter than the experimentally tractable time regime. In such an experimental setup, therefore, the theories for the short- and long-time regimes are both necessary to address the entire time domain. On the other hand, we note that the Lieb-Robinson time is too short in our numerical setup of Sec. II C for the short-time regime to be clearly visible, because our system size is small due to the limitation of numerical accessibility ($N \sim 15$). Our theory for the short-time regime as well as for the long-time regime is more relevant to real-experimental setups with current or the near-future technologies, where the system sizes can be much bigger than numerics.

IV. LONG-TIME REGIME

In this section we theoretically and numerically show the fluctuation theorem in the long-time regime. In Sec. IV A we discuss the long-time average and the relaxation time to the long-time average. In Sec. IV B we outline the proof of the fluctuation theorem. In Sec. IV C we show our numerical results are consistent to our theory.

A. Long-time average

In this subsection, we consider the long-time average of $\delta G_S = G_S - G_S^{can}$ and discuss the relaxation times of δG_S and δG_I.

To analyze the long-time average of G_S, we first write G_S as follows:

$$G_S := \text{tr}[e^{\beta H_S} U e^{-\beta H_S} \rho_B(0) U^\dagger \text{tr}_B[U \rho_S(0) \otimes \rho_B(0) U^\dagger]].$$
(39)

We note that Eq. (39) contains four unitary operators.

Then, under no degeneracy and the nonresonance condition [i.e., $E_a - E_b + E_c - E_d = 0$ holds only when the pair of indexes (a, c) equals (b, d)], the long-time average of G_S

can be written as $\overline{G_S} = \overline{G_{S1}} + \overline{G_{S2}}$. Here $\overline{G_{S1}}$ and $\overline{G_{S2}}$ are defined as

$$\overline{G_{S1}} := Z_S \sum_{a,c} \mathrm{tr}_S\big[e^{\beta H_S}\mathrm{tr}_B\big[\pi_a \rho_S^{\mathrm{can}} \otimes \rho_B(0)\pi_a\big]$$

$$\times \mathrm{tr}_B[\pi_c \rho_S(0) \otimes \rho_B(0)\pi_c]\big], \tag{40}$$

$$\overline{G_{S2}} := Z_S \sum_{\substack{a,b \\ a \neq b}} \mathrm{tr}_S\big[e^{\beta H_S}\mathrm{tr}_B\big[\pi_a \rho_S^{\mathrm{can}} \otimes \rho_B(0)\pi_b\big]$$

$$\times \mathrm{tr}_B[\pi_b \rho_S(0) \otimes \rho_B(0)\pi_a]\big], \tag{41}$$

where $\pi_a := |E_a\rangle\langle E_a|$. We refer to $\overline{G_{S1}}$ as the diagonal term because it contains the diagonal ensemble [76], which is the long-time average of the density operator and will be defined in Sec. IV B. We also refer to $\overline{G_{S2}}$ as the off-diagonal term, because $\overline{G_{S2}}$ contains the off-diagonal matrix elements with respect to energy eigenstates of H. In a similar manner, we write $\overline{G_S^{\mathrm{can}}} = \overline{G_{S1}^{\mathrm{can}}} + \overline{G_{S2}^{\mathrm{can}}}$.

In this paper, we prove

$$|\overline{\delta G_S}| = o(1). \tag{42}$$

From the foregoing discussion, the left-hand side is divided as $\overline{\delta G_S} = (\overline{G_{S1}} - \overline{G_{S1}^{\mathrm{can}}}) + (\overline{G_{S2}} - \overline{G_{S2}^{\mathrm{can}}})$. In the next subsection, we will show that

$$|\overline{G_{S1}} - \overline{G_{S1}^{\mathrm{can}}}| = o(1), \tag{43}$$

$$|\overline{G_{S2}}| = o(1), \tag{44}$$

$$|\overline{G_{S2}^{\mathrm{can}}}| = o(1), \tag{45}$$

which imply Eq. (42).

We now discuss the relaxation time of δG_S. From Eq. (39), we can write G_S as

$$G_S = \sum_{i,j,k,l} e^{\beta(E_i^S - E_j^S)} p_k^S \langle \psi_k(t)|q_S^{il}|\psi_k(t)\rangle \langle \phi_j(t)|q_S^{li}|\phi_j(t)\rangle, \tag{46}$$

$$|\psi_k(0)\rangle := |p_k^S\rangle \otimes |E_{\mathrm{ini}}^B\rangle, \tag{47}$$

$$|\phi_j(0)\rangle := |E_j^S\rangle \otimes |E_{\mathrm{ini}}^B\rangle, \tag{48}$$

where the spectral decomposition of $\rho_S(0)$ is $\rho_S(0) = \sum_k p_k^S|p_k^S\rangle\langle p_k^S|$. As shown in Eq. (46), G_S is written as a combination of the expectation values of q_S^{il}. Also, the same rewrite is possible for G_S^{can}. In the total system SB, if the initial state of bath B is the microcanonical ensemble, the relaxation time of an observable of system S to its long-time average is independent of the bath size under some conditions on the matrix elements of the operator and the initial state of SB [67]. Those conditions are satisfied if the amplitudes of the matrix elements obey the Gaussian distribution, which is numerically confirmed in various quantum many-body systems satisfying the off-diagonal ETH [31,34]. Then we argue that the relaxation time of any system operator does not depend on the bath size also in our setup assuming the off-diagonal ETH. Since G_S and G_S^{can} are written as a combination of the expectation values of operators in S, the relaxation time of δG_S is also independent of the bath size.

We also discuss the relaxation time of δG_I to $\overline{\delta G_I}$. Since the interaction H_I is local, we argue that the relaxation time is $\Theta(1)$ by the same argument as above.

B. Outline of the proof

This subsection outlines the proof that the diagonal contribution $|\overline{G_{S1}} - \overline{G_{S1}^{\mathrm{can}}}|$ and the off-diagonal contributions $|\overline{G_{S2}}|$ and $|\overline{G_{S2}^{\mathrm{can}}}|$ vanish in the thermodynamic limit. See Appendix F for the complete proof. We also note that the interaction-induced error $|\overline{\delta G_I}|$ vanishes in the thermodynamic limit under the rotating wave approximation (see Appendix C for the proof).

First, we show the outline of the proof of $|\overline{G_{S1}} - \overline{G_{S1}^{\mathrm{can}}}| = o(1)$. We note that $\overline{G_{S1}}$ and $\overline{G_{S1}^{\mathrm{can}}}$ can be written as the expectation values of the observables of S. That is, the following relations hold:

$$\overline{G_{S1}} = \mathrm{tr}[O_S \rho^{\mathrm{DE}}], \tag{49}$$

$$\overline{G_{S1}^{\mathrm{can}}} = \mathrm{tr}\big[O_S^{\mathrm{can}} \rho^{\mathrm{can,DE}}\big], \tag{50}$$

where $\rho^{\mathrm{DE}} := \sum_a \pi_a \rho_S^{\mathrm{can}} \otimes \rho_B(0)\pi_a$ is the diagonal ensemble of $\rho_S^{\mathrm{can}} \otimes \rho_B(0)$ with respect to H and $O_S := Z_S \overline{\rho_S(t)}e^{\beta H_S}$ is an operator of system S. Similarly, $\rho^{\mathrm{can,DE}} := \sum_a \pi_a \rho_S^{\mathrm{can}} \otimes \rho_B^{\mathrm{can}}\pi_a$ is the diagonal ensemble of $\rho_S^{\mathrm{can}} \otimes \rho_B^{\mathrm{can}}$ with respect to H, and $O_S^{\mathrm{can}} := Z_S \overline{\rho_S^{\mathrm{can}}(t)}e^{\beta H_S}$ is an operator of system S. It is clear that the energy width of $\rho_B(0)$ is $\Theta(1)$. For the case of ρ_B^{can}, the large-deviation-type upper bound for the energy distribution has been obtained, and the energy width of ρ_B^{can} is $\Theta(N^{1/2})$ [77]. Both of these energy widths are smaller than $\Delta = \Theta(N^\alpha)(1/2 < \alpha < 1)$. Thus, we can define an energy shell with the energy width Δ, which includes both of the supports of $\rho_B(0)$ and ρ_B^{can}, and the contribution out of the shell is negligible in the thermodynamic limit.

When the energy eigenstates of H satisfy the strong diagonal ETH for any observable of system S, $O_S \simeq O_S^{\mathrm{can}}$ and $\mathrm{tr}_B[\rho^{\mathrm{DE}}] \simeq \mathrm{tr}_B[\rho^{\mathrm{can,DE}}]$ hold, implying that any observable of system S relaxes to the long-time average and the initial states of bath B cannot be distinguished by looking at system S alone. Then we obtain

$$\overline{G_{S1}} = \mathrm{tr}[O_S \rho^{\mathrm{DE}}] \simeq \mathrm{tr}\big[O_S^{\mathrm{can}} \rho^{\mathrm{can,DE}}\big] = \overline{G_{S1}^{\mathrm{can}}}, \tag{51}$$

which leads to $|\overline{G_{S1}} - \overline{G_{S1}^{\mathrm{can}}}| = o(1)$.

When the interaction between system S and bath B is weak, the above result can also be interpreted as follows. We rewrite O_S and O_S^{can} as $O_S = \overline{\rho_S(t)}(\rho_S^{\mathrm{can}})^{-1}$ and $O_S^{\mathrm{can}} = \overline{\rho_S^{\mathrm{can}}(t)}(\rho_S^{\mathrm{can}})^{-1}$, respectively. If the state of S relaxes to the canonical ensemble of H_S, i.e., $\overline{\rho_S(t)} \simeq \rho_S^{\mathrm{can}}$ and $\overline{\rho_S^{\mathrm{can}}(t)} \simeq \rho_S^{\mathrm{can}}$, $O_S \simeq 1_S$ and $O_S^{\mathrm{can}} \simeq 1_S$ hold, which leads to $\overline{G_{S1}} \simeq \overline{G_{S1}^{\mathrm{can}}}$. That is, $|\overline{G_{S1}} - \overline{G_{S1}^{\mathrm{can}}}| = o(1)$ holds when the system-bath interaction is weak and the state of system S relaxes to the canonical ensemble.

We next prove $|\overline{G_{S2}}| = o(1)$ and $|\overline{G_{S2}^{\mathrm{can}}}| = o(1)$ by using the off-diagonal ETH for the energy eigenstates of H and

observables in S. We first approximate $\overline{G_{S2}}$ as

$$\overline{G_{S2}} \simeq Z_S \sum_{\substack{a,b \in \text{sh} \\ a \neq b \\ i,j}} (q_S^{ij})_{ab}(q_S^{ji})_{ba}(\rho_S^{\text{can}} \otimes \rho_B(0))_{ab}$$

$$\times (\rho_S(0) \otimes \rho_B(0))_{ba}, \tag{52}$$

where the summation over the energy eigenstates of H is restricted to the energy shell. If the off-diagonal ETH for q_S^{ij} holds, the bath-size dependence of the off-diagonal matrix elements $(q_S^{ij})_{ab}$ is written as

$$(q_S^{ij})_{ab} = O(1)/\sqrt{D'}. \tag{53}$$

In addition, by using the Cauchy-Schwartz inequality and the fact that the purity of any state ρ is not larger than 1, we obtain

$$|\overline{G_{S2}}| \leqslant \frac{\Theta(1)}{D'} = o(1). \tag{54}$$

See details in Appendix F 2. In a similar manner, $|\overline{G_{S2}^{\text{can}}}| = o(1)$ is proved.

We next comment on the interaction-induced error δG_I defined in (32). As mentioned in Sec. III A, the interaction-induced error is zero when the condition (17) is satisfied. Even when the condition (17) is not satisfied, we can prove that

$$|\overline{\delta G_I}| = o(1) \tag{55}$$

holds under the rotating wave approximation (see Appendix C).

We here summarize the assumptions used in the foregoing proofs. In the proof of $|\overline{G_{S1}} - \overline{G_{S1}^{\text{can}}}| = o(1)$ [Eq. (43)], we used the (strong) diagonal ETH (14) of H for all observables of S, and the fact that the energy width of the canonical ensemble is narrower than $\Theta(N^a)$ ($1/2 < a < 1$). In the proof of $|\overline{G_{S2}}| = o(1)$ and $|\overline{G_{S2}^{\text{can}}}| = o(1)$ [Eqs. (44) and (45)], we used the (strong) off-diagonal ETH (15) of H for all observables of S.

C. Numerical results

In this subsection, we numerically show that $|\overline{\delta G_S}| = o(1)$ [Eq. (42)] and $|\overline{\delta G_I}| = o(1)$ [Eq. (55)]. To numerically obtain $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$, we first analytically obtain the expressions like Eqs. (40) and (41), and then numerically evaluate these expressions.

First, we show the numerical results for the long-time average of δG_S and δG_I. We note that $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$ are calculated for each initial eigenstate of bath B. Since we consider D eigenstates of B in the energy shell, D data respectively for $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$ are obtained in our numerical calculation. Figure 5 shows the boxplot of the dependence of $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$ on the bath size, which represents the distribution of $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$. Each datum in the boxplot corresponds to each initial eigenstate of bath B in the energy shell $[E - \Delta E, E]$. The middle line of the box represents the median, and the top (bottom) of the box represents the upper (lower) quartile. The half of the eigenstates in $[E - \Delta E, E]$ are included in the box. Figure 5 shows that $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$ decreases as N increases.

To investigate the N dependence of $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$ in more detail, we fit the numerical data of $\log |\overline{\delta G_S}|$ and $\log |\overline{\delta G_I}|$
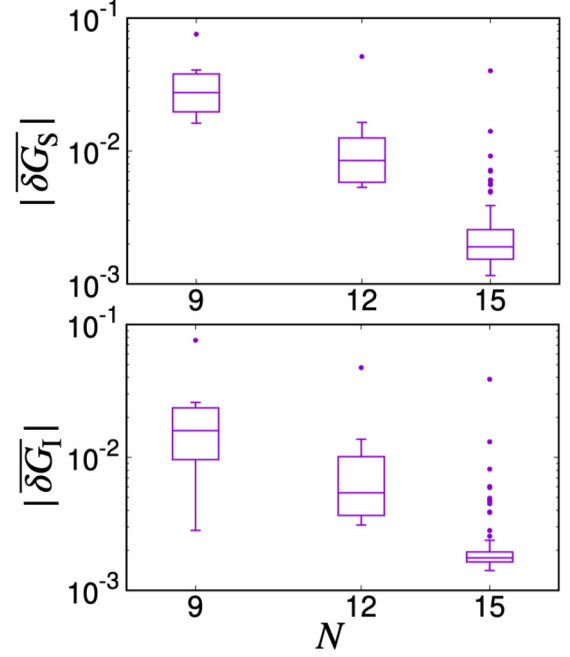

FIG. 5. The boxplot of $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$ to represent the dependence on the bath size and the initial state. Each datum in the boxplot corresponds to each initial eigenstate of bath B in the energy shell $[E - \Delta E, E]$. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\beta = 0.1$, $\gamma' = \gamma$. Both $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$ tend to decrease as N increases.

against a fitting function $-a \log N + b$ with the fitting parameters a and b. The positive a implies $|\overline{\delta G_S}| = o(1)$ and $|\overline{\delta G_I}| = o(1)$. Figure 6 shows the γ' dependence of a, where we recall that γ' is the coupling strength of the interaction between system S and bath B. Our result shows that a for $|\overline{\delta G_S}|$ is indeed positive for any γ'. Thus, $|\overline{\delta G_S}| = o(1)$ holds, which is consistent with the theory (42).

We note that system S is decoupled from bath B in the limit of $\gamma' \to 0$. Also, system S and the support of H_I are decoupled from the rest of the total system SB in the limit of $\gamma' \to \infty$. These decouplings imply that the ETH is no longer relevant to system S, and thus the fluctuation theorem does not hold. Thus, it is reasonable that a is small in the region of $\gamma'/\gamma \simeq 0$ or $\gamma'/\gamma \gg 1$ because of the large finite-size effect. If the bath size is much larger than N currently used, the finite-size effect is expected to become less significant, as a similar mechanism observed in Ref. [42].

For the interaction-induced error $|\overline{\delta G_I}|$, the fitting parameter is positive within the margin of numerical errors as shown in Fig. 6, which is consistent with (55), while the numerical errors are larger when γ' is small. We consider this is because the finite-size effect of the ETH is larger there.

V. SHORT-TIME REGIME

In this section we investigate the fluctuation theorem with the energy eigenstate bath in the short-time regime. While Ref. [44] theoretically showed that $|\delta G_S| = o(1)$ holds on the basis of the Lieb-Robinson bound and the ETH, no quantitative numerical calculations of the bath-size dependence of the error were made. In the present paper, we numerically inves-
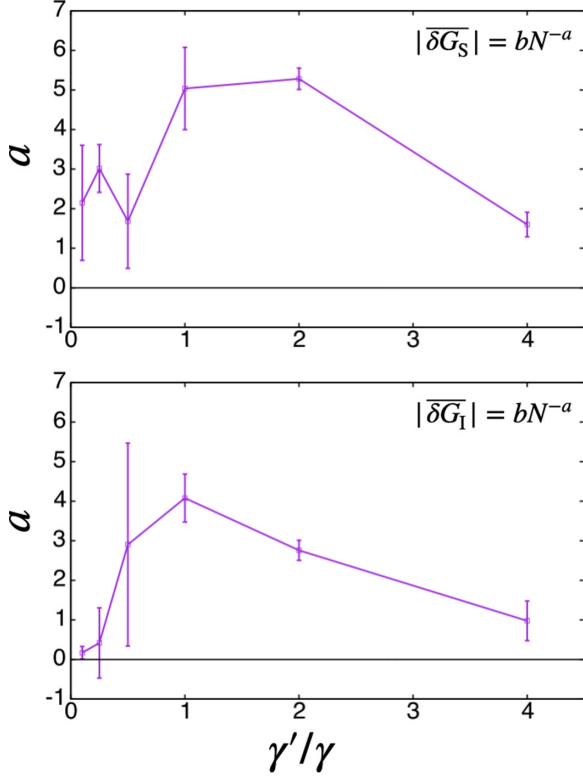
FIG. 6. The γ' dependence of the exponent a, which is obtained by fitting the medians of $\log|\overline{\delta G_S}|$ and $\log|\overline{\delta G_I}|$ against $-a \log N + b$. The fitting is performed with error, and the error bars of a comes from it. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\beta = 0.1$. For both $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$, we obtain $a > 0$, which the theories (42) and (55) support. When γ' is near 0 or γ' is too large, the finite-size effect is large and a tends to be small.

tigate the dependence of the error of the fluctuation theorem on the bath size and the initial state. In Sec. V A we define the error and briefly overview the theory developed in Ref. [44]. In Sec. V B we show our numerical results, which support the theory.

A. Definition of the error

In this subsection, we discuss the error of the fluctuation theorem in the short-time regime. Specifically, we decompose δG_S introduced in Eq. (30) into the term that can be evaluated by the Lieb-Robinson bound and the term that can be evaluated by the ETH.

From the Lieb-Robinson bound, when considering the dynamics of the short-time regime, the effect of the far side of bath B (B_1 in Fig. 4) on system S is negligible. Using H_{B_0} which is given by restricting H_B to B_0, we define the truncated Hamiltonian H_T and the corresponding time evolution operator U_T as $H_T := H_S + H_I + H_{B_0}$ and $U_T := \exp(-iH_Tt/\hbar)$. Then we define G_S^T and $G_S^{\text{can,T}}$ as

$$G_S^T := \text{tr}[e^{\beta H_S}U_Te^{-\beta H_S}\rho_B(0)U_T^\dagger \rho_S(t)], \tag{56}$$

$$G_S^{\text{can,T}} := \text{tr}[e^{\beta H_S}U_Te^{-\beta H_S}\rho_B^{\text{can}}U_T^\dagger \rho_S^{\text{can}}(t)]. \tag{57}$$

Due to the Lieb-Robinson bound, $G_S^T \simeq G_S$ and $G_S^{\text{can,T}} \simeq G_S^{\text{can}}$ hold in $t \ll \tau_{\text{LR}}$. Using these arguments, we decompose δG_S as

$$\delta G_S = \delta G_{\text{LR}} + \delta G_{\text{ETH}}, \tag{58}$$

where we define

$$\delta G_{\text{LR}} := \delta G_{\text{LR}}^{(1)} + \delta G_{\text{LR}}^{(2)}, \tag{59}$$

$$\delta G_{\text{LR}}^{(1)} := G_S - G_S^T, \tag{60}$$

$$\delta G_{\text{LR}}^{(2)} := G_S^{\text{can,T}} - G_S^{\text{can}}, \tag{61}$$

$$\delta G_{\text{ETH}} := \delta G_S - \delta G_{\text{LR}} = G_S^T - G_S^{\text{can,T}}. \tag{62}$$

We can evaluate these terms as follows.

Using the Lieb-Robinson bound [45,46], we can show that

$$|\delta G_{\text{LR}}| = o(1) \ (t \ll \tau_{\text{LR}}) \tag{63}$$

holds. In this evaluation, the ETH is not used, while it is essential that H_B and H_I are local. In Ref. [44] it has been shown that

$$|\delta G_{\text{LR}}^{(1,2)}| \leqslant Ce^{-\kappa\ell}t^2 \tag{64}$$

holds in $t \ll \tau_{\text{LR}}$, where $\ell := \text{dist}(\tilde{S}, \partial B)$ is the distance between $\tilde{S}$ and ∂B, and $\tilde{S}$ is the area that consists of system S and the support of H_I. The positive constants C and κ are independent of N. Since ℓ is increasing as the bath size increases, Eq. (64) implies Eq. (63). Furthermore, using the improved Lieb-Robinson bound [47], we can show that

$$|\delta G_{\text{LR}}^{(1,2)}| \leqslant C'\frac{t^\ell}{\ell!} \simeq C'e^{-\ell(\log\ell-1)}t^\ell \tag{65}$$

holds in $t \ll \tau_{\text{LR}}$ with C' being a positive constant independent of N. The evaluation of (65) is tighter than (64), where the power of t depends on ℓ.

We next discuss δG_{ETH}. Under the assumption that the initial state $|E_a^B\rangle$ of bath B satisfies the ETH of H_B for observables in B_0, we show that

$$|\delta G_{\text{ETH}}| \simeq |\text{tr}_{B_0}[L_{B_0}\text{tr}_{B_1}[\rho_B - \rho_B^{\text{can}}]]| \tag{66}$$

$$= o(1), \tag{67}$$

where $L_{B_0} := \text{tr}_S[U_T^\dagger \rho_S(t)e^{\beta H_S}U_Te^{-\beta H_S}]$. In Eq. (67), we use the ETH for L_{B_0}. The initial state of bath B must be indistinguishable from the canonical ensemble, because the time evolution is restricted to U_T and the far part B_1 can be neglected.

If the bath size is small [specifically, $O(1)$], τ_{LR} and τ_{relax} becomes comparable [both $O(1)$], which cannot be distinguished in practice. Therefore, in our numerical calculations in the next subsection, we focus only on the time regime $t \lesssim \tau_{\text{relax}}$ in which the error of the fluctuation theorem is proportionate to t^2. We calculate the coefficients in our numerical calculations, a_{ETH}, a_I and a_{LR} defined as

$$|\delta G_{\text{ETH}}| = a_{\text{ETH}}t^2, \tag{68}$$

$$|\delta G_I| = a_It^2, \tag{69}$$

$$|\delta G_{\text{LR}}| = a_{\text{LR}}t^b, \tag{70}$$

where b is a fitting parameter determined by numerical calculation. We will show the t^2 dependence in Eqs. (68) and (69) in Appendix E. Since Eq. (65) is an inequality, b does not necessarily equal ℓ. We note that b can depend on the bath size. Thus, we find that the initial t^2 behavior in δG_S originates from δG_{ETH}, not from δG_{LR}.

In order for the time dependence (68) and (70) to continue until the relaxation time of $\Theta(1)$ and for Eqs. (63) and (67) to hold,

$$a_{LR} = o(1), \qquad (71)$$

$$a_{ETH} = o(1) \qquad (72)$$

must be satisfied. We will numerically show these equations in the next subsection.

Using the perturbation theory and the off-diagonal ETH (see Appendix G 2), we can show that

$$|\delta G_I| = o(1). \qquad (73)$$

holds in $t \ll \tau_{LR}$. In order for the time dependence (69) to continue until the relaxation time of $\Theta(1)$ and for Eq. (73) to hold,

$$a_I = o(1) \qquad (74)$$

must be satisfied. This will be numerically shown in the next subsection.

We here remark on the two separate timescales appeared above. The time interval satisfying $\tau_{relax} \lesssim t \ll \tau_{LR}$ clearly exists for a sufficiently large bath size N, while such a large bath size is not accessible by our numerical calculations but may be achieved by real experiments with ultracold atoms. In our numerical simulation (see Fig. 13 in Appendix H), only $\tau_{relax} \lesssim t \lesssim \tau_{LR}$ can be guaranteed for the bath size $N = 15$. We should further remark that τ_{relax} can depend on the system-bath coupling γ', and the bath size N required for the existence of the time interval $\tau_{relax} \lesssim t \lesssim \tau_{LR}$ increases as γ' decreases. In our numerical simulation, such a time interval exists for $\gamma'/\gamma = 1, 1.5, 2, 3, 4$, while it does not clearly exist for $\gamma'/\gamma = 0.05, 0.1, 0.4$ where our N is not sufficiently large.

B. Numerical results

In this subsection, we numerically show that Eqs. (71), (72), and (74) hold. We choose three sites close to system S as B_0 as shown in Fig. 7.

Using the fitting functions $|\delta G_{ETH}| = a_{ETH} t^{b_{ETH}}$ and $|\delta G_I| = a_I t^{b_I}$ and numerical data in $10^{-3} < \gamma t < 10^{-2}$, we determine the fitting parameters a_{ETH}, b_{ETH}, a_I, and b_I. Because the effect of the subleading terms are negligible in this time region, we choose this time region for simplicity. Regardless of the parameters of the Hamiltonian, we obtain $b_{ETH} = 2$ and $b_I = 2$ with negligibly small numerical errors. Figure 8 is a boxplot showing the dependence of a_{ETH} and a_I on the bath size and the initial state. Each datum in the boxplot corresponds to each energy eigenstate in the energy shell $[E - \Delta E, E]$. Both a_{ETH} and a_I tend to decrease as the bath size increases.

To investigate the N dependence of a_{ETH} and a_I, we fit the medians of $\log a_{ETH}$ and $\log a_I$ against the fitting functions $\log a_{ETH} = c_{ETH} - \eta_{ETH} \log N$ and $\log a_I = c_I - \eta_I \log N$. As
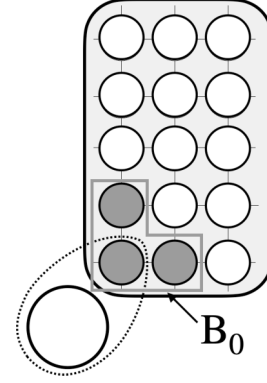

FIG. 7. B_0 used in our numerical simulation in Sec. V B.

a result, we obtain $\eta_{ETH} = 1.4 \pm 0.8$, $\eta_I = 2.0 \pm 0.8$ for $\gamma' = \gamma$, $g = 0.1\gamma$ and $\eta_{ETH} = 1.8 \pm 0.2$, $\eta_I = 2.1 \pm 0.1$ for $\gamma' = \gamma$, $g = 0.4\gamma$. These results support the theories based on the ETH (72) and (74). We note that these values are almost independent of γ'. This is understood from the fact that $a_{ETH} = o(1)$ is brought by the ETH of H_B, which is independent of the interaction between system S and bath B.

We next show our numerical results on the errors related to the Lieb-Robinson bound. We fit the median of $|\delta G_{LR}^{(1)}|$ against the function $a_{LR}^{(1)} t^{b_{LR}^{(1)}}$. Because $\delta G_{LR}^{(2)}$ cannot be fitted by the fitting function $a_{LR}^{(2)} t^{b_{LR}^{(2)}}$, we numerically find $a_{LR}^{(2)} t^{b_{LR}^{(2)}}$ only satisfying $|\delta G_{LR}^{(2)}| \leqslant a_{LR}^{(2)} t^{b_{LR}^{(2)}}$ in $10^{-3} < \gamma t < 10^{-2}$. As a result, we obtain $b_{LR}^{(1,2)} = 4$, which is different from $b = 2$ in Eq. (63). Figure 9 is a boxplot showing the dependence of $a_{LR}^{(1)}$
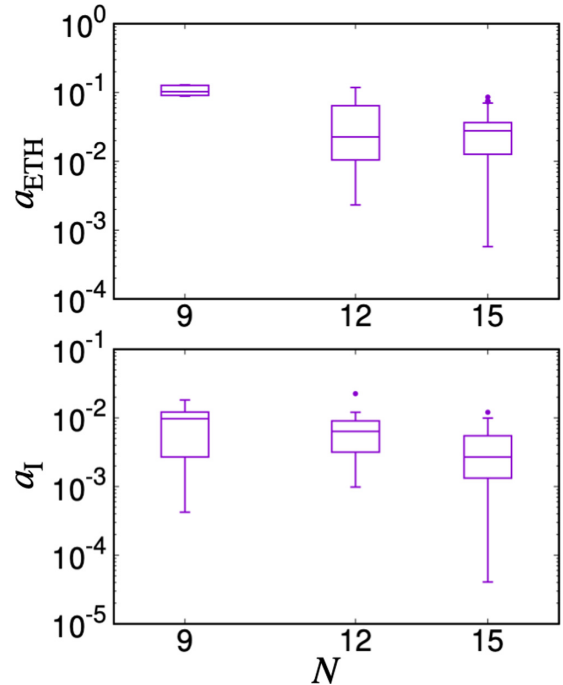

FIG. 8. The dependence of a_{ETH} and a_I on the bath size and the initial state. Each datum in the boxplot corresponds to each initial eigenstate of bath B in the energy shell $[E - \Delta E, E]$. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\gamma' = \gamma$, $\beta = 0.1$. Both a_{ETH} and a_I tend to decrease as N increases.
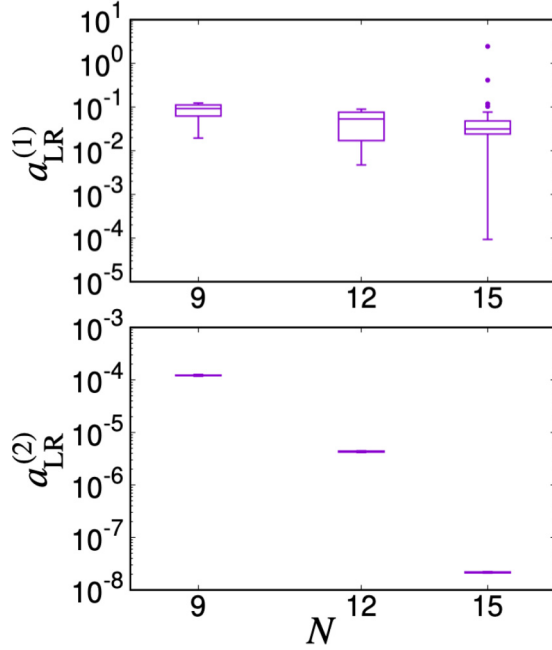
FIG. 9. The bath-size dependence of $a_{LR}^{(1)}$ and $a_{LR}^{(2)}$, which relate to the Lieb-Robinson bound. Each datum in the boxplot corresponds to each initial eigenstate of bath B in the energy shell $[E - \Delta E, E]$. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\gamma' = \gamma$, $\beta = 0.1$. Both $a_{LR}^{(1)}$ and $a_{LR}^{(2)}$ tend to decrease as N increases. For $a_{LR}^{(1)}$ at $N = 15$, we note that the vertical range of the graph is widened because there are outliers on the smaller error side. The initial state dependence is small for $a_{LR}^{(2)}$, implying that the change of $a_{LR}^{(2)}$ with respect to β is small when the initial state of bath B is the canonical ensemble.

and $a_{LR}^{(2)}$ on the bath size and the initial state. Both $a_{LR}^{(1)}$ and $a_{LR}^{(2)}$ tend to decrease as N increases.

Finally, we compare the time dependence of the errors of the fluctuation theorem with that of an ordinary observable of system S. As mentioned in Sec. III, the change of the occupation number in system S (23) initially rises in the form of $a_n t^2$, which is apparently similar to the case of the fluctuation theorem (E1). Figure 10 shows the dependence of a_n on the bath size and the initial state of B. We observe that a_n does not depend on the bath size, i.e., $a_n = \Theta(1)$. This bath-size independence is reasonable because there is no physical mechanism for the change of the occupation number in system S to be suppressed in the thermodynamic limit. This result $a_n = \Theta(1)$ is in contrast with $a_{ETH} = o(1)$ and $a_{LR} = o(1)$ about the fluctuation theorem. Then, while the initial rise of the error of the fluctuation theorem is also proportionate to t^2, this is due to a different physical mechanism from the case of ordinary observables such as the occupation number. Thus, the above numerical result and discussion again support the scenario that the fluctuation theorem in the short-time regime holds with the nontrivial thermal mechanism based on the Lieb-Robinson bound and the ETH.

VI. SUMMARY AND DISCUSSION

In this study, we have investigated the fluctuation theorem in the long- and short-time regimes, when the initial state of
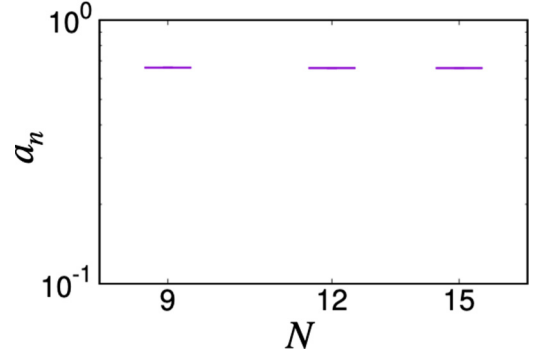

FIG. 10. The bath-size dependence of a_n. Each datum in the boxplot corresponds to each initial eigenstate of bath B in the energy shell $[E - \Delta E, E]$. We note that the boxes are squashed because the initial-state dependence of a_n is quite small. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\gamma' = \gamma$, $\beta = 0.1$. a_n is independent of the bath size N, which is reasonable because the size of system S does not depend on the bath size N.

bath B is an energy eigenstate and the time evolution of the total system is unitary. Our results theoretically and numerically show that the fluctuation theorem holds in the entire time domain.

In the long-time regime, we have considered the long-time average of $\langle e^{-\sigma} \rangle$. We theoretically showed that the error of the fluctuation theorem introduced in Eq. (30) vanishes in the thermodynamic limit of the heat bath (Sec. IV). The main assumptions used are the diagonal and off-diagonal ETH (14) (15). We have also shown that the interaction-induced error (32) decreases with increasing the bath size by using the rotating wave approximation (C2). We numerically investigated the dependence of the error of the fluctuation theorem on the bath size and the initial state. Figures 5 and 6 support the above theory.

We remark that the fluctuation theorem in the long-time regime can also be derived from other assumptions than ours. For example, Ref. [71] is based on some natural assumptions on transition probabilities. While those assumptions are not equivalent to our assumptions such as the ETH of H discussed in Sec. IV, their result and ours are both correct and would play complementary roles.

In the short-time regime, the fluctuation theorem has been theoretically shown in Ref. [44] on the basis of the ETH and the Lieb-Robinson bound. We performed systematic numerical calculations to show the validity of the fluctuation theorem in the short-time regime (Sec. V). In particular, we focus on the dependence of the errors of the fluctuation theorem on the bath size, which establishes that the validity of the fluctuation theorem is due to the theoretically proposed scenario based on the ETH and the Lieb-Robinson bound [44], rather than a trivial scenario argued in the last paragraph of Sec. V. Figures 8–10 show that our numerical results support this theoretical scenario.

The two time regimes play key roles in this study: The long- and short-time regimes. The long-time regime is defined by that $\langle e^{-\sigma} \rangle$ nearly equals the long-time average $\overline{\langle e^{-\sigma} \rangle}$, which is independent of the bath size as discussed in Sec. III B. The short-time regime is defined by the system in-

formation's not reaching the far part of bath B, which becomes longer as the bath size increases. Therefore, the long- and short-time regimes overlap and cover the entire time domain when the bath size is large enough. We again emphasize, however, that the fluctuation theorem has been shown independently in these time regimes.

We remark that our result highlights the connection between information and thermodynamics [78–80]. The informational entropy (the von Neumann entropy) and the thermodynamic quantity (heat) are quantitatively connected to each other in the second law of thermodynamics (1), as historically demonstrated by the Szilard engine [81] and the Landauer principle [82]. While in the conventional theory [9] this connection between the informational entropy and heat relies on the assumption that the initial state of the bath is canonical and thus has the maximum entropy, the present work shows that the same connection emerges even when the initial state of the bath is a single energy eigenstate. Therefore, our result serves as a theoretical foundation of thermodynamics of information beyond the conventional canonical setup.

Meanwhile, in our numerical calculation, we used a specific model of the two-dimensional system introduced in Sec. II C. However, our theory ensures that the fluctuation theorem should hold for a much broader class of models that satisfy the ETH (Table I). Since it has been numerically confirmed that the ETH holds for various nonintegrable systems [29–42], the fluctuation theorem holds for these systems.

We consider that our theory can be experimentally verified. In particular, the dynamics of isolated nonintegrable many-body systems have been investigated using artificial quantum systems such as cold atoms [19–24], trapped ions [25], and superconducting qubits [26]. Not only local physical quantities but also informational entropy are experimentally measurable [20,26]. Because our numerical calculation is performed with a small bath size of 15 sites using numerically exact diagonalization, we cannot observe the separation between the two timescales, the Lieb-Robinson time τ_{LR} and the relaxation time τ_{relax}, as mentioned in Sec. V. However, experiments with around 100 to 400 sites are currently accessible [22–24]. Therefore, the separation of the above timescales would be experimentally observable. It is an interesting future issue to directly verify the theory of the long- and short-time regimes by such real experiments, which would open up the experimental investigation of the emergence of thermodynamics from quantum mechanics.

Note added. Recently we became aware of the results of J. Gemmer *et al.* [71], which are closely related to the present work.

APPENDIX A: ABSOLUTE IRREVERSIBILITY

Absolute irreversibility means a violation of the fluctuation theorem, which occurs when the initial state of system S is singular in the sense that it does not have the full support in the Hilbert space (e.g., the initial state is pure) [48,49]. This is because the final state of the reverse process is not necessarily in the support of the initial state of the forward process. The violation of the fluctuation theorem can be described by a correction term as described later. In this Appendix we consider the generalized fluctuation theorem including the correction term in the presence of absolute irreversibility.

First, we consider the conventional case where the initial state of bath B is the canonical ensemble. We denote the projection operator onto the subspace which supports $\rho_{\mathrm{S}}(0)$ by $P_{\mathrm{ini}}^{\mathrm{S}}$. When $D_{\mathrm{S}} > \mathrm{rank}[P_{\mathrm{ini}}^{\mathrm{S}}]$, the fluctuation theorem is accompanied by the correction term due to absolute irreversibility. In particular, if $\rho_{\mathrm{S}}(0)$ is pure, $\mathrm{rank}[P_{\mathrm{ini}}^{\mathrm{S}}] = 1 < D_{\mathrm{S}}$ holds. Then it has been shown in Refs. [48,49] that

$$\langle e^{-\sigma} \rangle = 1 - \lambda^{\mathrm{can}}(t), \tag{A1}$$

$$\lambda^{\mathrm{can}}(t) := \mathrm{tr}\big[\big(1 - P_{\mathrm{ini}}^{\mathrm{S}}\big)U^{\dagger}\rho_{\mathrm{R}}^{\mathrm{can}}(0)U\big], \tag{A2}$$

where $\rho_{\mathrm{R}}^{\mathrm{can}}(0) := \rho_{\mathrm{S}}^{\mathrm{can}}(t) \otimes \rho_{\mathrm{B}}^{\mathrm{can}}$ is the initial state of the reverse process. We note that one can measure $\langle e^{-\sigma} \rangle$ and $\lambda^{\mathrm{can}}(t)$ in the forward and the reverse processes independently. In fact, Eq. (A1) has been confirmed experimentally using projection measurements on the system [55].

The correction term (A2) does not appear if the singularity is removed by regularizing the initial state of system S as follows:

$$\rho_{\mathrm{S}}^{\mathrm{reg}}(0) := \frac{\tilde{\rho}_{\mathrm{S}}(0)}{\mathrm{tr}_{\mathrm{S}}[\tilde{\rho}_{\mathrm{S}}(0)]}, \tag{A3}$$

$$\tilde{\rho}_{\mathrm{S}}(0) := \rho_{\mathrm{S}}(0) + \varepsilon\big(1 - P_{\mathrm{ini}}^{\mathrm{S}}\big) \quad (\varepsilon > 0), \tag{A4}$$

which is essentially the same approach as adopted in Ref. [71] (see also Ref. [83]). If we take the limit of $\varepsilon \to 0$ after calculating $\langle e^{-\sigma} \rangle$, then $\langle e^{-\sigma} \rangle = 1$ holds. On the other hand, if we take the limit $\varepsilon \to 0$ first and then calculate $\langle e^{-\sigma} \rangle$, Eq. (A1) holds. This is not just the problem of the (unphysical) order of the limits. As mentioned above, however, Eq. (A1) has been experimentally observed, and thus the effect of absolute irreversibility is physically relevant. On the other hand, in the case where small noise is unavoidable in the initial state, the regularization approach would be relevant. Therefore, these two approaches are both physically reasonable and play complementary roles.

We next consider the fluctuation theorem with the energy eigenstate bath in the long-time regime in the presence of absolute irreversibility. We define the correction term as

$$\lambda(t) := \mathrm{tr}\big[\big(1 - P_{\mathrm{ini}}^{\mathrm{S}}\big)U^{\dagger}\rho_{\mathrm{R}}(0)U\big], \tag{A5}$$

where $\rho_{\mathrm{R}}(0) := \rho_{\mathrm{S}}(t) \otimes \rho_{\mathrm{B}}(0)$. We will prove in Appendix F that

$$\langle e^{-\sigma} \rangle = 1 - \lambda(t) \tag{A6}$$

holds in the thermodynamic limit. Because we have already discussed the long-time average of the left-hand side in
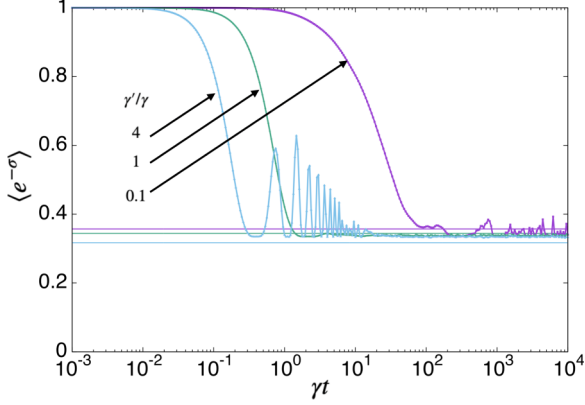
FIG. 11. The time dependence of $\langle e^{-\sigma} \rangle$ when the initial state of system S is pure ($p = 1$). Parameters: $g = 0.1\gamma$, $\gamma' = 0.1\gamma$ (purple), γ (green), 4γ (blue). The initial state of bath B is an energy eigenstate of H_B, whose energy is maximum in the energy shell at $\beta = 0.1$. The onsite potential ω is determined by $\text{tr}_B[n_B \rho_B^{\text{can}}] = N_P$. The horizontal lines represent the numerically obtained values of $1 - \overline{\lambda(t)}$.

Sec. IV B, we only need to show

$$|\overline{\lambda^{\text{can}}(t)} - \overline{\lambda(t)}| = o(1). \quad (A7)$$

We will prove Eq. (A7) by using the ETH of the total system SB in Appendix F.

We here show the validity of Eq. (A6) by numerical simulation. Figure 11 shows the time dependence of $\langle e^{-\sigma} \rangle$ when the initial state of system S is a pure state ($p = 1$). In the long-time regime, $\langle e^{-\sigma} \rangle$ deviates from 1 because absolute irreversibility occurs. The convergent value nearly equals $1 - \overline{\lambda(t)}$, implying the fluctuation theorem with the correction term (A6) holds. We note that in the numerics of Ref. [44] ($L_x = L_y = 4$, $N_P = 4$), we have numerically checked that Eq. (A6) holds by performing the same analysis as in this paper.

APPENDIX B: SUPPLEMENT ON THE INTERACTION-INDUCED ERROR

In this Appendix we discuss the form of the interaction-induced error. Specifically, we compare δG_I defined in Eq. (32) and the interaction-induced error defined in Ref. [44]. We show that δG_I reflects how thermal the initial state of bath B is.

First, we consider the characteristic function of the entropy production:

$$\mathcal{G}(u; \rho_B) := \text{tr}[U e^{-iu\beta H_B} e^{iu \ln \rho_S(0)}$$
$$\times \rho_S(0) \otimes \rho_B U^\dagger e^{-iu \ln \rho_S(t)} e^{iu\beta H_B}], \quad (B1)$$

which is the Fourier transformation of the probability distribution of the stochastic entropy production σ. We note that $\langle e^{-\sigma} \rangle = \mathcal{G}(i; \rho_B)$ holds. Let us next consider the following function by replacing H_B by $-H_S$, as in G_S defined in Eq. (26):

$$\mathcal{G}_S(u; \rho_B) := \text{tr}[U e^{iu\beta H_S} e^{iu \ln \rho_S(0)}$$
$$\times \rho_S(0) \otimes \rho_B U^\dagger e^{-iu \ln \rho_S(t)} e^{-iu\beta H_S}]. \quad (B2)$$

The interaction-induced error comes from the difference between Eqs. (B1) and (B2). By noting that the fluctuation

theorem holds if the initial state of B is canonical, we define the interaction-induced error as

$$\delta G_I := \mathcal{G}(i; \rho_B(0)) - \mathcal{G}_S(i; \rho_B(0)) + \mathcal{G}_S(i; \rho_B^{\text{can}}) - \mathcal{G}(i; \rho_B^{\text{can}}). \quad (B3)$$

By definition, $\delta G_I = 0$ holds when the initial state of bath B is the canonical ensemble [$\rho_B(0) = \rho_B^{\text{can}}$].

As an alternative definition, in the supplemental information of Ref. [44], the interaction-induced error is introduced using the operator norm as

$$\delta_I(u) := \|e^{iu\beta H_B} U e^{-iu\beta H_B} - e^{-iu\beta H_S} U e^{iu\beta H_S}\|. \quad (B4)$$

Using $\delta_I(u)$, the error of the fluctuation theorem is evaluated as

$$|\mathcal{G}(u; \rho_B(0)) - \mathcal{G}_S(i; \rho_B(0))| \leqslant \delta_I(u) C(u), \quad (B5)$$

where $C(u)$ is an N-independent constant and written as

$$C(u) = \begin{cases} (p_b(0))^{-|u_I - 1|} > 1 & (1 < u_I) \\ 1 & (0 \leqslant u_I \leqslant 1) \\ (p_b(t))^{-|u_I|} > 1 & (u_I < 0), \end{cases} \quad (B6)$$

where $p_b(t)$ is the minimum eigenvalue of $\rho_S(t)$ and $u_I := \text{Im}[u]$. In the supplemental information of Ref. [44], $C(u)$ in the right-hand side of (B5) was considered to be 1, while the modification as above is required for $1 < u_I$ and $u_I < 0$. We note that only the case of $u_I = 1$ was used in the proof of the fluctuation theorem of the form (4) in Ref. [44], and therefore the proof of it is correct as it is.

In the above evaluation, $\delta_I(u)$ is independent of the initial state of bath B. On the other hand, $\delta G_I = 0$ holds when the initial state of bath B is the canonical ensemble. Therefore, δG_I is a more proper measure of the error, because it can reflect how thermal the initial state of bath B is.

APPENDIX C: ROTATING WAVE APPROXIMATION AND THE INTERACTION-INDUCED ERROR

In this Appendix we show that Eqs. (18) and (36) hold under the rotating wave approximation and the off-diagonal ETH. We first introduce the rotating wave approximation and show Eq. (18) in Appendix C 1. Then we show Eq. (36) in Appendix C 2.

1. Rotating wave approximation and the proof of Eq. (18)

We write the interaction Hamiltonian as

$$H_I = \sum_{ab} (H_I)_{ab} |E_a\rangle\langle E_b|. \quad (C1)$$

Let Ω be the cutoff frequency for the rotating wave approximation, which is assumed to be independent of the system size. We then introduce the rotating wave approximation of H_I as

$$\tilde{H}_I = \sum_{ab}' (H_I)_{ab} |E_a\rangle\langle E_b|, \quad (C2)$$

where the sum is taken over (a, b) satisfying $|E_a - E_b| < \Omega$. The rotating wave approximation holds in a long-time regime with $t \gg \Omega^{-1}$.

To show Eq. (18), we assume the off-diagonal ETH in the following form [28,37]:

$$(\tilde{H}_I)_{ab} = \frac{1}{\sqrt{D'}} g_{ab} r_{ab}, \tag{C3}$$

where g_{ab} characterizes the dependence of $(\tilde{H}_I)_{ab}$ on E_a and E_b, and r_{ab} is a random variable whose mean is 0 and variance is 1. We note that Eq. (C3) is stronger than the off-diagonal ETH in the form of Eq. (15). We use Eq. (C3) only in this Appendix.

Under the rotating wave approximation and the off-diagonal ETH, the commutator in Eq. (18) is written as

$$[H_S + H_B, \tilde{H}_I] = {\sum_{ab}}' |E_a\rangle\langle E_b| \frac{R_{ab}}{\sqrt{D'}}, \tag{C4}$$

where $R_{ab} := (E_a - E_b) g_{ab} r_{ab}$. We note that R_{ab} is the matrix elements of the band random matrix and the variance of R_{ab} is less than $\Omega^2 \|\tilde{H}_I\|^2$. From the Wigner semicircle law, the maximum eigenvalue of a random matrix, whose dimension is D and variance of the matrix elements is σ^2, is about $2\sqrt{D}\sigma$. Similarly, the maximum eigenvalue of a band random matrix with band width W is evaluated to be $2\sqrt{W}\sigma$, which is mathematically shown for some band random matrix [84]. We evaluate the band width of the band random matrix R as $D'\Omega/\|\tilde{H}\|$, where we defined $\tilde{H} := H_S + H_B + \tilde{H}_I$. Then the operator norm of Eq. (C4) is evaluated as

$$\|[H_S + H_B, \tilde{H}_I]\| \leqslant \frac{c\Omega\|\tilde{H}_I\|}{\sqrt{N}}, \tag{C5}$$

where we used $\|\tilde{H}\| = \Theta(N)$ and introduced $c := \Theta(1)$. Thus, Eq. (18) is proved.

In the same manner, we can show that

$$\|f_n\| \leqslant \frac{c}{\sqrt{N}} \|\tilde{H}_I\| \Omega^n, \tag{C6}$$

$$\|g_n^a\| \leqslant \frac{ca}{\sqrt{N}} \|\tilde{H}_I\|^a \Omega^n, \tag{C7}$$

where we inductively define

$$f_{n+1} := [\tilde{H}, f_n], \quad f_0 = A, \tag{C8}$$

$$g_{n+1}^a := [\tilde{H}, g_n^a], \quad g_0^a = \tilde{H}_I^a, \tag{C9}$$

$$A := [\tilde{H}_I, \tilde{U}^\dagger], \tag{C10}$$

where $\tilde{U} = \exp(-i\tilde{H}t/\hbar)$. We note that Eq. (C5) is a special case ($a = n = 1$) of Eq. (C7).

2. Proof of Eq. (36)

In this subsection, we show Eq. (36). We first focus on $\delta G_I^{(1)}$:

$$\delta G_I^{(1)} = \text{tr}[\delta\tilde{U}_\beta \rho \tilde{U}^\dagger \Lambda_S(t)], \tag{C11}$$

where $\delta\tilde{U}_\beta := e^{-\beta H_0}\tilde{U}e^{\beta H_0} - \tilde{U}$, $H_0 := H_S + H_B$, $\rho := \rho_S^{\text{can}} \otimes \rho_B(0)$, and $\Lambda_S(t) := \rho_S(t)(\rho_S^{\text{can}})^{-1}$. Using the Cauchy-Schwartz inequality, we bound $(\delta G_I^{(1)})^2$ from above as

$$(\delta G_I^{(1)})^2 \leqslant \text{tr}[\tilde{U}\rho\tilde{U}^\dagger \Lambda_S(t)\Lambda_S^\dagger(t)]\text{tr}[\delta\tilde{U}_\beta \rho \delta\tilde{U}_\beta^\dagger].$$

Since $|\text{tr}[\tilde{U}\rho\tilde{U}^\dagger \Lambda_S(t)\Lambda_S^\dagger(t)]| = \Theta(1)$, we focus on

$$\text{tr}[\delta\tilde{U}_\beta \rho \delta\tilde{U}_\beta^\dagger] = \text{tr}[\delta\tilde{U}_{2\beta}^\dagger \tilde{U}\rho] - 2\text{tr}[\delta\tilde{U}_\beta^\dagger \tilde{U}\rho]$$

$$=: \sum_{n=2} \beta^n B_n. \tag{C12}$$

In the first line, we used $[H_0, \rho] = 0$. In the second line, we used the Baker-Campbell-Hausdorff formula, respectively, for $\delta\tilde{U}_{2\beta}^\dagger$ and $\delta\tilde{U}_\beta^\dagger$. We will show that $|B_n| = o(1)$ in the following.

We first consider $|B_2|$. We can write B_2 as

$$B_2 = \text{tr}[[H_0, [H_0, \tilde{U}^\dagger]]\tilde{U}\rho]$$

$$= -\text{tr}[(f_1 - \tilde{H}_I f_0 + f_0 \tilde{H}_I)U\rho]. \tag{C13}$$

Then we obtain

$$|B_2| \leqslant (\|f_1\| + 2\|\tilde{H}_I\|\|f_0\|)\|\tilde{U}\|\|\rho\|_1$$

$$\leqslant \frac{c}{\sqrt{N}}\|\tilde{H}_I\|(\Omega + \|\tilde{H}_I\|)$$

$$= o(1), \tag{C14}$$

where we used Eq. (C6) in the second line.

We next consider the case of $n \geqslant 3$. B_n is written as a linear combination of terms consisting of the product of f_m, g_m^a, $\tilde{H}_I$ ($m \leqslant n$) and each term contains a single f_n. Since Eqs. (C6) and (C7), the leading terms with respect to N do not contain g_n^a. Then, from the straightforward calculation, we obtain

$$|B_n| \leqslant \frac{1}{\sqrt{N}} \frac{\|\tilde{H}_I\|}{y} \frac{(6\beta y)^n}{n!} + o\left(\frac{1}{\sqrt{N}}\right), \tag{C15}$$

$$y := \max(\Omega, \|\tilde{H}_I\|). \tag{C16}$$

Summing $\beta^n B_n$ over n, we obtain

$$\left|\sum_{n=2} \beta^n B_n\right| \leqslant \frac{\|\tilde{H}_I\|}{y}\left(e^{6\beta y} - 6\beta y - 1\right)\frac{c}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right). \tag{C17}$$

From the foregoing arguments, $|\delta G_I^{(1)}| = o(1)$ is shown. In the same manner, $|\delta G_I^{(2)}| = o(1)$ is also shown. Thus, Eq. (36) is shown under the rotating wave approximation (C2) and the off-diagonal ETH (C3). Furthermore, Eq. (55) is shown since $\Omega = \Theta(1)$.

APPENDIX D: NAIVE APPROACH TO EVALUATE THE ERROR OF THE FLUCTUATION THEOREM

In this Appendix we discuss another naive approach to evaluate the error of the fluctuation without Eq. (29). In this approach, however, we cannot show that the error of the fluctuation theorem vanishes in the thermodynamic limit.

We note that $G - G^{\text{can}}$ can be written as the difference between the expectation values of $O_B(t)$, which is nonlocal

for $t > 0$, as follows:

$$G - G^{\text{can}} = \text{tr}_{\text{B}}\big[O_{\text{B}}(t)\big(\rho_{\text{B}} - \rho_{\text{B}}^{\text{MC}}\big)\big]$$
$$+ \text{tr}_{\text{B}}\big[O_{\text{B}}(t)\big(\rho_{\text{B}}^{\text{MC}} - \rho_{\text{B}}^{\text{can}}\big)\big] \qquad \text{(D1)}$$

$$O_{\text{B}}(t) := Z_{\text{B}} \text{tr}_{\text{S}}\big[U^{\dagger}\rho_{\text{S}}(t) \otimes \rho_{\text{B}}^{\text{can}} U\big] e^{\beta H_{\text{B}}}. \qquad \text{(D2)}$$

The first and second terms on the right-hand side of Eq. (D1) are regarded as the errors associated with the ETH and the equivalence of ensembles for $O_{\text{B}}(t)$, respectively. Below we discuss them in detail.

The equivalence of ensembles has been theoretically shown for (quasi)local operators, whose support is at most $N^{1/2}$ and operator norm does not increase with respect to N [77]. The error of the equivalence of ensembles only polynomially decreases with respect to N. If we assume that the equivalence of ensembles in the same form as [77] holds for $O_{\text{B}}(t)$, we cannot show the error of the equivalence of ensembles vanishes in the thermodynamic limit due to the exponential increase of the norm of $O_{\text{B}}(t)$.

The ETH for highly nonlocal operators has been investigated theoretically [85] and numerically [86]. In Ref. [85] the ETH is theoretically shown to hold with the error decreasing as $1/\sqrt{D}$ for typical many-body operators, whose operator norm does not grow exponentially. In Ref. [86] the same scaling as [85] was numerically observed. However, the operator norm of $O_{\text{B}}(t)$ exponentially increases and is bounded by $D\sqrt{N}$ [77] [see also Eq. (F37)]. Then the error associated with the ETH is bounded by $\sqrt{DN}$, and we cannot show that the error of the fluctuation theorem associated with the ETH vanishes in the thermodynamic limit.

For the above reasons, instead of the direct evaluation of $G - G_{\text{can}}$, in the main text we have adopted the decomposition (29) and show the fluctuation theorem based on plausible assumptions such as the ETH and the equivalence of ensembles for (quasi)local physical quantities.

APPENDIX E: INITIAL RISE OF THE ERROR OF THE FLUCTUATION THEOREM

In this Appendix we show that the errors of the fluctuation theorem initially rises in proportionate to t^2 as mentioned in Secs. III and V.

The error of the fluctuation theorem $\langle e^{-\sigma} \rangle - 1$ initially rises in proportionate to t^2, which is shown as

$$\langle e^{-\sigma} \rangle - 1 = \text{tr}[e^{-\beta H_{\text{B}}} U e^{\beta H_{\text{B}}} \rho_{\text{B}}(0) U^{\dagger} \rho_{\text{S}}(t)] - 1$$
$$= \text{tr}\Big[e^{-\beta H_{\text{B}}}\Big(-\frac{itH}{\hbar}\Big)e^{\beta H_{\text{B}}} \rho_{\text{S}}(t) \otimes \rho_{\text{B}}(0)\Big]$$
$$+ \text{tr}\Big[\rho_{\text{B}}(0)\Big(\frac{itH}{\hbar}\Big)\rho_{\text{S}}(t)\Big] + O(t^2)$$
$$= \text{tr}\Big[\frac{-itH}{\hbar}\big(e^{\beta H_{\text{B}}}\rho_t e^{-\beta H_{\text{B}}} - \rho_t\big)\Big] + O(t^2)$$
$$= O(t^2), \qquad \text{(E1)}$$

where we defined $\rho_t := \rho_{\text{S}}(t) \otimes \rho_{\text{B}}(0)$. In the second line, we expand the time evolution operator with respect to t. In the last line, we used $[H_{\text{B}}, \rho(0)] = 0$. We note that Eq. (17) is not assumed here.

Similarly, by using $[H_{\text{S}}, \rho(0)] = 0$ in addition to $[H_{\text{B}}, \rho(0)] = 0$, we show that δG_{S} and δG_{I} initially rise in t^2. In fact,

$$\delta G_{\text{S}} = \text{tr}[e^{\beta H_{\text{S}}} U e^{-\beta H_{\text{S}}} \rho_{\text{B}}(0) U^{\dagger} \rho_{\text{S}}(t)]$$
$$- \text{tr}[e^{\beta H_{\text{S}}} U e^{-\beta H_{\text{S}}} \rho_{\text{B}}^{\text{can}} U^{\dagger} \rho_{\text{S}}^{\text{can}}(t)]$$
$$= \text{tr}\Big[\Big(-\frac{itH}{\hbar}\Big)(e^{-\beta H_{\text{S}}}\rho_0 e^{\beta H_{\text{S}}} - \rho_0)\Big]$$
$$- \text{tr}\Big[\Big(-\frac{itH}{\hbar}\Big)(e^{-\beta H_{\text{S}}}\rho_0^{\text{can}} e^{\beta H_{\text{S}}} - \rho_0^{\text{can}})\Big] + O(t^2)$$
$$= O(t^2), \qquad \text{(E2)}$$

where $\rho_t^{\text{can}} := \rho_{\text{S}}(t) \otimes \rho_{\text{B}}^{\text{can}}$. Next, the interaction-induced error δG_{I} is decomposed into $\delta G_{\text{I}}^{(1)}$ and $\delta G_{\text{I}}^{(2)}$ as in Eq. (33). Then

$$\delta G_{\text{I}}^{(1)} = \text{tr}[e^{-\beta H_{\text{B}}} U e^{\beta H_{\text{B}}} \rho_{\text{B}}(0) U^{\dagger} \rho_{\text{S}}(t)]$$
$$- \text{tr}[e^{\beta H_{\text{S}}} U e^{-\beta H_{\text{S}}} \rho_{\text{B}}(0) U^{\dagger} \rho_{\text{S}}(t)]$$
$$= \text{tr}\Big[\Big(-\frac{itH}{\hbar}\Big)(e^{\beta H_{\text{B}}}\rho_t e^{-\beta H_{\text{B}}} - \rho_t)\Big]$$
$$- \text{tr}\Big[\Big(-\frac{itH}{\hbar}\Big)(e^{-\beta H_{\text{S}}}\rho_t e^{\beta H_{\text{S}}} - \rho_t)\Big] + O(t^2)$$
$$= O(t^2), \qquad \text{(E3)}$$

where we used the fact that ρ_0 and ρ_0^{can} commute with both of H_{S} and H_{B}. In the same manner as in Eq. (E3), it is shown that the initial rise of $\delta G_{\text{I}}^{(2)}$ defined in Eq. (35) is also proportionate to t^2.

Furthermore, we can show that δG_{ETH} defined in Sec. V initially rises as t^2 by replacing U in Eq. (E2) by U_{T}. Thus, Eqs. (68) and (69) are confirmed.

APPENDIX F: PROOF IN THE LONG-TIME REGIME

This Appendix shows the details of the proof of the fluctuation theorem in the long-time regime, which was discussed in Sec. IV. We show $|\overline{G_{\text{S1}}} - \overline{G_{\text{S1}}^{\text{can}}}| = o(1)$ [Eq. (43)] in Appendix F 1 and $|\overline{G_{\text{S2}}}| = o(1)$ and $|\overline{G_{\text{S2}}^{\text{can}}}| = o(1)$ [Eqs. (44) and (45)] in Appendix F 2. In Appendix F 3 we show that $|\overline{\lambda(t)} - \overline{\lambda^{\text{can}}(t)}| = o(1)$ [Eq. (A7)] holds, which complements the discussion about absolute irreversibility in Appendix A. Finally, we show that the temporal fluctuation of the error of the fluctuation theorem in the long-time regime vanishes in the thermodynamic limit in Appendix F 4.

1. Proof of Eq. (43) in Sec. IV

In this subsection, we show that $|\overline{G_{\text{S1}}} - \overline{G_{\text{S1}}^{\text{can}}}| = o(1)$ [Eq. (43)] holds. The assumption used here is the diagonal ETH of H for the observable of system S [Eq. (14)]. We also assume Eqs. (F4) and (F5) below, which state that the contribution from the outside of the energy shell is negligible in the thermodynamic limit.

We first define the energy shell of H as $[E' - \Delta, E' + \Delta]$, where $E' = \text{tr}[H\rho^{\text{DE}}]$ and $\Delta = \Theta(N^a)$ ($1/2 < a < 1$). We denote by $\sum_{a \in \text{sh}}$ the sum over a such that $E' - \Delta \leqslant E_a \leqslant E' + \Delta$. For the energy shell, we define $\rho^{\text{DE-sh}} := \sum_{a \in \text{sh}} \pi_a \rho^{\text{DE}} \pi_a$ and $\rho^{\text{can,DE-sh}} := \sum_{a \in \text{sh}} \pi_a \rho^{\text{can,DE}} \pi_a$, where

$\rho^{\text{DE-sh}}$ and $\rho^{\text{can,DE-sh}}$ are not normalized. Also, we define the microcanonical ensemble $\rho_{\text{MC}} := \sum_{a \in \text{sh}} \pi_a / D'$, where D' is the dimension of the energy shell.

From Eqs. (49) and (50), we evaluate $|\overline{G_{\text{S1}}} - \overline{G_{\text{S1}}^{\text{can}}}|$ by using the triangle inequality as follows:

$$\left| \overline{G_{\text{S1}}} - \overline{G_{\text{S1}}^{\text{can}}} \right| \tag{F1}$$

$$= \left| \text{tr}[O_{\text{S}} \rho^{\text{DE}}] - \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{can,DE}}] \right| \tag{F2}$$

$$\leqslant |\text{tr}[O_{\text{S}} \rho^{\text{DE}}] - \text{tr}[O_{\text{S}} \rho^{\text{DE-sh}}]|$$
$$+ |\text{tr}[O_{\text{S}} \rho^{\text{DE-sh}}] - \text{tr}[O_{\text{S}} \rho^{\text{MC}}]|$$
$$+ \left| \text{tr}[O_{\text{S}} \rho^{\text{MC}}] - \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{MC}}] \right|$$
$$+ \left| \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{MC}}] - \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{can,DE-sh}}] \right|$$
$$+ \left| \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{can,DE-sh}}] - \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{can,DE}}] \right|. \tag{F3}$$

The energy widths of $\rho_{\text{B}}(0)$ and $\rho_{\text{B}}^{\text{can}}$ are $\Theta(1)$ and $\Theta(N^{1/2})$ [77], respectively. Since they are narrower than Δ for sufficiently large N, the truncation error, which originates when ρ^{DE} and $\rho^{\text{can,DE}}$ are restricted to the energy shell, is negligible in the thermodynamic limit [50]:

$$\|\rho^{\text{DE}} - \rho^{\text{DE-sh}}\|_1 = o(1), \tag{F4}$$

$$\|\rho^{\text{can,DE}} - \rho^{\text{can,DE-sh}}\|_1 = o(1). \tag{F5}$$

From Eqs. (F4) and (F5), $|1 - \text{tr}[\rho^{\text{DE-sh}}]| = o(1)$ and $|1 - \text{tr}[\rho^{\text{can,DE-sh}}]| = o(1)$ hold. Then the first and fifth terms on the right-hand side of Eq. (F3) are evaluated as

$$|\text{tr}[O_{\text{S}} \rho^{\text{DE}}] - \text{tr}[O_{\text{S}} \rho^{\text{DE-sh}}]| \leqslant \|O_{\text{S}}\| \|\rho^{\text{DE}} - \rho^{\text{DE-sh}}\|_1 = o(1), \tag{F6}$$

$$\left| \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{can,DE-sh}}] - \text{tr}[O_{\text{S}}^{\text{can}} \rho^{\text{can,DE}}] \right|$$
$$\leqslant \|O_{\text{S}}^{\text{can}}\| \|\rho^{\text{can,DE-sh}} - \rho^{\text{can,DE}}\|_1 = o(1). \tag{F7}$$

We next evaluate the second term on the right-hand side of Eq. (F3) by using the ETH of H for O_{S}. We write the spectral decomposition of $\rho^{\text{DE-sh}}$ as $\rho^{\text{DE-sh}} = \sum_{a \in \text{sh}} p_a^{\text{DE}} \pi_a$. Then we can evaluate the second term on the right-hand side of Eq. (F3) as follows:

$$|\text{tr}[O_{\text{S}}(\rho^{\text{DE-sh}} - \rho^{\text{MC}})]| = \left| \sum_{a \in \text{sh}} p_a^{\text{DE}}(O_{\text{S}})_{aa} - \langle O_{\text{S}} \rangle_{\text{MC}} \right| \tag{F8}$$

$$\leqslant \max_{a \in \text{sh}} |(O_{\text{S}})_{aa} - \langle O_{\text{S}} \rangle_{\text{MC}}| \sum_{a \in \text{sh}} p_a^{\text{DE}}$$

$$+ \left(1 - \sum_{a \in \text{sh}} p_a^{\text{DE}} \right) |\langle O_{\text{S}} \rangle_{\text{MC}}| \tag{F9}$$

$$= o(1). \tag{F10}$$

In the last line, we used the ETH of H for O_{S} and Eq. (F4). In the same manner, we show that the fourth term on the right-hand side of Eq. (F3) is $o(1)$ by using the ETH for $O_{\text{S}}^{\text{can}}$.

The third term on the right-hand side of Eq. (F3) can be bounded from above by $\|\rho_{\text{MC}}\| \|O_{\text{S}} - O_{\text{S}}^{\text{can}}\|_1$. Besides, we can evaluate $\|O_{\text{S}} - O_{\text{S}}^{\text{can}}\|_1$ as $\|O_{\text{S}} - O_{\text{S}}^{\text{can}}\|_1 \leqslant Z_{\text{S}} \|e^{\beta H_{\text{S}}}\| \|\text{tr}_{\text{B}}[\rho^{\text{DE}} - \rho^{\text{can,DE}}]\|_1$. Then, using the triangle

inequality, we obtain

$$\|\text{tr}_{\text{B}}[\rho^{\text{DE}} - \rho^{\text{can,DE}}]\|_1 \leqslant \|\text{tr}_{\text{B}}[\rho^{\text{DE}} - \rho^{\text{DE-sh}}]\|_1$$
$$+ \|\text{tr}_{\text{B}}[\rho^{\text{DE-sh}} - \rho^{\text{MC}}]\|_1$$
$$+ \|\text{tr}_{\text{B}}[\rho^{\text{MC}} - \rho^{\text{can,DE-sh}}]\|_1$$
$$+ \|\text{tr}_{\text{B}}[\rho^{\text{can,DE-sh}} - \rho^{\text{can,DE}}]\|_1. \tag{F11}$$

Using Eqs. (F4) and (F5), we can show that the first and fourth terms on the right-hand side are both $o(1)$. From the ETH for any operator of system S, we show that the second and third terms on the right-hand side are also $o(1)$.

By summing up the foregoing arguments, $|\overline{G_{\text{S1}}} - \overline{G_{\text{S1}}^{\text{can}}}| = o(1)$ is proved.

2. Proof of Eqs. (44) and (45) in Sec. IV

This Appendix shows that Eqs. (44) and (45) hold by using the off-diagonal ETH for all operators of system S (53) and the assumption on the initial energy distribution (F4) and (F5).

From Eqs. (F4) and (F5), we can neglect the contribution from the outside of the energy shell and obtain Eq. (52) from Eq. (41). Then, along with the off-diagonal ETH (53), we show that

$$|\overline{G_{\text{S2}}}| \leqslant \frac{\Theta(1)}{D'} \left| \sum_{\substack{a,b \in \text{sh} \\ a \neq b \\ i,j}} e^{\beta E_i^{\text{S}}} (\rho_1)_{ab} (\rho_2)_{ba} \right| \tag{F12}$$

holds, where $\rho_1 := \rho_{\text{S}}^{\text{can}} \otimes \rho_{\text{B}}(0)$ and $\rho_2 := \rho_{\text{S}}(0) \otimes \rho_{\text{B}}(0)$. Using the Cauchy-Schwartz inequality, we obtain

$$\left| \sum_{\substack{a,b \in \text{sh} \\ a \neq b}} (\rho_1)_{ab} (\rho_2)_{ba} \right| \leqslant \sqrt{\text{tr}[\rho_1^2] \text{tr}[\rho_2^2]} \leqslant 1.$$

From the above,

$$|\overline{G_{\text{S2}}}| \leqslant \frac{\Theta(1)}{D'} \sum_{i,j} e^{\beta E_i^{\text{S}}} = o(1) \tag{F13}$$

is proved. In the same manner, $|\overline{G_{\text{S2}}^{\text{can}}}| \leqslant o(1)$ is proved.

3. Proof of Eq. (A7) in Appendix A

In this Appendix we show Eq. (A7) in Appendix A. As in Sec. IV, the long-time averages of $\lambda(t)$ and $\lambda^{\text{can}}(t)$ are written as the sums of the diagonal term and the off-diagonal term: $\overline{\lambda(t)} = \overline{\lambda_1} + \overline{\lambda_2}$ and $\overline{\lambda^{\text{can}}(t)} = \overline{\lambda_1^{\text{can}}} + \overline{\lambda_2^{\text{can}}}$.

First, we evaluate the diagonal term. We define $Q^{\text{S}} := 1 - P_{\text{ini}}^{\text{S}}$. Then $\overline{\lambda_1}$ is written as

$$\overline{\lambda_1} = \sum_a \text{tr}[Q^{\text{S}} |E_a\rangle\langle E_a| \rho_{\text{S}}^{\text{DE}} \otimes \rho_{\text{B}}(0) |E_a\rangle\langle E_a|] \tag{F14}$$

$$= \langle Q^{\text{S}} \rangle_{\text{MC}} + \delta\lambda_1, \tag{F15}$$

$$\delta\lambda_1 := \sum_a ((Q^{\text{S}})_{aa} - \langle Q^{\text{S}} \rangle_{\text{MC}}) (\rho_{\text{S}}^{\text{DE}} \otimes \rho_{\text{B}}(0))_{aa}, \tag{F16}$$

where $\langle Q^{\text{S}} \rangle_{\text{MC}}$ only depends on the energy shell and does not depend on the microscopic details of the initial state. On the

other hand, $\delta\lambda_1$ can depend on the details. By neglecting the outside of the energy shell, we obtain

$$|\delta\lambda_1| \simeq \left| \sum_{a\in\text{sh}} ((Q^S)_{aa} - \langle Q^S\rangle_{\text{MC}})(\rho_S^{\text{DE}} \otimes \rho_B(0))_{aa} \right| \quad \text{(F17)}$$

$$\leqslant \max_{a\in\text{sh}} |(Q^S)_{aa} - \langle Q^S\rangle_{\text{MC}}| \sum_b (\rho_S^{\text{DE}} \otimes \rho_B(0))_{bb} \quad \text{(F18)}$$

$$= o(1), \quad \text{(F19)}$$

where we used the ETH for Q^S. We can show $|\overline{\lambda_1^{\text{can}}} - \langle Q^S\rangle_{\text{MC}}| = o(1)$ in the same manner. Thus, $|\overline{\lambda_1} - \overline{\lambda_1^{\text{can}}}| = o(1)$ is proved.

The off-diagonal term $\overline{\lambda_2}$ can be evaluated as follows:

$$\overline{\lambda_2} \simeq \sum_{\substack{a,b\in\text{sh}\\a\neq b\\i,j}} (Q^S)_{ba}(q_S^{ij})_{ab}(q_S^{ji})_{ba}(\rho(0))_{ab} \quad \text{(F20)}$$

$$= \sum_{\substack{a,b\in\text{sh}\\a\neq b}} \frac{\Theta(1)}{D'}(q_S)_{ba}(\rho(0))_{ab}, \quad \text{(F21)}$$

where we used the off-diagonal ETH for Q^S and q_S^{ij} and defined $q_S := \sum_{ji} q_S^{ji}$, which is a Hermitian operator of system S. Using the Cauchy-Schwartz inequality, we obtain

$$|\overline{\lambda_2}| \leqslant \frac{\Theta(1)}{D'} \sqrt{\left| \sum_{\substack{a,b\in\text{sh}\\a\neq b}} (q_S)_{ba}(q_S)_{ab} \right| \left| \sum_{\substack{a,b\in\text{sh}\\a\neq b}} (\rho(0))_{ab}(\rho(0))_{ba} \right|}. \quad \text{(F22)}$$

The right-hand side of this inequality is evaluated as

$$\left| \sum_{\substack{a,b\in\text{sh}\\a\neq b}} (q_S)_{ba}(q_S)_{ab} \right| \leqslant D'\left|\text{tr}_S[(q_S)^2]\right| = \Theta(D),$$

$$\sum_{\substack{a,b\in\text{sh}\\a\neq b}} (\rho(0))_{ab}(\rho(0))_{ba} \leqslant \text{tr}[(\rho(0))^2] \leqslant 1. \quad \text{(F23)}$$

Then

$$|\overline{\lambda_2}| \leqslant \frac{\Theta(1)}{\sqrt{D'}} = o(1) \quad \text{(F24)}$$

is proved. In the same manner, $|\overline{\lambda_2^{\text{can}}}| = o(1)$ is proved.

From the foregoing argument, $|\overline{\lambda(t)} - \overline{\lambda^{\text{can}}(t)}| = o(1)$ is proved. We note that if the ETH holds, the correction of absolute irreversibility in the long-time regime is given by $\langle Q^S\rangle_{\text{MC}} = 1 - \langle P_{\text{ini}}^S\rangle_{\text{MC}}$ regardless of the initial state of bath B.

4. Temporal fluctuation

In this subsection, we show that the temporal fluctuation around the long-time average of the error of the fluctuation theorem vanishes in the thermodynamic limit. Together with the fact that the long-time average of the error is $o(1)$, we show that the fluctuation theorem holds for almost all the times
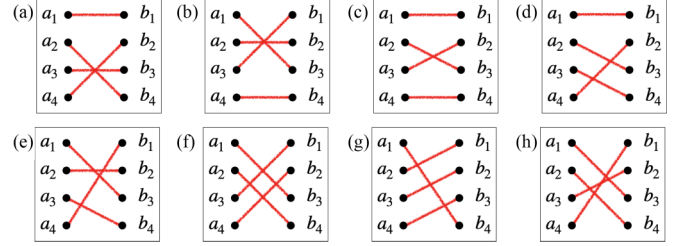

FIG. 12. The terms that appear in the calculation of the temporal fluctuation. The indexes connected by the lines are equal. For example, (a) shows $a_1 = b_1$, $a_2 = b_4$, $a_3 = b_3$, and $a_4 = b_2$.

after the relaxation time, except for the effect of quantum recurrence [72–74].

We define the temporal fluctuation around the long-time average $\overline{A}$ as

$$\Delta_T(A) := \overline{(A - \overline{A})^2} \quad \text{(F25)}$$

and the cross correlation as

$$\Delta_T(A, B) := \overline{AB} - \overline{A}\,\overline{B}. \quad \text{(F26)}$$

The temporal fluctuations of $\delta G := G - 1$, δG_S and δG_I are written as

$$\Delta_T(\delta G) = \Delta_T(G), \quad \text{(F27)}$$

$$\Delta_T(\delta G_S) = \Delta_T(G_S) + \Delta_T(G_S^{\text{can}}) - 2\Delta_T(G_S, G_S^{\text{can}}), \quad \text{(F28)}$$

$$\Delta_T(\delta G_I) = \Delta_T(\delta G_S) + \Delta_T(G)$$
$$- 2\Delta_T(G, G_S) + 2\Delta_T(G, G_S^{\text{can}}), \quad \text{(F29)}$$

respectively. From the Cauchy-Schwartz inequality, $|\Delta_T(A, B)|^2 \leqslant \Delta_T(A)\Delta_T(B)$ holds. Then, if all of the equations $\Delta_T(G_S) = o(1)$, $\Delta_T(G_S^{\text{can}}) = o(1)$ and $\Delta_T(G) = o(1)$ hold, the temporal fluctuations (F27), (F28), and (F29) vanish in the thermodynamic limit. Below we will show $\Delta_T(G_S) = o(1)$, $\Delta_T(G_S^{\text{can}}) = o(1)$ and $\Delta_T(G) = o(1)$.

First, to show $\Delta_T(G_S) = o(1)$, we evaluate $\overline{G_S^2}$. In the same manner as in Sec. IV A, we obtain

$$\overline{G_S^2} = \sum_{\substack{(a_1,a_2,a_3,a_4)=(b_1,b_2,b_3,b_4)\\i,j,k,l}} \rho_{a_1,b_1}\rho'_{a_2,b_2}\rho_{a_3,b_3}\rho'_{a_4,b_4}$$

$$\times (A_S^{ij})_{b_1,a_1}(B_S^{ij})_{b_2,a_2}(A_S^{kl})_{b_3,a_3}(B_S^{kl})_{b_4,a_4}, \quad \text{(F30)}$$

where $\rho := \rho_S^{\text{can}} \otimes \rho_B(0)$, $\rho' := \rho_S(0) \otimes \rho_B(0)$, $A_S^{ij} := Z_S q_S^{ij} e^{\beta H_S}$, and $B_S^{ij} := Z_S q_S^{ji}$. We note that the summation is taken over the case of $(a_1, a_2, a_3, a_4) = (b_1, b_2, b_3, b_4)$, i.e., a permutation of (a_1, a_2, a_3, a_4) equals (b_1, b_2, b_3, b_4). In the calculation of $(\overline{G_S})^2$, only the combination of the form $(a_1, a_2) = (b_1, b_2)$ and $(a_3, a_4) = (b_3, b_4)$ appears in the sum. Then this combination does not appear in the temporal fluctuation $\overline{G_S^2} - (\overline{G_S})^2$. In Fig. 12 we show that combinations of (a_1, a_2, a_3, a_4) and (b_1, b_2, b_3, b_4) that should be evaluated. The combinations that do not appear in Fig. 12 are equivalent to some of the combinations in Fig. 12. In the following evaluation, the diagonal matrix elements for the observable of

system S are evaluated using the diagonal ETH as

$$\left(A_{\mathrm{S}}^{ij}\right)_{a,a} \simeq \left\langle A_{\mathrm{S}}^{ij}\right\rangle_{\mathrm{MC}} = \Theta(1). \tag{F31}$$

In $\Delta_T(G_{\mathrm{S}})$, the term corresponding to Fig. 12(a) is evaluated as

$$\sum_{\substack{a_1,a_2,a_3,a_4 \\ i,j,k,l}} \rho_{a_1,a_1} \rho'_{a_2,a_4} \rho_{a_3,a_3} \rho'_{a_4,a_2}$$

$$\times \left(A_{\mathrm{S}}^{ij}\right)_{a_1,a_1} \left(B_{\mathrm{S}}^{ij}\right)_{a_4,a_2} \left(A_{\mathrm{S}}^{kl}\right)_{a_3,a_3} \left(B_{\mathrm{S}}^{kl}\right)_{a_2,a_4} \tag{F32}$$

$$= \frac{\Theta(1)}{D'} \sum_{\substack{a_1,a_2,a_3,a_4 \\ i,j,k,l}} \rho_{a_1,a_1} \rho'_{a_2,a_4} \rho_{a_3,a_3} \rho'_{a_4,a_2} \tag{F33}$$

$$= \frac{\Theta(1)}{D'} (\mathrm{tr}[\rho])^2 \mathrm{tr}[(\rho')^2] \tag{F34}$$

$$= o(1). \tag{F35}$$

We used the off-diagonal ETH for B_{S}^{ij} and B_{S}^{kl} and the diagonal ETH for A_{S}^{ij} and A_{S}^{kl}. Because we can evaluate the other terms in the similar manner, $\Delta_T(G_{\mathrm{S}}) = o(1)$ is proved.

In the evaluation of $\Delta_T(G_{\mathrm{S}})$, only $\mathrm{tr}[\rho] = 1$ and $\mathrm{tr}[\rho^2] \leqslant 1$ are used as the properties of the density operator. Then $\Delta_T(G_{\mathrm{S}}^{\mathrm{can}}) = o(1)$ is proved by setting $\rho = \rho_{\mathrm{S}}^{\mathrm{can}} \otimes \rho_{\mathrm{B}}^{\mathrm{can}}$ and $\rho' = \rho_{\mathrm{S}}(0) \otimes \rho_{\mathrm{B}}^{\mathrm{can}}$.

We next evaluate $\Delta_T(G)$. As in Eq. (F30), we can write $\overline{G^2}$ as

$$\overline{G^2} = \left(D_{\mathrm{S}} Z_{\mathrm{B}} e^{\beta E_{\mathrm{ini}}^{\mathrm{B}}}\right)^2$$

$$\times \sum_{\substack{(a_1,a_2,a_3,a_4)=(b_1,b_2,b_3,b_4) \\ i,j,k,l}} \tilde{\rho}_{a_1,b_1} \rho'_{a_2,b_2} \tilde{\rho}_{a_3,b_3} \rho'_{a_4,b_4}$$

$$\times (\rho^{ij})_{b_1,a_1} \left(B_{\mathrm{S}}^{ij}\right)_{b_2,a_2} (\rho^{kl})_{b_3,a_3} \left(B_{\mathrm{S}}^{kl}\right)_{b_4,a_4}, \tag{F36}$$

where $E_{\mathrm{ini}}^{\mathrm{B}}$ is the energy of the initial state of bath B, $\tilde{\rho} := (1_{\mathrm{S}}/D_{\mathrm{S}}) \otimes \rho_{\mathrm{B}}(0)$ and $\rho^{ij} := q_{\mathrm{S}}^{ij} \otimes \rho_{\mathrm{B}}^{\mathrm{can}}$.

Regarding the partition function, the following relation holds [77]:

$$\left(Z_{\mathrm{B}} e^{\beta E_{\mathrm{ini}}^{\mathrm{B}}}\right)^2 \leqslant D^2 N. \tag{F37}$$

We also have

$$\mathrm{tr}[(\rho^{ij})^2] = \delta_{ij}/D_{\mathrm{eff}}\left[\rho_{\mathrm{B}}^{\mathrm{can}}\right], \tag{F38}$$

where $D_{\mathrm{eff}}[\rho_{\mathrm{B}}^{\mathrm{can}}]$ is the effective dimension of $\rho_{\mathrm{B}}^{\mathrm{can}}$ with respect to H_{B}. For simplicity, when there is no degeneracy in the eigenenergy and the state is diagonal with the energy eigenstates as $\rho = \sum_i p_i |E_i\rangle\langle E_i|$, the effective dimension is written as

$$D_{\mathrm{eff}}[\rho] := \left(\sum_i |p_i|^2\right)^{-1}. \tag{F39}$$

We note that $D_{\mathrm{eff}}[\rho_{\mathrm{B}}^{\mathrm{can}}]$ increases exponentially with respect to the bath size N [50]. Also, we assume that the matrix elements of the density operator ρ' and $\tilde{\rho}$ are $O(D'^{-1})$:

$$|(\rho')_{ab}| = O(D'^{-1}), \tag{F40}$$

$$|(\tilde{\rho})_{ab}| = O(D'^{-1}). \tag{F41}$$

This is not satisfied when the density operators ρ' and $\tilde{\rho}$, which are product states as in Eq. (3), are localized to some specific eigenstates of H. When the initial state is in the form of Eq. (3), we argue that $O(D'^{-1})$ is naturally satisfied if the Hamiltonian H is chaotic and mixes system S and bath B sufficiently.

In $\Delta_T(G_{\mathrm{S}})$, the term corresponding to Fig. 12(b) is evaluated as

$$\left| \left(D_{\mathrm{S}} Z_{\mathrm{B}} e^{\beta E_{\mathrm{ini}}^{\mathrm{B}}}\right)^2 \sum_{\substack{a_1,a_2,a_3,a_4 \\ i,j,k,l}} \tilde{\rho}_{a_1,a_3} \rho'_{a_2,a_2} \tilde{\rho}_{a_3,a_1} \rho'_{a_4,a_4} \right.$$

$$\left. \times (\rho^{ij})_{a_3,a_1} \left(B_{\mathrm{S}}^{ij}\right)_{a_2,a_2} (\rho^{kl})_{a_1,a_3} \left(B_{\mathrm{S}}^{kl}\right)_{a_4,a_4} \right| \tag{F42}$$

$$\leqslant \Theta(N) \left| \sum_{\substack{a_1,a_2,a_3,a_4 \\ i,j,k,l}} \rho'_{a_2,a_2} \rho'_{a_4,a_4} (\rho^{ij})_{a_3,a_1} (\rho^{kl})_{a_1,a_3} \right| \tag{F43}$$

$$\leqslant \Theta(N) (\mathrm{tr}[\rho'])^2 \sum_{i,j,k,l} \sqrt{\mathrm{tr}[(\rho^{ij})^2] \mathrm{tr}[(\rho^{kl})^2]} \tag{F44}$$

$$= \frac{\Theta(N)}{D_{\mathrm{eff}}[\rho_{\mathrm{B}}^{\mathrm{can}}]} = o(1). \tag{F45}$$

In the first inequality, we used Eqs. (F31), (F37), and (F40). In the second inequality, we used the Cauchy-Schwartz inequality. In the last inequality, we used Eq. (F38). Because we can evaluate the other terms in the same manner, $\Delta_T(G) = o(1)$ is proved.

APPENDIX G: PERTURBATION THEORY ON THE INTERACTION-INDUCED ERROR

In this Appendix, by using the perturbation theory, we show that the interaction-induced error does not grow significantly in the long-time regime and vanishes in the thermodynamic limit in the short-time regime.

1. The long-time regime

We show that $|\overline{\delta G_{\mathrm{I}}}| \leqslant \Theta(1)$ holds at high temperature using the perturbation theory. The long-time average of the interaction-induced error (32) is written as

$$\overline{\delta G_{\mathrm{I}}^{(1)}} := \overline{G} - \overline{G_{\mathrm{S}}}, \tag{G1}$$

$$\overline{\delta G_{\mathrm{I}}^{(2)}} := \overline{G_{\mathrm{S}}^{\mathrm{can}}} - \overline{G^{\mathrm{can}}}. \tag{G2}$$

First, we divide $\overline{\delta G_{\mathrm{I}}^{(1)}}$ into the diagonal term $\overline{\delta G_{\mathrm{I1}}^{(1)}}$ and the off-diagonal term $\overline{\delta G_{\mathrm{I2}}^{(1)}}$. The diagonal term is written as

$$\overline{\delta G_{\mathrm{I1}}^{(1)}} = \sum_a \mathrm{tr}\left[e^{-\beta H_{\mathrm{B}}} \pi_a e^{\beta H_{\mathrm{B}}} \rho_{\mathrm{B}}(0) \pi_a \rho_{\mathrm{S}}^{\mathrm{DE}}\right]$$

$$- \sum_a \mathrm{tr}\left[e^{\beta H_{\mathrm{S}}} \pi_a e^{-\beta H_{\mathrm{S}}} \rho_{\mathrm{B}}(0) \pi_a \rho_{\mathrm{S}}^{\mathrm{DE}}\right], \tag{G3}$$

where $\rho_{\mathrm{S}}^{\mathrm{DE}} := \mathrm{tr}_{\mathrm{B}}[\rho^{\mathrm{DE}}]$. By using the first-order perturbation theory for the eigenstates, we obtain

$$\overline{\delta G_{\mathrm{I1}}^{(1)}} \simeq \sum_{i\alpha} \sum_{j\beta}{}' \sum_{k\gamma}{}' \frac{\left|\langle E_j^{\mathrm{S}} E_\beta^{\mathrm{B}}|H_{\mathrm{I}}|E_i^{\mathrm{S}} E_\alpha^{\mathrm{B}}\rangle\right|^2}{\left(\delta E_{ji}^{\mathrm{S}} + \delta E_{\beta\alpha}^{\mathrm{B}}\right)\left(\delta E_{ki}^{\mathrm{S}} + \delta E_{\beta\alpha}^{\mathrm{B}}\right)} \delta_{\beta\gamma}$$

$$\times \left[\langle E_j^S | \rho_S^{DE} | E_k^S \rangle \delta_{\alpha,\text{ini}} \left(e^{-\beta \delta E_{\beta\alpha}^B} - e^{-\beta \delta E_{ki}^S} \right) \right.$$
$$\left. + \langle E_j^S | \rho_S^{DE} | E_k^S \rangle \delta_{\alpha,\text{ini}} \left(e^{-\beta \delta E_{\beta\alpha}^B} - e^{-\beta \delta E_{ki}^S} \right) \right], \quad (G4)$$

where the summation over j, k, β, γ are restricted to $j \neq i, k \neq i, \beta \neq \alpha, \gamma \neq \alpha$ and $\delta_{\alpha,\text{ini}}$ means α equals the index of the initial state of bath B. We also define $\delta E_{ij}^S := E_i^S - E_j^S$ and $\delta E_{\beta\alpha}^B := E_\beta^B - E_\alpha^B$.

Here we assume the following relation as the off-diagonal ETH instead of Eq. (53).

$$\left| \langle E_j^S E_\beta^B | H_I | E_i^S E_\alpha^B \rangle \right|^2 \sim \frac{\Theta(1)}{D} e^{-\nu |\delta E_{\beta\alpha}^B|}, \quad (G5)$$

where ν is an N-independent constant. The exponential decay with respect to energy difference $\delta E_{\beta\alpha}^B$ in Eq. (G5) has been observed in numerical calculations of the off-diagonal ETH [37]. Also, the exponential decay is theoretically shown when the Hamiltonian is local [87]. Then the terms that can diverge in Eq. (G4) are written as $e^{-(\nu\pm\beta)|\delta E_{\beta\alpha}^B|}$. At high temperature $\nu > \beta$, $\overline{\delta G_{I1}^{(1)}} \leqslant \Theta(1)$ is shown.

We next show that the off-diagonal term vanishes as N increases: $|\overline{\delta G_{I2}^{(1)}}| = o(1)$. Since $|\overline{G_{S2}}| = o(1)$ is already shown, it is sufficient to show $|\overline{G_2}| = o(1)$, where $\overline{G_2}$ is the off-diagonal contribution of $\overline{G}$. In the same manner as Appendix F, we show $|\overline{G_2}| = o(1)$ by using the off-diagonal ETH (53) and the conditions on the density operators (F40) and (F41).

We divide $\overline{\delta G_I^{(2)}}$ into the diagonal term $\overline{\delta G_{I1}^{(2)}}$ and the off-diagonal term $\overline{\delta G_{I1}^{(2)}}$. Since $\overline{G_{S2}^{can}} = o(1)$ is already shown, $|\overline{\delta G_{I2}^{(2)}}| = o(1)$ holds. By using the diagonal ETH for the operator of system S, the diagonal term $\overline{\delta G_{I1}^{(2)}}$ is evaluated as

$$\overline{\delta G_{I1}^{(2)}} \simeq 1 - \left\langle \rho_S^{DE} \left(\rho_S^{can} \right)^{-1} \right\rangle_{MC}. \quad (G6)$$

If the density operator of system S in the long-time regime ρ_S^{DE} relaxes to the canonical ensemble of H_S, the second term on the right-hand side equals 1. Then we argue that $\overline{\delta G_{I1}^{(2)}} = o(1)$ holds when ρ_S^{DE} relaxes to the canonical ensemble. On the other hand, ρ_S^{DE} does not necessarily relax to the canonical ensemble of H_S for general interactions. Then we assume $\overline{\delta G_{I1}^{(2)}} \leqslant \Theta(1)$ in general.

From the foregoing argument, $|\overline{\delta G_I}| \leqslant \Theta(1)$ is shown.

2. The short-time regime

We show that $|\delta G_I| = o(1)$ holds in the short-time regime ($t \ll \tau_{LR}$) at high temperature using the perturbation theory.

The interaction-induced error is defined in Eq. (32). Because $\rho_S(t) \simeq \rho_S^{can}(t)$ in $t \ll \tau_{LR}$ holds due to the Lieb-Robinson bound and the ETH, we obtain

$$\delta G_I \simeq \text{tr}[e^{-\beta H_B} U e^{\beta H_B} \delta \rho_B U^\dagger \rho_S(t)]$$
$$- \text{tr}[e^{\beta H_S} U e^{-\beta H_S} \delta \rho_B U^\dagger \rho_S(t)], \quad (G7)$$

where $\delta \rho_B := \rho_B(0) - \rho_B^{can}$. By using the Lieb-Robinson bound and the ETH again, the second line is shown to be $o(1)$. Below we evaluate the first line using the perturbation calculation for the time evolution operator.

From the first-order perturbation theory, we obtain

$$U \simeq U_0 + U_1, \quad (G8)$$
$$U_0 := e^{-iH_0 t/\hbar}, \quad (G9)$$
$$U_1 := -iU_0 \int_0^t d\tau e^{-iH_0\tau/\hbar} H_I e^{iH_0\tau/\hbar}, \quad (G10)$$

where we define $H_0 := H_S + H_B$. We also define

$$\delta G_I^B[U, V^\dagger] := \text{tr}[e^{-\beta H_B} U e^{\beta H_B} \delta \rho_B V^\dagger \rho_S(t)], \quad (G11)$$
$$\delta G_{I,(i,j)}^B := \delta G_I^B[U_i, U_j^\dagger]. \quad (G12)$$

Then the first line of Eq. (G7) is approximated as $\sum_{i=0}^1 \sum_{j=0}^1 \delta G_{I,(i,j)}^B$.

We first evaluate $\delta G_{I,(0,0)}^B$ as

$$\delta G_{I,(0,0)}^B = \text{tr}[e^{-\beta H_B} U_0 e^{\beta H_B} \delta \rho_B U_0^\dagger \rho_S(t)]$$
$$= \text{tr}[\rho_S(t) \otimes \delta \rho_B]$$
$$= 0, \quad (G13)$$

where we used $[U_0, H_B] = 0$ and $[U_0, \delta \rho_B] = 0$. We next calculate $\delta G_{I,(0,1)}^B$ as

$$\left| \delta G_{I,(0,1)}^B \right| = \left| \text{tr}[e^{-\beta H_B} U_0 e^{\beta H_B} \delta \rho_B U_1^\dagger \rho_S(t)] \right|$$
$$= \left| \int_0^t d\tau \text{tr}[U_0 H_I \rho_S(t, \tau) \otimes \delta \rho_B] \right|$$
$$\leqslant \int_0^t d\tau \|U_0\| \|H_I \rho_S(t, \tau) \otimes \delta \rho_B\|_1, \quad (G14)$$

where we define $\rho_S(t, \tau) := e^{iH_S(t-\tau)} \rho_S(t) e^{-iH_S(t-\tau)}$. We note that $\|U_0\| = 1$ holds, and $\|H_I \rho_S(t, \tau) \otimes \delta \rho_B\|_1 = O(D^{-1/2})$ is satisfied from the ETH. Then, in the short-time regime $t \ll \tau_{LR}$, we obtain

$$\left| \delta G_{I,(0,1)}^B \right| \leqslant \frac{\tau_{LR}}{\Theta(D^{1/2})} = \frac{\Theta(N^\mu)}{\Theta(D^{1/2})} = o(1). \quad (G15)$$

Also, $\delta G_{I,(1,0)}^B$ is evaluated as

$$\delta G_{I,(1,0)}^B = -i \int_0^t d\tau \text{tr}[U_0 \rho_S(t, \tau) \otimes \delta \rho H_I]. \quad (G16)$$

Then $\delta G_{I,(1,0)}^B = o(1)$ in $t \ll \tau_{LR}$ is shown to hold in the same manner as $\delta G_{I,(0,1)}^B$.

We finally calculate $\delta G_{I,(1,1)}^B$ as

$$\delta G_{I,(1,1)}^B = \int_0^t \int_0^t d\tau d\tau' \text{tr}[e^{-\beta H_B} e^{iH_0(t-\tau)} H_I$$
$$\times e^{-iH_0\tau} e^{\beta H_B} \delta \rho_B e^{iH_0\tau'} H_I e^{iH_0(t-\tau')} \rho_S(t)]$$
$$\leqslant t^2 \left| \sum_{a,b,c} e^{-\beta \delta E_B^{a_B b_B}} (H_I)_{ab} (H_I)_{bc} (\delta \rho_B)_{bb} (\rho_S(t))_{ca} \right|$$
$$= t^2 \left| \sum_{a,b,c} M(a, b, c, t) \right|, \quad (G17)$$

where

$$M(a, b, c, t) := e^{-\beta \delta E_B^{a_B b_B}} (H_I)_{ab} (H_I)_{bc} (\delta \rho_B)_{bb} (\rho_S(t))_{ca}. \quad (G18)$$

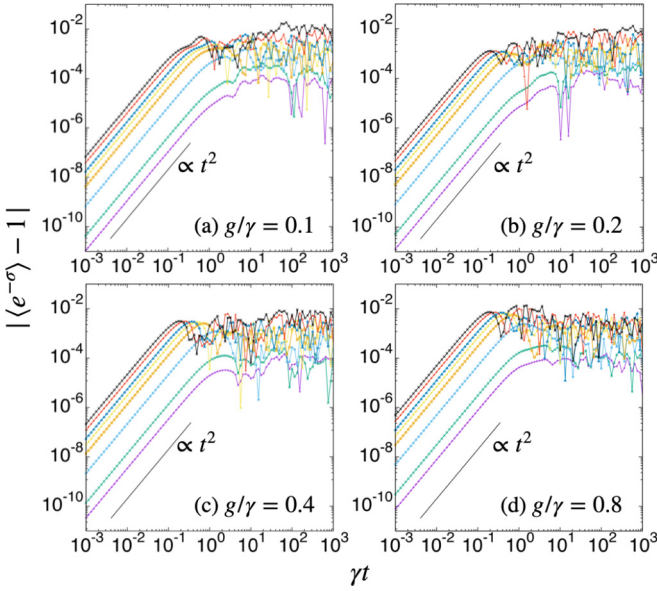FIG. 13. The time dependence of the error of the fluctuation theorem $|\langle e^{-\sigma}\rangle - 1|$, whose initial rise is proportionate to t^2. The initial state of bath B is the energy eigenstates of H_B, whose energy is maximum in the energy shell at $\beta = 0.1$. Parameters: $\gamma'/\gamma = 0.05, 0.1, 0.4, 1, 1.5, 2, 3, 4$ (from bottom to top) and $p = 0.99$. The onsite potential ω is determined by $\mathrm{tr}_B[n_B \rho_B^{can}] = N_P$.

We note that the index a represents a pair of indexes of the eigenstates of system S and bath B and is written as $a = (a_S, a_B)$. We write the corresponding eigenenergy as $E^a = E_S^{a_S} + E_B^{a_B}$. The same applies to b and c. We write the matrix elements as $(A)_{ab} = \langle E_S^{a_S} E_B^{a_B}|A|E_S^{b_S} E_B^{b_B}\rangle$, $(A)_{a_S b_S} = \langle E_S^{a_S}|A|E_S^{b_S}\rangle$, and $(A)_{a_B b_B} = \langle E_B^{a_B}|A|E_B^{b_B}\rangle$. We also write the energy change of B as $\delta E_B^{a_B b_B}$.

We define the energy ranges of bath B as follows. First, we define $\Lambda_1 := [E - \Delta, E + \Delta]$, $E = \mathrm{tr}_B[H_B \rho_B(0)] = \mathrm{tr}_B[H_B \rho_B^{can}]$, and $\Delta = \Theta(N^a)$ ($1/2 < a < 1$). We note that the energy widths of $\rho_B(0)$ and ρ_B^{can} are included in Λ_1. We also define $\Lambda_2 := [E - 2\Delta, E + 2\Delta]$ and denote the outside of Λ_2 by $\overline{\Lambda_2}$. We write the sum range as $a \in \Lambda_1$, which means that $\sum_{a \in \Lambda_1} = \sum_{a_S} \sum_{a_B \in \Lambda_1}$.

We evaluate $\sum_{a,b,c} M(a, b, c, t)$ for the following three cases: (1) $E_a \in \overline{\Lambda_2}$, (2) $E_a \in \Lambda_2, a_B \neq b_B$, and (3) $E_a \in \Lambda_2, a_B = b_B$. We note that $a_B = c_B$ always holds since $\rho_S(t)$ is the operator of system S.
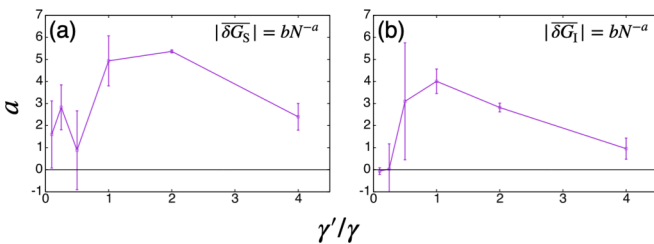

FIG. 14. The γ' dependence of the exponent a, which is obtained by fitting $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$. Parameters: $p = 0.9$, $g = 0.1\gamma$, $\beta = 0.1$, $\gamma' = \gamma$. The left figure shows $|\overline{\delta G_S}|$, and the right figure shows $|\overline{\delta G_I}|$.
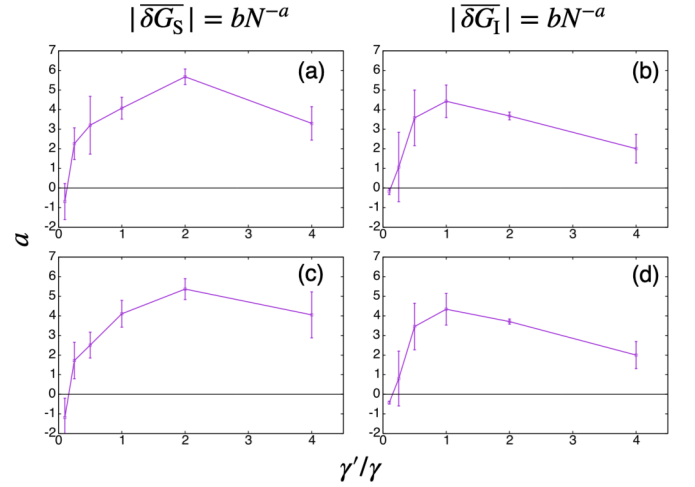

FIG. 15. The γ' dependence of the exponent a, which is obtained by fitting $|\overline{\delta G_S}|$ and $|\overline{\delta G_I}|$. Parameters: $g = 0.4\gamma$, $\beta = 0.1$, $\gamma' = \gamma$. $p = 0.99$ (upper panels) and $p = 0.9$ (lower panels). The left panels show $|\overline{\delta G_S}|$, and the right panels show $|\overline{\delta G_I}|$.

First, we consider case (1):

$$\sum_{\substack{a \in \overline{\Lambda_2} \\ b \in \Lambda_1 \\ c; c_B = a_B}} M(a, b, c, t) = \frac{\Theta(1)}{D} \sum_{\substack{a \in \overline{\Lambda_2} \\ b \in \Lambda_1 \\ c; c_B = a_B}} (\delta\rho_B)_{bb}(\rho_S(t))_{ca}$$
$$\times e^{-\nu|\delta E_B^{a_B b_B}|} e^{-\nu|\delta E_B^{b_B c_B}|} e^{-\beta \delta E_B^{a_B b_B}}, \tag{G19}$$

where we used the off-diagonal ETH for H_I. We write b_B maximizing $e^{-\nu|\delta E_B^{a_B b_B}|}$ and $e^{-\beta \delta E_B^{a_B b_B}}$ as b_B^ν and b_B^β, respectively. By using $e^{-\nu|\delta E_B^{b_B c_B}|} \leqslant 1$ and $|(\rho_S(t))_{ca}| \leqslant 1$, the absolute value of the right-hand side of Eq. (G19) is bounded from above as follows:

$$\frac{D_S}{D} \sum_{a \in \overline{\Lambda_2}} e^{-\nu|\delta E_B^{a_B b_B^\nu}|} e^{+\beta|\delta E_B^{a_B b_B^\beta}|} \sum_{b \in \Lambda} |\delta\rho_{bb}|$$
$$\leqslant \frac{D_S D_{\overline{\Lambda_2}}}{D} \max_{a_B} \left(e^{-\nu|\delta E_B^{a_B b_B^\nu}|} e^{+\beta|\delta E_B^{a_B b_B^\beta}|}\right), \tag{G20}$$

where we used $\sum_{b \in \Lambda} |\delta\rho_{bb}| \leqslant 1$ and $D_{\overline{\Lambda_2}}$ is the dimension of $\overline{\Lambda_2}$. The right-hand side exponentially decays as $e^{-(\nu-\beta)\Theta(\Delta)}$ if $\beta < \nu$. Then the right-hand side of Eq. (G17) in this case is evaluated as $\tau_{LR}^2 e^{-(\nu-\beta)\Theta(\Delta)} = o(1)$.

Next, we consider case (2). When $a_B \neq b_B$, $c_B \neq b_B$ holds from $a_B = c_B$. Then, using the off-diagonal ETH, we evaluate $\sum_{a,b,c} M(a, b, c, t)$ restricted to case (2) as

$$\frac{1}{\Theta(D)} \sum_{\substack{a \in \overline{\Lambda_2} \\ b \in \Lambda_1; b_B \neq a_B \\ c; c_B = a_B}} (\delta\rho_B)_{bb}(\rho_S(t))_{ca}$$
$$\times e^{-\beta \delta E_B^{a_B b_B} - \nu(|\delta E_B^{a_B b_B}| + |\delta E_B^{c_B b_B}|)}, \tag{G21}$$

which is shown to exponentially decay as $e^{-(\nu-\beta)\Theta(\Delta)}$ in $\beta < \nu$ using $\sum_{b \in \Lambda} |\delta\rho_{bb}| \leqslant 1$. Then the right-hand side of Eq. (G17) in case (2) is $o(1)$.
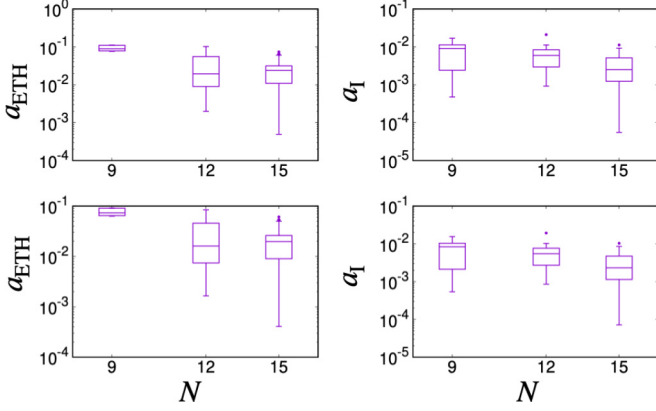
FIG. 16. The dependence of a_{ETH} and a_{I} on the bath size and the initial state. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\gamma' = 4\gamma$ (upper panels) and 0.1γ (lower panels), $\beta = 0.1$. Left panels are a_{ETH}, and right panels are a_{I}.

Finally, we consider case (3). We evaluate $\sum_{a,b,c} M(a, b, c, t)$ restricted to this case as

$$
\sum_{\substack{b_{\mathrm{B}} \in \Lambda'_1 \\ a_{\mathrm{S}}, b_{\mathrm{S}}, c_{\mathrm{S}}}} \left(J_{\mathrm{I}}^{a_{\mathrm{S}} b_{\mathrm{S}}}\right)_{b_{\mathrm{B}} b_{\mathrm{B}}} \left(J_{\mathrm{I}}^{b_{\mathrm{S}} c_{\mathrm{S}}}\right)_{b_{\mathrm{B}} b_{\mathrm{B}}} (\rho_{\mathrm{S}}(t))_{c_{\mathrm{S}} a_{\mathrm{S}}} (\delta\rho_{\mathrm{B}})_{b_{\mathrm{B}} b_{\mathrm{B}}}
$$

$$
\simeq \sum_{a_{\mathrm{S}} b_{\mathrm{S}} c_{\mathrm{S}}} \langle J_{\mathrm{I}}^{a_{\mathrm{S}} b_{\mathrm{S}}} \rangle_{\mathrm{MC}} \langle J_{\mathrm{I}}^{b_{\mathrm{S}} c_{\mathrm{S}}} \rangle_{\mathrm{MC}} (\rho_{\mathrm{S}}(t))_{c_{\mathrm{S}} a_{\mathrm{S}}} \sum_{b_{\mathrm{B}} \in \Lambda_1} (\delta\rho_{\mathrm{B}})_{b_{\mathrm{B}} b_{\mathrm{B}}},
$$

(G22)

where we define $J_{\mathrm{I}}^{a_{\mathrm{S}} b_{\mathrm{S}}} := \langle E_{\mathrm{S}}^{a_{\mathrm{S}}} | H_{\mathrm{I}} | E_{\mathrm{S}}^{b_{\mathrm{S}}} \rangle$ and use the ETH for $J_{\mathrm{I}}^{a_{\mathrm{S}} b_{\mathrm{S}}}$. Also, $|\sum_{b_{\mathrm{B}} \in \Lambda_1} (\delta\rho_{\mathrm{B}})_{b_{\mathrm{B}} b_{\mathrm{B}}}|$ is the contribution of $\rho_{\mathrm{B}}^{\mathrm{can}}$ from outside Λ_1, which decays as $e^{-\Theta(\Delta)}$. Therefore, the right-hand side of Eq. (G17) in case (3) is also $o(1)$.
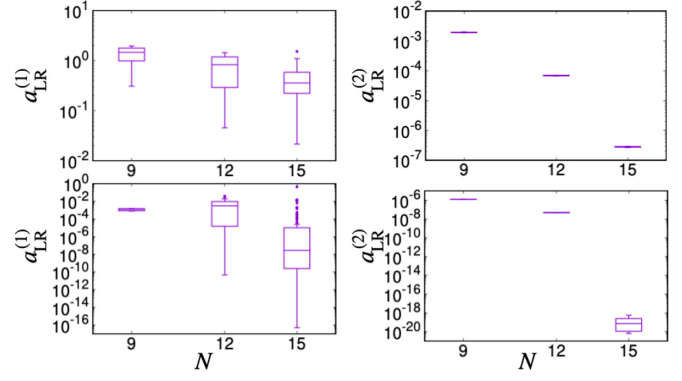

FIG. 17. The dependence of $a_{\mathrm{LR}}^{(1)}$ and $a_{\mathrm{LR}}^{(2)}$ on the bath size and the initial state. Parameters: $p = 0.99$, $g = 0.1\gamma$, $\gamma' = 4\gamma$ (upper panels) and 0.1γ (lower panels), $\beta = 0.1$. Left panels are $a_{\mathrm{LR}}^{(1)}$, and right panels are $a_{\mathrm{LR}}^{(2)}$.

From the foregoing argument, $|\delta G_{\mathrm{I}}| = o(1)$ is shown in the short-time regime ($t \ll \tau_{\mathrm{LR}}$).

APPENDIX H: SUPPLEMENTARY NUMERICAL RESULTS

We first show the time dependence of the error of the fluctuation theorem $|\langle e^{-\sigma} \rangle - 1|$ for various interaction parameters. As in Fig. 13, we observe qualitatively the same behavior as the inset of Fig. 2 as mentioned in Sec. III.

We next show the supplemental data in the long-time regime. Figure 14 shows the γ' dependence of $|\overline{\delta G_{\mathrm{S}}}|$ and $|\overline{\delta G_{\mathrm{I}}}|$ with $p = 0.9$. Also, Fig. 15 shows the γ' dependence of $|\overline{\delta G_{\mathrm{S}}}|$ and $|\overline{\delta G_{\mathrm{I}}}|$ with $g = 0.4\gamma$, showing a similar result as in the main text.

Finally, we show the supplemental data in the short-time regime. Figure 16 shows the N dependence of a_{ETH} and a_{I} with $\gamma'/\gamma = 4$ and 0.1. Both a_{ETH} and a_{I} decrease as N increases. Figure 17 shows the N dependence of $a_{\mathrm{LR}}^{(1,2)}$. We see that both $a_{\mathrm{LR}}^{(1)}$ and $a_{\mathrm{LR}}^{(2)}$ decrease in N.

[1] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).

[2] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. **78**, 2690 (1997).

[3] C. Jarzynski, Hamiltonian derivation of a detailed fluctuation theorem, J. Stat. Phys. **98**, 77 (2000).

[4] G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Phys. Rev. E **60**, 2721 (1999).

[5] J. Kurchan, A quantum fluctuation theorem, arXiv:cond-mat/0007360 (2000).

[6] H. Tasaki, Jarzynski relations for quantum systems and some applications, arXiv:cond-mat/0009244 (2000).

[7] M. Esposito, U. Harbola, and S. Mukamel, Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems, Rev. Mod. Phys. **81**, 1665 (2009).

[8] M. Campisi, P. Hänggi, and P. Talkner, Colloquium: Quantum fluctuation relations: Foundations and applications, Rev. Mod. Phys. **83**, 771 (2011).

[9] T. Sagawa, Second law-Like inequalities with quantum relative entropy an introduction, in *Lectures on Quantum Computing, Thermodynamics and Statistical Physics*, Kinki University Series on Quantum Computing Vol. 8 (World Scientific, Singapore, 2012), p. 127.

[10] Á. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry, M. Campisi, P. Hänggi, and P. Talkner, Fluctuating Work: From Quantum Thermodynamical Identities to a Second Law Equality, Phys. Rev. X **6**, 041017 (2016).

[11] R. Dorner, S. R. Clark, L. Heaney, R. Fazio, J. Goold, and V. Vedral, Extracting Quantum Work Statistics and Fluctuation Theorems by Single-Qubit Interferometry, Phys. Rev. Lett. **110**, 230601 (2013).

[12] S. An, J.-N. Zhang, M. Um, D. Lv, Y. Lu, J. Zhang, Z.-Q. Yin, H. T. Quan, and K. Kim, Experimental test of the quantum Jarzynski equality with a trapped-ion system, Nat. Phys. **11**, 193 (2015).

[13] T. P. Xiong, L. L. Yan, F. Zhou, K. Rehan, D. F. Liang, L. Chen, W. L. Yang, Z. H. Ma, M. Feng, and V. Vedral, Experimen-

tal Verification of a Jarzynski-Related Information-Theoretic Equality by a Single Trapped Ion, Phys. Rev. Lett. **120**, 010601 (2018).

[14] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, Colloquium: Nonequilibrium dynamics of closed interacting quantum systems, Rev. Mod. Phys. **83**, 863 (2011).

[15] J. Eisert, M. Friesdorf, and C. Gogolin, Quantum many-body systems out of equilibrium, Nat. Phys. **11**, 124 (2015).

[16] C. Gogolin and J. Eisert, Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems, Rep. Prog. Phys. **79**, 056001 (2016).

[17] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics, Adv. Phys. **65**, 239 (2016).

[18] T. Mori, T. N. Ikeda, E. Kaminishi, and M. Ueda, Thermalization and prethermalization in isolated quantum systems: A theoretical overview, J. Phys. B **51**, 112001 (2018).

[19] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck J. Eisert, and I. Bloch, Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas, Nat. Phys. **8**, 325 (2012).

[20] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, Science **353**, 794 (2016).

[21] C. Gross and I. Bloch, Quantum simulations with ultracold atoms in optical lattices, Science **357**, 995 (2017).

[22] M. F. Parsons, F. Huber, A. Mazurenko, C. S. Chiu, W. Setiawan, K. Wooley-Brown, S. Blatt, and M. Greiner, Site-Resolved Imaging of Fermionic ^{6}Li in an Optical Lattice, Phys. Rev. Lett. **114**, 213002 (2015).

[23] D. Ohl de Mello, D. Schäffner, J. Werkmann, T. Preuschoff, L. Kohfahl, M. Schlosser, and G. Birkl, Defect-Free Assembly of 2D Clusters of More than 100 Single-Atom Quantum Systems, Phys. Rev. Lett. **122**, 203601 (2019).

[24] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho *et al.*, Quantum phases of matter on a 256-atom programmable quantum simulator, Nature (London) **595**, 227 (2021).

[25] G. Clos, D. Porras, U. Warring, and T. Schaetz, Time-Resolved Observation of Thermalization in an Isolated Quantum System, Phys. Rev. Lett. **117**, 170401 (2016).

[26] C. Neill, P. Roushan, M. Fang, Y. Chen, M. Kolodrubetz, Z. Chen, A. Megrant, R. Barends, B. Campbell, B. Chiaro *et al.* Ergodic dynamics and thermalization in an isolated quantum system, Nat. Phys. **12**, 1037 (2016).

[27] J. M. Deutsch, Quantum statistical mechanics in a closed system, Phys. Rev. A **43**, 2046 (1991).

[28] M. Srednicki, Chaos and quantum thermalization, Phys. Rev. E **50**, 888 (1994).

[29] M. Rigol, V. Dunjko, and M. Olshanii, Thermalization and its mechanism for generic isolated quantum systems, Nature (London) **452**, 854 (2008).

[30] G. Biroli, C. Kollath, and A. M. Läuchli, Effect of Rare Fluctuations on the Thermalization of Isolated Quantum Systems, Phys. Rev. Lett. **105**, 250401 (2010).

[31] R. Steinigeweg, J. Herbrych, and P. Prelovsek, Eigenstate thermalization within isolated spin-chain systems, Phys. Rev. E **87**, 012118 (2013).

[32] H. Kim, T. N. Ikeda, and D. A. Huse, Testing whether all eigenstates obey the eigenstate thermalization hypothesis, Phys. Rev. E **90**, 052105 (2014).

[33] W. Beugeling, R. Moessner, and M. Haque, Finite-size scaling of eigenstate thermalization, Phys. Rev. E **89**, 042112 (2014).

[34] W. Beugeling, R. Moessner, and M. Haque, Off-diagonal matrix elements of local operators in many-body quantum systems, Phys. Rev. E **91**, 012144 (2015).

[35] K. R. Fratus and M. Srednicki, Eigenstate thermalization in systems with spontaneously broken symmetry, Phys. Rev. E **92**, 040103(R) (2015).

[36] R. Mondaini, K. R. Fratus, M. Srednicki, and M. Rigol, Eigenstate thermalization in the two-dimensional transverse field Ising model, Phys. Rev. E **93**, 032104 (2016).

[37] R. Mondaini and M. Rigol, Eigenstate thermalization in the two-dimensional transverse field Ising model. II. Off-diagonal matrix elements of observables, Phys. Rev. E **96**, 012157 (2017).

[38] T. Yoshizawa, E. Iyoda, and T. Sagawa, Numerical Large Deviation Analysis of the Eigenstate Thermalization Hypothesis, Phys. Rev. Lett. **120**, 200604 (2018).

[39] J. R. Garrison and T. Grover, Does a Single Eigenstate Encode the Full Hamiltonian? Phys. Rev. X **8**, 021026 (2018).

[40] A. Dymarsky, N. Lashkari, and H. Liu, Subsystem eigenstate thermalization hypothesis, Phys. Rev. E **97**, 012140 (2018).

[41] I. M. Khaymovich, M. Haque, and P. A. McClarty, Eigenstate Thermalization, Random Matrix Theory, and Behemoths, Phys. Rev. Lett. **122**, 070601 (2019).

[42] M. Brenes, T. LeBlond, J. Goold, and M. Rigol, Eigenstate Thermalization in a Locally Perturbed Integrable System, Phys. Rev. Lett. **125**, 070605 (2020).

[43] K. Kaneko, E. Iyoda, and T. Sagawa, Characterizing complexity of many-body quantum dynamics by higher-order eigenstate thermalization, Phys. Rev. A **101**, 042126 (2020).

[44] E. Iyoda, K. Kaneko, and T. Sagawa, Fluctuation Theorem for Many-Body Pure Quantum States, Phys. Rev. Lett. **119**, 100601 (2017).

[45] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems, Commun. Math. Phys. **28**, 251 (1972).

[46] M. B. Hastings and T. Koma, Spectral gap and exponential decay of correlations, Commun. Math. Phys. **265**, 781 (2006).

[47] J. Haah, M. B. Hastings, R. Kothari, and G. H. Low, Quantum algorithm for simulating real time evolution of lattice Hamiltonians, SIAM J. Comput. (2018), doi: 10.1137/18M1231511.

[48] Y. Murashita, K. Funo, and M. Ueda, Nonequilibrium equalities in absolutely irreversible processes, Phys. Rev. E **90**, 042110 (2014).

[49] K. Funo, Y. Murashita, and M. Ueda, Quantum nonequilibrium equalities with absolute irreversibility, New J. Phys. **17**, 075005 (2015).

[50] K. Kaneko, E. Iyoda, and T. Sagawa, Saturation of entropy production in quantum many-body systems, Phys. Rev. E **96**, 062148 (2017).

[51] F. Jin, R. Steinigeweg, H. De Raedt, K. Michielsen, M. Campisi, and J. Gemmer, Eigenstate thermalization hypothesis and quantum Jarzynski relation for pure initial states, Phys. Rev. E **94**, 012125 (2016).

[52] D. Schmidtke, L. Knipschild, M. Campisi, R. Steinigeweg, and J. Gemmer, Stiffness of probability distributions of work and

Jarzynski relation for non-Gibbsian initial states, Phys. Rev. E **98**, 012123 (2018).

[53] G. Manzano, J. M. Horowitz, and J. M. R. Parrondo, Quantum Fluctuation Theorems for Arbitrary Environments: Adiabatic and Nonadiabatic Entropy Production, Phys. Rev. X **8**, 031037 (2018).

[54] G. T. Landi and M. Paternostro, Irreversible entropy production, from quantum to classical, Rev. Mod. Phys. **93**, 035008 (2021).

[55] Y. Masuyama, K. Funo, Y. Murashita, A. Noguchi, S. Kono, Y. Tabuchi, R. Yamazaki, M. Ueda, and Y. Nakamura, Information-to-work conversion by Maxwell's demon in a superconducting circuit quantum electrodynamical system, Nat. Commun. **9**, 1291 (2018).

[56] N. Cottet, S. Jezouin, L. Bretheau, P. C. Ibarcq, Q. Ficheux, J. Anders, A. Auffèves, R. Azouit, P. Rouchon, and B. Huard, Observing a quantum Maxwell demon at work, Proc. Natl. Acad. Sci. USA **114**, 7561 (2017).

[57] K. Vogel and H. Risken, Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, Phys. Rev. A **40**, 2847 (1989).

[58] D. Ruelle, *Statistical Mechanics: Rigorous Results* (World Scientific, Singapore, 1999).

[59] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papic, Weak ergodicity breaking from quantum many-body scars, Nat. Phys. **14**, 745 (2018).

[60] Y. Ogata, Large Deviations in quantum spin chains, Commun. Math. Phys. **296**, 35 (2010).

[61] T. Mori, Weak eigenstate thermalization with large deviation bound, arXiv:1609.09776 (2016).

[62] T. LeBlond, K. Mallayya, L. Vidmar, and M. Rigol, Entanglement and matrix elements of observables in interacting integrable systems, Phys. Rev. E **100**, 062134 (2019).

[63] L. Vidmar, L. Hackl, E. Bianchi, and M. Rigol, Entanglement Entropy of Eigenstates of Quadratic Fermionic Hamiltonians, Phys. Rev. Lett. **119**, 020601 (2017).

[64] A. Pal and D. A. Huse, Many-body localization phase transition, Phys. Rev. B **82**, 174411 (2010).

[65] R. Nandkishore and D. A. Huse, Many-body localization and thermalization in quantum statistical mechanics, Annu. Rev. Condens. Matter Phys. **6**, 15 (2015).

[66] H.-P. Breuer, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2007).

[67] L. P. Garcia-Pintos, N. Linden, A. S. L. Malabarba, A. J. Short, and A. Winter, Equilibration Time Scales of Physically Relevant Observables, Phys. Rev. X **7**, 031027 (2017).

[68] P. Reimann, Typical fast thermalization processes in closed many-body systems, Nat. Commun. **7**, 10821 (2016).

[69] D. Hetterich, M. Fuchs, and B. Trauzettel, Equilibration in closed quantum systems: Application to spin qubits, Phys. Rev. B **92**, 155314 (2015).

[70] M. Rigol, Quantum quenches and thermalization in one- dimensional fermionic systems, Phys. Rev. A **80**, 053607 (2009).

[71] R. Heveling, J. Wang, and J. Gemmer, Integral fluctuation theorem for microcanonical and pure states, arXiv:2102.12294 (2021).

[72] P. Bocchieri and A. Loinger, Quantum recurrence theorem, Phys. Rev. **107**, 337 (1957).

[73] E. C. Percival, Almost periodicity and the quantal *H* theorem, J. Math. Phys. **2**, 235 (1961).

[74] L. C. Venuti, The recurrence time in quantum mechanics, arXiv:1509.04352 (2015).

[75] J.-Y. Choi, S. Hild, J. Zeiher, P. Schass, A. R.-Abadal, T. Yefsah, V. Kehmani, D. A. Huse, I. Bloch, and C. Gross, Exploring the many-body localization transition in two dimensions, Science **352**, 1547 (2016).

[76] P. Reimann, Foundation of Statistical Mechanics under Experimentally Realistic Conditions, Phys. Rev. Lett. **101**, 190403 (2008).

[77] H. Tasaki, On the local equivalence between the canonical and the microcanonical ensembles for quantum spin systems, J. Stat. Phys. **172**, 905 (2018).

[78] T. Sagawa, Thermodynamics of information processing in small systems, Prog. Theor. Phys. **127**, 1 (2012).

[79] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, Quantum and Information Thermodynamics: A Unifying Framework Based on Repeated Interactions, Phys. Rev. X **7**, 021003 (2017).

[80] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[81] L. Szilard, On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings, Z. Phys. **53**, 840 (1929).

[82] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[83] T. Sagawa and M. Ueda, Nonequilibrium thermodynamics of feedback control, Phys. Rev. E **85**, 021104 (2012).

[84] S. Sodin, The spectral edge of some random band matrices, Ann. Math. **172**, 2223 (2010).

[85] R. Hamazaki and M. Ueda, Atypicality of Most Few-Body Observables, Phys. Rev. Lett. **120**, 080603 (2018).

[86] K. Kaneko, E. Iyoda, and T. Sagawa, Work extraction from a single energy eigenstate, Phys. Rev. E **99**, 032128 (2019).

[87] I. Arad, T. Kuwahara, and Z. Landau, Connecting global and local energy distributions in quantum spin models on a lattice, J. Stat. Mech. **2016**, 033301 (2016).