# Entropy production of a closed Hamiltonian system via the detailed fluctuation relation

Yûto Murashita ⬤,[1] Naoto Kura,[1] and Masahito Ueda[1,2,3]

[1]*Department of Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[2]*Institute for Physics of Intelligence, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[3]*RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan*

We revisit the problem of emergent irreversibility in a closed Hamiltonian system in light of the detailed fluctuation relation by invoking an imperfect Loschmidt demon that performs a time-reversal operation with a finite precision. The imperfect time reversal can be utilized to evaluate entropy production à la Kolmogorov and Sinai in a manner consistent with the fluctuation theorem with absolute irreversibility.

## I. INTRODUCTION

Understanding emergent irreversibility in macrophysics from reversible microphysics has been the holy grail in statistical physics. Boltzmann studied the kinetic theory of gases to show the H-theorem [1]. However, his proof relies on the molecular chaos assumption, and therefore it has remained elusive why macroscopic irreversibility emerges from reversible microscopic dynamics as epitomized by the Loschmidt irreversibility paradox [2–4]. In the mid-20th century, researchers revisited the issues of how entropy should be defined and why irreversibility emerges in chaotic dynamical systems [5,6]. In a related context, the Kolmogorov-Sinai theory introduces a minimal accessible length scale to coarse-grain phase space and evaluates scale-invariant emergent irreversibility in terms of the Kolmogorov-Sinai entropy [5,6]. Meanwhile, the fluctuation theorem allows a quantitative estimate of far-from-equilibrium irreversibility [7,8], where the scale separation is implicit due to the assumption of a built-in equilibration mechanism. Indeed, initial [9,10] or continuous [11,12] contact of a system with an ideal heat bath is assumed in the derivation of the standard fluctuation theorem. One is naturally led to ask whether the key idea of the Kolmogorov-Sinai theory can justify the built-in assumption of the fluctuation theorem, thereby allowing one to evaluate emergent irreversibility and entropy production in a closed Hamiltonian system.

In this paper, we investigate how irreversibility emerges in a closed chaotic Hamiltonian system by incorporating a minimal accessible length scale à la Kolmogorov and Sinai. We utilize a unified framework of the fluctuation theorem [8,13] and evaluate an entropy production by constructing a scale-dependent dual process with an imperfect Loschmidt demon,

which executes the time-reversal operation with a finite precision. The crucial observation here is that while the value of entropy production itself depends on the precision, its growth rate does not and agrees with the scale-invariant Kolmogorov-Sinai entropy. In a spirit similar to the Kolmogorov-Sinai theory, we introduce a nonzero coarse-graining length scale in phase space to extract scale-invariant intrinsic irreversibility in chaotic dynamics. Thus the imperfect Loschmidt demon naturally integrates the key idea of the Kolmogorov-Sinai theory with fluctuation theorems to evaluate emergent irreversibility of a closed Hamiltonian system.

In Sec. II, we define an entropy production via the detailed fluctuation relation by invoking the time-reversal process by the imperfect Loschmidt demon. In Sec. III, we consider the Bunimovich billiard as a prototypical example and demonstrate that the entropy production exhibits time evolution analogous to the Kolmogorov-Sinai entropy. In Sec. IV, we conclude the paper. In Appendix A, we elaborate on the details and precision of numerical simulations. In Appendix B, we discuss our conjecture that the growth rate of the entropy production coincides with the Kolmogorov-Sinai entropy for a generic closed chaotic Hamiltonian system.

## II. ENTROPY PRODUCTION VIA THE DETAILED FLUCTUATION RELATION

We quantify irreversibility on the basis of a variant of the finite-time fluctuation theorem [9–12]. We follow Refs. [8,13] to consider a dual version of the original nonequilibrium process and compare the probabilities of the original and dual processes to define a formal entropy production $\sigma$ by the detailed fluctuation relation

$$\sigma = -\ln \frac{\tilde{P}[\Gamma]}{P[\Gamma]}, \qquad (1)$$

where $\Gamma$ and $\mathcal{P}$ represent a path in phase space and the path probability, respectively, and the tilde indicates that the process is dual. In the standard formulation, the physical meaning of the entropy production can be given if we assume that the system is either continuously attached to a heat bath or
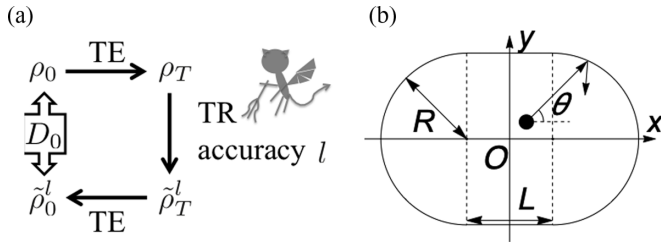
FIG. 1. Schematic illustration of a time-reversal (TR) test by an imperfect Loschmidt demon and the Bunimovich billiard. (a) Protocol of the TR test. An initial state $\rho_0$ undergoes time evolution (TE) over time $T$ to reach a final state $\rho_T$. The Loschmidt demon conducts the TR operation $\mathcal{T}$ with a nonzero resolution $l$ to produce the coarse-grained state $\tilde{\rho}_T^l$, which then evolves during time $T$ to arrive at the final state $\tilde{\rho}_0^l$. By comparing it with $\rho_0$, one can evaluate the Rényi-0 divergence $D_0$ which gives the lower bound for the average entropy production. See the text for details. (b) Geometry of the Bunimovich stadium. A single particle ballistically moves and elastically bounces on the boundary.

initially equilibrated with it (and virtually attached to it at the end as explained in Ref. [9]). We note that, if we formally apply the fluctuation theorem to a Hamiltonian system with no heat bath by setting the dual process to the time-reversal one that starts from the final state of the forward process, we obtain a trivial result with vanishing entropy production.

We here consider a nonequilibrium process in a closed Hamiltonian system. Let $d_E$ be the dimension of phase space. The system starts from the initial state $\rho_0$ and evolves over time $T$ into the final state $\rho_T$ [see Fig. 1(a)]. Then, the Loschmidt demon performs the time-reversal test [14,15]. We assume that the demon perturbs the final state $\rho_T$ by introducing a nonzero spatial resolution $l$ into the state $\mathcal{C}_l[\rho_T]$, where $\mathcal{C}_l[\rho]$ denotes the convolution of $\rho$ with the $d_E$-dimensional isotropic Gaussian with the standard deviation $l$. Thus the initial state of the time-reversed process is $\tilde{\rho}_T^l = \mathcal{T}\mathcal{C}_l[\rho_T]$, where $\mathcal{T}$ represents the velocity reversal. Then, the demon lets the system evolve over time $T$ to reach the final state $\tilde{\rho}_0^l$. Our protocol is reminiscent of the Loschmidt echo [16], which is used to characterize decoherence in classically chaotic quantum systems. The crucial distinction is that the state $\rho_T$ is perturbed in our protocol, whereas the backward Hamiltonian is perturbed in the Loschmidt echo.

When the system is chaotic, the final state of the backward process $\tilde{\rho}_0^l$ deviates more from the initial state of the forward process $\rho_0$ as the evolution time $T$ becomes longer. Although this qualitative behavior is not surprising due to chaoticity, we can exploit the behavior to quantitatively define entropy production via the detailed fluctuation relation as detailed below.

We consider the dual dynamics as the above-mentioned imperfect time-reversal process $\tilde{\rho}_T^l \to \tilde{\rho}_0^l$ and define an entropy production by Eq. (1). Since the dynamics is deterministic, the path probability is determined only by the initial phase-space point $\gamma_0$ and therefore we obtain $\sigma = -\ln[\mathcal{T}\tilde{\rho}_0^l(\gamma_0)/\rho_0(\gamma_0)]$. Note that the ensemble average $\langle\sigma\rangle$ is not the thermodynamic entropy production but an information-theoretic measure known as the Kullback-Leibler divergence or the Rényi-1 divergence [17]: $D_1(\rho_0||\mathcal{T}\tilde{\rho}_0^l) :=$

$\langle\ln[\rho_0(\gamma)/\mathcal{T}\tilde{\rho}_0^l(\gamma)]\rangle := \int \rho_0(\gamma)\ln[\rho_0(\gamma)/\mathcal{T}\tilde{\rho}_0^l(\gamma)]d\gamma$. Since $D_1(\rho_0||\mathcal{T}\tilde{\rho}_0^l) = D_1(\rho_T||\mathcal{T}\tilde{\rho}_T^l)$, it gives a (quasi)distance between the final state $\rho_T$ and the state $\mathcal{T}\tilde{\rho}_T^l = \mathcal{C}_l[\rho_T]$ that one observes with accuracy $l$.

As is usual in the standard framework of fluctuation theorems, entropy production $\sigma$ represents irreversibility of the forward process alone rather than the reversal operation or the backward process. Therefore the entropy production discussed here does not involve irreversibility that is caused by the imperfect Loschmidt demon but represents irreversibility generated in the forward isolated process. Moreover, as we demonstrate later, the scale $l$ only parametrically changes the value of entropy production but does not change its growth rate. Thus the growth rate represents scale-invariant intrinsic irreversibility of the forward isolated dynamics. We stress that this length scale must be introduced in any realistic experiments or numerical simulations to be consistent with the measurement accuracy or the numerical precision, since structures originally below the resolution limit grow to be observed by chaotic exponential expansion after a certain time. Thus the external perturbation due to the imperfect Loschmidt demon should be considered as a technical tool to extract intrinsic irreversibility in a controlled manner.

Notably, entropy production (1) does not always satisfy the conventional integral fluctuation theorem, i.e., $\langle e^{-\sigma}\rangle \neq 1$. This is because $\sigma$ is divergent when the initial phase-space point $\gamma_0$ is outside the support of $\rho_0$ as $\sigma = -\ln[\mathcal{T}\tilde{\rho}_0^l(\gamma_0)/0]$. This phenomenon with divergent entropy production should be distinguished from ordinary irreversibility due to its singular nature and is referred to as absolute irreversibility [18]. In the presence of absolute irreversibility, the integral fluctuation theorem should be modified [18] as

$$\langle e^{-\sigma}\rangle = 1 - \lambda, \tag{2}$$

where $\lambda$ is the fraction of the singular part of $\tilde{\rho}_0^l$ with respect to $\rho_0$ in the measure-theoretic sense (see, e.g., Ref. [19]). In this case, since $\lambda = \int_{\rho(\gamma)=0} \mathcal{T}\tilde{\rho}_0^l(\gamma)d\gamma$, we have $1 - \lambda = e^{-D_0}$, where $D_0$ is an information-theoretic measure called the Rényi-0 divergence which is defined as $D_0(\rho||\mathcal{C}_l[\rho]) := -\ln\int_{\rho(\gamma)>0}\mathcal{C}_l[\rho](\gamma)d\gamma$ [17]. By applying the Jensen inequality to Eq. (2), we obtain

$$\langle\sigma\rangle = D_1(\rho_T||\mathcal{C}_l[\rho_T]) \geqslant -\ln(1 - \lambda) = D_0(\rho_T||\mathcal{C}_l[\rho_T]). \tag{3}$$

We note that the magnitude relation between $D_0$ and $D_1$ is well established in information theory [17].

The absolute irreversibility discussed in Ref. [18] originates from a built-in thermalization mechanism or measurement in the forward process. In particular, the reversal operation to derive the Jarzynski-type work relation consists of thermal equilibration at the final parameter value of the forward process and velocity reversal. Here, absolute irreversibility originates directly from the assumed equilibration and is not emergent. In contrast, absolute irreversibility discussed in this paper dynamically arises from nonzero precision with no assumption on any thermalization mechanism.
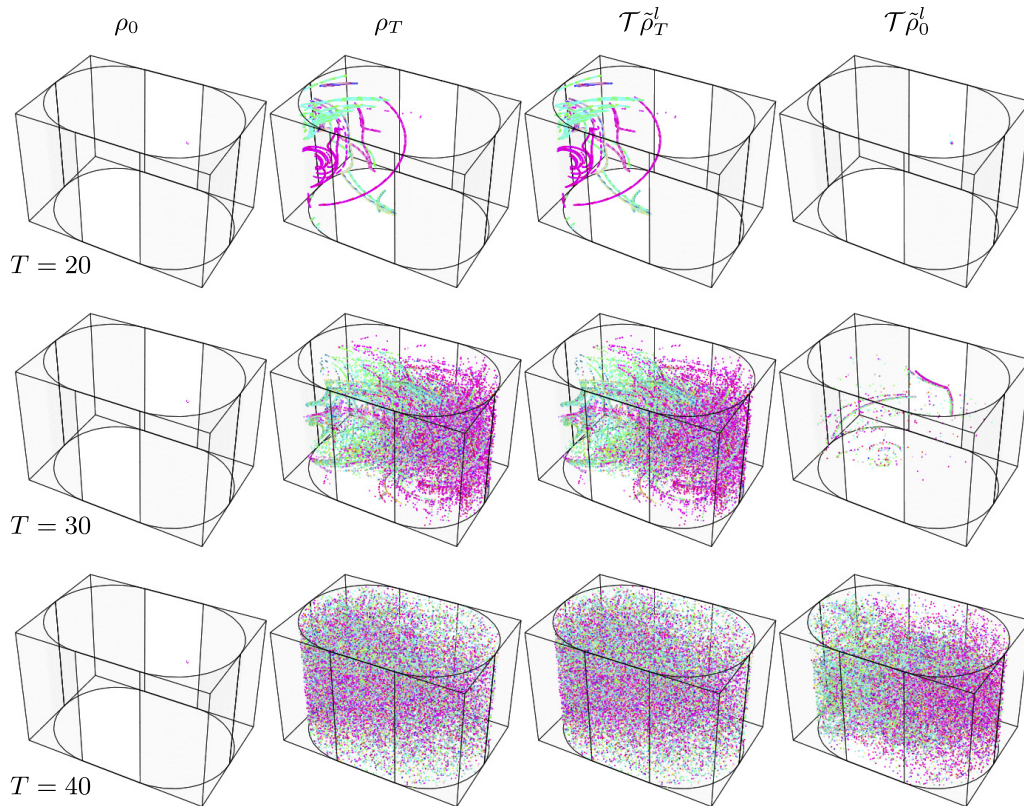
FIG. 2. Time-reversal test by an imperfect Loschmidt demon. The state starting from a small region undergoes time evolution over time $T$ and then traces the time-reversed path after coarse graining. The RGB values of colors correspond to the slightly different three-dimensional positions in the initial cubic state. The vertical axis corresponds to the direction of the velocity $\theta$ as indicated in Fig. 1. When the evolution time $T$ is 20, $\mathcal{T}\tilde{\rho}_0^l$ is almost identical to the initial state $\rho_0$. However, as $T$ increases further, $\mathcal{T}\tilde{\rho}_0^l$ deviates dramatically from $\rho_0$ as measured by an increase in the Rényi-0 divergence between them. Although $\rho_T$ and $\mathcal{T}\tilde{\rho}_T^l$ appear identical, they actually differ for large $T$, and this difference can efficiently be evaluated from the difference between $\rho_0$ and $\mathcal{T}\tilde{\rho}_0^l$. Thus the time-reversal test permits evaluation of $D_0(\rho_T||\mathcal{T}\tilde{\rho}_T^l)$, which would otherwise be extremely challenging to calculate. See Appendix A for details of numerical simulations.

## III. BUNIMOVICH BILLIARD AS A PROTOTYPICAL EXAMPLE

As a prototypical example, we consider an elastic billiard in the Bunimovich stadium [20] [see Fig. 1(b)]. The phase space of the Bunimovich billiard can be represented as a three-dimensional space ($d_E = 3$) due to the conservation of energy; a point in the stadium is characterized by the Cartesian coordinates ($x$ and $y$) and the direction of the velocity ($\theta$).

We use the Monte Carlo method to numerically evaluate entropy production. First of all, we sample a phase-space point $\gamma_0$ from the probability distribution function at an initial time $\rho_0$, which is prepared to be a uniform distribution in a small cube. Then, we let it evolve over time $T$ to obtain $\gamma_T$, where the ensemble of these points is the final state $\rho_T$. An imperfect Loschmidt demon adds to $\gamma_T$ a Gaussian noise with the standard deviation $l$ to obtain $\gamma_T'$ and inverts its velocity $\tilde{\gamma}_T = \mathcal{T}\gamma_T'$ to form an ensemble $\tilde{\rho}_T^l = \mathcal{T}\mathcal{C}_l[\rho_T]$. Finally, the demon lets every phase-space point in this ensemble evolve over time $T$ to reach the point $\tilde{\gamma}_0$ belonging to the corresponding ensemble $\tilde{\rho}_0^l$. We note that $D_0$ is far more numerically tractable than $D_1$ and that the evaluation of $D_{0/1}$ is not feasible at the final time because $\rho_T$ has an extremely complex support with an exponentially stretched boundary.

When $l$ is small, the difference between $\rho_T$ and $\mathcal{T}\tilde{\rho}_T^l$ cannot be discerned (see Fig. 2). However, after the reversed time evolution, the difference between $\rho_0$ and $\mathcal{T}\tilde{\rho}_0^l$ becomes significant. For $T = 20$, the state $\mathcal{T}\tilde{\rho}_0^l$ returns to the initial state almost completely. However, as $T$ increases further, $\mathcal{T}\tilde{\rho}_0^l$ deviates from the initial state $\rho_0$ and expands over the entire phase space due to chaoticity of the system, resulting in a larger value of $D_0$.

Figure 3(a) plots the Rényi-0 divergence $D_0$ as a function of $l$ and $T$. For intermediate values of $l$ and $T$, we find that $D_0$ lies on a tilted plane. To understand this behavior, we examine the box-counting dimension $d$, which is a key fractal dimension for the final state $\rho_T$. Then, the Rényi-0 divergence is evaluated as

$$D_0(\rho_T||\mathcal{C}_l[\rho_T]) = d_C \ln(l/l_0), \qquad (4)$$

where $l_0$ is the minimal length scale of $\rho_T$ and we define the complement of the box-counting dimension with respect to the embedding dimension as the codimension: $d_C := d_E - d$. See Appendix B for details. Here, we assume that $l$ is larger than $l_0$ but sufficiently smaller than the system size. Consequently, we can extract the dimension of $\rho_T$ by observing the behaviors of the Rényi-0 divergence against the variations of the length scale $l$. By substituting $l_0 \simeq we^{-\Lambda T}$ into Eq. (4),
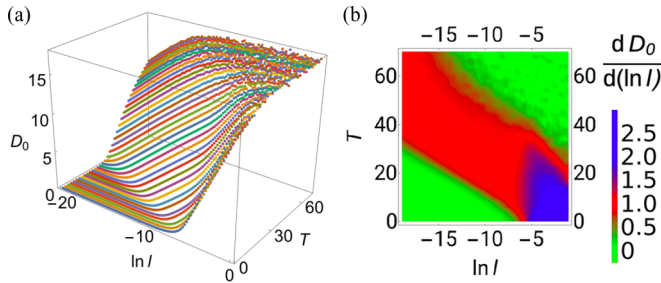
FIG. 3. Numerically calculated Rényi-0 divergence. (a) Rényi-0 divergence $D_0$ against the minimal length scale $l$ and the evolution time $T$. A linear dependence of $D_0$ on $\ln l$ in an intermediate length scale vindicates Eq. (4). (b) Derivative of the Rényi-0 divergence $D_0$ with respect to $\ln l$. The red plateau with a constant value demonstrates that Eq. (4) is applicable to the transient region with a fixed length scale. The height of the plateau is $d_C = 1$, which indicates $d = 2$. See Appendix A for details of numerical simulations.

we obtain

$$D_0 \simeq d_C \ln l + d_C \Lambda T + \text{const.} \quad (5)$$

To see the $l$ dependence more clearly, Fig. 3(b) gives the derivative of $D_0$ with respect to $\ln l$. The plateau at the height $d_C = 1$ indicates that $\rho_T$ has the dimension of $d = 2$. As a result, the proportionality constant for $T$ gives $\Lambda$, which coincides with the Kolmogorov-Sinai entropy $h_{KS}$ of this system, as the Pesin formula [21] states that the Kolmogorov-Sinai entropy is equal to the sum of the positive Lyapunov exponents. We expect that the same linear growth, $h_{KS}T$, is generic for a closed chaotic Hamiltonian system as discussed in Appendix B. A similar growth in an intermediate time regime was found for a coarse-grained Shannon entropy in chaotic

conservative maps [22]. We also note that in an open volume-preserving system the growth rate of the coarse-grained entropy production converges to a similar positive constant in the limit of infinite system size and infinite precision [6,23]. See Appendix A for a discussion on numerical accuracy.

In Fig. 4(a), we numerically calculate the probability distribution function of $\sigma$ and verify the fluctuation theorem with absolute irreversibility in Fig. 4(b). Figures 4(c) and 4(d) compare $\langle \sigma \rangle$ with the Rényi-0 divergence $D_0 = -\ln(1 - \lambda)$ in logarithmic and linear scales, respectively. We see that $\langle \sigma \rangle$ is bounded from below by $D_0$.

Moreover, we can see that three different temporal regions exist. Initially, $D_0$ exponentially increases as $e^{\Lambda T}$, where $\Lambda$ is a positive Lyapunov exponent [see Fig. 4(c)]. Then, in an intermediate regime, $D_0$ grows linearly in time with a proportionality constant $h_{KS}$ as shown in Fig. 4(d) [see Eq. (5)]. Finally, $D_0$ gets saturated at a value determined by the ratio of the initial-state volume to the entire phase-space volume. We note that the behaviors in these three regimes are analogous to those for the coarse-grained Shannon entropy [22].

We can delay the occurrence of this finite-size effect to a later time by reducing the initial volume as shown in Fig. 4(e). The linear growth in the transient time scale is reminiscent of the constant entropy production in steady states of dissipative systems. As discussed in Appendix B, we expect that a generic chaotic closed Hamiltonian system exhibits a transient linear growth of entropy with a rate equal to the Kolmogorov-Sinai entropy until finite-size effects enter in significantly.

## IV. CONCLUSION

We have demonstrated that we can evaluate emerging entropy production of the nonequilibrium dynamics of a closed
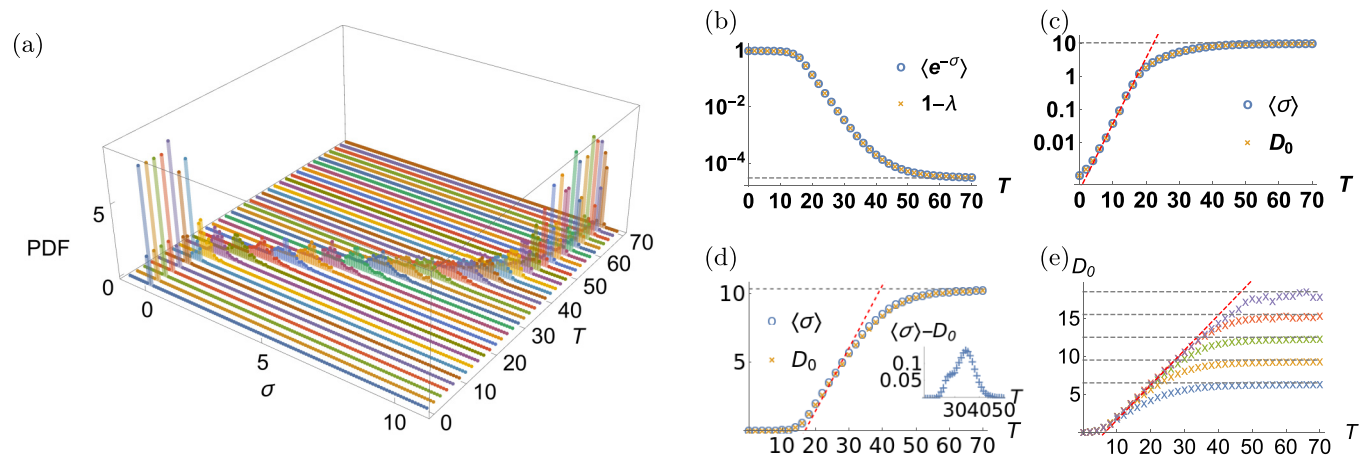


FIG. 4. Entropy production and fluctuation theorem with absolute irreversibility. (a) Time evolution of the probability distribution function (PDF) of entropy production $\sigma$ in Eq. (1). The PDF shifts with time toward larger $\sigma$, indicating that stronger irreversibility should occur in the time-reversal test for longer $T$. (b) Verification of the fluctuation theorem with absolute irreversibility in Eq. (2). The exponential average of entropy production $\langle e^{-\sigma} \rangle$ calculated from the PDF of $\sigma$ and the value of absolute irreversibility $\lambda$ are plotted against time. An excellent agreement between them indicates the validity of the fluctuation theorem with absolute irreversibility. The dashed line indicates the saturated value due to a finite-size effect. (c) Bound of the second-law-like inequality (3) shown in the logarithmic scale. The red dashed line represents an exponential growth in time. (d) Same data as in (c) in the linear scale. The red dashed line shows the linear growth in Eq. (5), which emerges after the exponential growth. Inset: Difference between $\langle \sigma \rangle$ and $D_0$, which is non-negative according to Eq. (3). (e) Saturation and finite-size effects. By reducing the initial phase-space volume (from bottom to top), we can increase the saturated value due to the finite-size effect. See Appendix A for details of numerical simulations.

Hamiltonian system by invoking the Kolmogorov-Sinai theory with the assistance of an imperfect Loschmidt demon. By identifying the process performed by the demon as a dual process for the detailed fluctuation relation, we have evaluated entropy production $\sigma$ and shown that it satisfies the fluctuation theorem with absolute irreversibility. The Rényi divergence tells us the difference between $\rho_T$ and $\mathcal{T}\tilde{\rho}_T^l = \mathcal{C}_l[\rho_T]$, where the latter is the state the demon actually observes due to precision $l$. We find that the Rényi divergence grows linearly in the quasisteady state in accordance with the growth rate of the Kolmogorov-Sinai entropy. It would be interesting to investigate how this information-theoretic irreversibility can be related to the thermodynamic irreversibility. An extension to a quantum regime should merit further study because the length scale set by the Planck constant $\hbar$ should provide the fundamental cutoff in addition to $l$ and we expect an interesting interplay between them.

### APPENDIX A: DETAILS AND PRECISION OF NUMERICAL SIMULATIONS

#### 1. Details of numerical simulations

Throughout our numerical simulations, the geometry of the stadium is set to $R = 1$ and $L = 1$ [see Fig. 1(a)], and the velocity of the billiard is set to unity. The initial state $\rho_0$ is prepared to be a uniform distribution over the cube with the linear dimension of $w$ and the least-valued vertex $(x_0, y_0, \theta_0) = (0.5, 0.5, \pi/4)$.

In Fig. 2, we set $w = 0.01$ and $l = e^{-15}$.

In Fig. 3(a), we set $w = 0.01$ and calculate the Rényi-0 divergence $D_0$ by the Monte Carlo method as described in the main text, where we increase the length scale $l$ from $e^{-20}$ to 1 by the multiplication of $e^{0.2}$ and varying $T$ from 0 to 70 with an increment of 2. For each pair of $l$ and $T$, we sample $10^9$ events from a low-discrepancy sequence generated by the additive recurrence. Although the numerical precision decreases as the value of $D_0$ increases, the relative statistical error can be estimated to be at most 2%. In Fig. 3(b), we calculate $dD_0/d\ln l$ by locally applying the least-squares fitting to $D_0$ with respect to $\ln l$.

In Fig. 4(a), we calculate the probability distribution of entropy production by setting $w = 0.1$ and $l = e^{-10}$. To evaluate the entropy production, we should calculate the probability distribution $\tilde{\rho}_0^l$ in the support of $\rho_0$. To this aim, we divide the support of $\rho_0$ into $8^3$ subcubes. In each subcube, we generate $2^{24} \sim 2 \times 10^7$ samples as the initial state and perform the time-reversal test. The total sampling number is therefore $2^{33} \sim 9 \times 10^9$. By accumulating the events that terminate in

each subcube, we numerically evaluate the probability distribution $\tilde{\rho}_0^l$. We increase $T$ from 0 to 70 by an increment of 2. As $T$ increases, the statistical error increases because $\tilde{\rho}_0^l$ diffuses and the number of events that return to each subcube decreases. Nevertheless, the relative statistical errors are at most 5%. From the probability distribution of the empirical entropy production in Fig. 4(a), the data in Figs. 4(b)–4(d) are calculated. In Fig. 4(e), we set $l = e^{-3}w$ and decrease $w$ from $e^{-1}$ to $e^{-5}$ by the multiplication of $e^{-1}$. For each $w$ and $T$, the Rényi-0 divergence $D_0$ is evaluated from $10^9$ samples.

#### 2. Chaos and numerical precision

In this section, we argue that the double-precision arithmetics is enough to simulate our chaotic system under the present parameters. To numerically examine the effect of a finite precision, we conduct a time-reversal test with no noise. Let $10^{-p}$ be the relative precision of the floating number that we use in our simulations. The largest error originates from the error in the stable direction at the final state. When the velocity is reversed, the direction becomes unstable, and the error grows exponentially as $e^{\Lambda T}$. Therefore the typical error is expected to be

$$\epsilon \simeq 10^{-p} e^{\Lambda T} \simeq 10^{0.2T-p}, \tag{A1}$$

where we use $\Lambda = 0.46$ for the Bunimovich billiard with our parameters. We numerically confirm this estimate by varying the precision of floating numbers as shown in Fig. 5, which shows that the reliability condition for our numerical simulation is $w \gg \epsilon \simeq 10^{0.2T-p}$. Therefore, for $w = 0.01$ and the double precision ($p = 16$), the reliability is guaranteed for $T \lesssim 70$, which covers the range of Fig. 3.
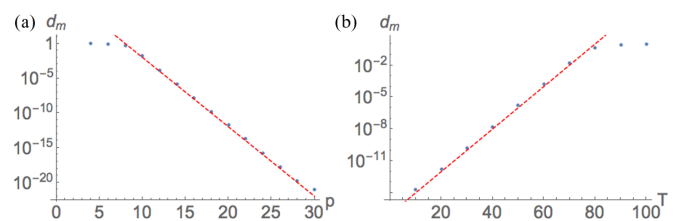


FIG. 5. Errors due to the numerical precision in the time-reversal test with no added noise. (a) Dependence of the median of the distances $d_{\rm m}$ on the numerical precision $p$, where $10^{-p}$ is the relative precision of the floating number. By conducting the time-reversal test with no noise, we obtain the distribution of the Euclidean distance between the initial state and the pulled-back state in phase space, from which we calculate their median. The numerical precision is varied with the evolution time held fixed at $T = 40$. (b) Dependence of the median of the distances $d_{\rm m}$ on the evolution time $T$. The evolution time is varied with the numerical precision fixed at $p = 16$, i.e., the double precision. The red dashed line in each panel represents the estimate of errors given by Eq. (A1). The unit of time is determined by the ratio of the extension of the stadium ($R$ and $L$) to the velocity of the billiard, both of which are set to unity.

## APPENDIX B: CONJECTURE FOR A GENERIC CLOSED CHAOTIC HAMILTONIAN SYSTEM

### 1. Rényi-0 divergence and fractal theory

In the main text, the box-counting dimension $d$ is determined from the scaling law of the Rényi-0 divergence:

$$D_0\big(\rho\|\mathcal{C}_l[\rho]\big) = d_C \ln(l/l_0),$$
$$d_C = d_E - d, \tag{B1}$$

where $d_E$ is the dimension of the embedding space and $l_0$ is the smallest length scale. Since the box-counting dimension is one of the fractal dimensions, we invoke fractal theory to conduct quantitative analyses. From a viewpoint of rigorous mathematics, fractals should have an infinitesimal structure ($l_0 \to 0$). However, from a viewpoint of physics, any fractal generated by a finite number of operations should have a nonzero (albeit extremely small) $l_0$. We refer to fractals with an infinitesimal structure as mathematical fractals and those with a nonzero minimal length scale as physical fractals. In this section, we prove Eq. (B1) for physical fractals on the basis of some natural assumptions.

First, we review how the fractal dimensions are mathematically defined. Let $S$ be a subset in the $d_E$-dimensional Euclidean space. For a fixed $a > 0$, we consider a hypercubic lattice with spacing $a$, which partitions the space into $d_E$-dimensional hypercubes (or "boxes") of volume $a^{d_E}$. We denote by $N(a)$ the number of boxes that have a nonzero overlap with $S$. The *box-counting dimension* $d_{box}$ is then defined as

$$d_{box} = -\lim_{a\to+0} \frac{\ln N(a)}{\ln a}. \tag{B2}$$

Given the probability distribution $\pi$ supported on $S$, it is possible to obtain generalized fractal dimensions [24] including the *correlation dimension*. For this purpose, we cover the fractal $S$ with $N(a)$ boxes, which are labeled with integers from 1 to $N(a)$. Let $\pi_i(a)$ be the probability with which a sample taken from $\pi$ belongs to the $i$th box. The correlation dimension $d_{cor}$ is then defined as

$$d_{cor} = \lim_{a\to+0} \frac{\ln C(a)}{\ln a}, \tag{B3}$$

where

$$C(a) = \sum_{i=1}^{N(a)} [\pi_i(a)]^2. \tag{B4}$$

Although $d_{box} = d_{cor}$ does not hold in general, it does in many practical cases. For instance, we may assume $\pi$ to be uniform in the sense that the relation

$$\pi_i(a) = \frac{1}{N(a)} + o\left(\frac{1}{N(a)}\right) \quad [i = 1, \ldots, N(a)] \tag{B5}$$

holds for all $a > 0$. In this case, we obtain $C(a) = 1/N(a) + o(1/N(a))$, and the two fractal dimensions coincide with each other. As we discuss later, the assumption of uniformity is consistent with the time-reversal test discussed in the main text. Henceforth, we assume that $\pi$ has a unique fractal dimension $d = d_{box} = d_{cor}$.

It is also known that the function $C(a)$ in Eq. (B3) can be replaced by the integral:

$$C'(a) = \iint \exp\left(-\frac{\|\gamma' - \gamma\|^2}{2a^2}\right) \pi(d\gamma)\pi(d\gamma'). \tag{B6}$$

Both $C(a)$ and $C'(a)$ roughly estimate the probability that two samples $\gamma$ and $\gamma'$, which are independently chosen from $\pi$, fall within the distance of $a$.

Now, we consider a physical fractal that reduces to a mathematical fractal $S$ as long as we consider a length scale larger than $l_0$. We define $S'$ to be the union of $N(l_0)$ boxes covering $S$. The length scale $l_0$ characterizes the smallest structure in the physical fractal, with which $S'$ can be considered to be the same as the mathematical fractal $S$.

Furthermore, we consider a uniform probability distribution over $S'$. We note that the uniformity of the distribution also holds for the final state $\rho_T$ in the time-reversal test, since the initial state is uniform and the dynamics is volume preserving. The probability density function $\rho$ can be written as

$$\rho(\gamma) = \frac{\mathbf{1}_{S'}(\gamma)}{l_0^{d_E} N(l_0)}, \tag{B7}$$

where $\mathbf{1}_{S'}$ is the indicator function of $S'$, which returns 1 for $\gamma \in S'$ and 0 for $\gamma \notin S'$.

The Rényi-0 divergence $D_0(\rho\|\mathcal{C}_l[\rho])$ can be computed as

$$D_0\big(\rho\|\mathcal{C}_l[\rho]\big) = -\ln \int_{\rho(\gamma)>0} \mathcal{C}_l[\rho](\gamma) d\gamma = -\ln \int \mathcal{C}_l[\rho](\gamma)\mathbf{1}_{S'}(\gamma) d\gamma \tag{B8}$$

$$= -\ln \iint \frac{1}{\left(\sqrt{2\pi}l\right)^{d_E}} \exp\left(-\frac{\|\gamma' - \gamma\|^2}{2l^2}\right) \rho(\gamma')\mathbf{1}_{S'}(\gamma) d\gamma d\gamma'. \tag{B9}$$

We further use (B7) to obtain

$$D_0\big(\rho\|\mathcal{C}_l[\rho]\big) = -\ln \iint \frac{l_0^{d_E} N(l_0)}{\left(\sqrt{2\pi}l\right)^{d_E}} \exp\left(-\frac{\|\gamma' - \gamma\|^2}{2a^2}\right) \rho(\gamma)\rho(\gamma') d\gamma d\gamma' \tag{B10}$$

$$= d_E \ln\left(\sqrt{2\pi}\frac{l}{l_0}\right) - \ln N(l_0) - \ln \iint \exp\left(-\frac{\|\gamma' - \gamma\|^2}{2l^2}\right) \rho(\gamma)\rho(\gamma') d\gamma d\gamma'. \tag{B11}$$

The integral in (B11) coincides with $C'(l)$ if we replace the measure $\rho(\gamma)d\gamma$ by $\pi(d\gamma)$. This replacement is justified when $l$ is sufficiently larger than $l_0$, because the difference between $\rho(\gamma)d\gamma$ and $\pi(d\gamma)$ enters in significantly only over a length scale below $l_0$. Therefore, for $1 \gg l \gg l_0$, the Rényi-0 divergence is evaluated as

$$D_0\big(\rho\,\|\,\mathcal{C}_l[\rho]\big) \approx d_{\mathrm{E}} \ln\left(\sqrt{2\pi}\,\frac{l}{l_0}\right) - \ln N(l_0) - \ln C'(l) \tag{B12}$$

$$\approx d_{\mathrm{E}} \ln(l/l_0) + d_{\mathrm{box}} \ln l_0 - d_{\mathrm{cor}} \ln l. \tag{B13}$$

Here, the approximations $\ln N(l_0) \approx -d_{\mathrm{box}} \ln l_0$ and $\ln C'(l) \approx \ln C(l) \approx d_{\mathrm{cor}} \ln l$ follow from the definition of the fractal dimensions (B2) and (B3). We obtain the desired relation (B1) by substituting $d = d_{\mathrm{box}} = d_{\mathrm{cor}}$ into (B13).

### 2. Time evolution of the Rényi-0 divergence in a Hamiltonian system with a generic dimension

We here consider a generic isolated time-reversal-invariant chaotic Hamiltonian system and argue that the Rényi-0 divergence shows a linear growth reminiscent of the Kolmogorov-Sinai theory.

Let $M$ be the phase space of the Hamiltonian system. Under the constraint of energy conservation, $M$ has the dimension of $d_{\mathrm{E}} = 2d_{\mathrm{S}} - 1$, with $d_{\mathrm{S}}$ being the spatial dimension, and can therefore be represented by some local coordinates $(x_1, \ldots, x_{2d_{\mathrm{S}}-1})$. In addition, we assume that the support $S_0$ of the initial state $\rho_0$ be sufficiently small. We require that the spread of $S_0$ is comparable in every direction, barring the possibility in which the dimension in one particular direction is much larger than that in any other direction. We exclude the case in which the extent of $S_0$ in one direction is by far larger than that in the other directions. Cubes and spheres satisfy this condition.

Following the procedure of the time-reversal test, we sample a phase-space point $\gamma_0$ from the initial state $\rho_0$, which evolves into $\gamma_T = \Phi_T(\gamma_0)$, according to the time evolution $\Phi_T : M \to M$ over time $T$. We may use different local coordinates to represent $\gamma_0$ and $\gamma_T$ as $d_{\mathrm{E}}$ vectors; we describe them as $\boldsymbol{\gamma}_0 = (x_1, \ldots, x_{2d_{\mathrm{S}}-1})$ and $\boldsymbol{\gamma}_T = (x'_1, \ldots, x'_{2d-1})$.

The behavior of the time evolution $\Phi_T$ near $\boldsymbol{\gamma}_0$ is linearly approximated as

$$\Phi_T(\boldsymbol{\gamma}_0 + \delta\boldsymbol{\gamma}) \approx \boldsymbol{\gamma}_T + \mathbf{J}_T \cdot \delta\boldsymbol{\gamma} \quad (\|\delta\boldsymbol{\gamma}\| \to 0), \tag{B14}$$

where $\mathbf{J}_T = \mathbf{J}_T(\boldsymbol{\gamma}_0)$ is a $(2d_{\mathrm{S}}-1) \times (2d_{\mathrm{S}}-1)$ matrix defined as

$$[\mathbf{J}_T(\boldsymbol{\gamma}_0)]_{j,k} = \frac{\partial x'_j}{\partial x_k}. \tag{B15}$$

The singular values $M_1 \leqslant \cdots \leqslant M_{2d_{\mathrm{S}}-1}$ of the matrix $\mathbf{J}_T$ can be obtained as the eigenvalues of $(\mathbf{J}_T^{\mathrm{t}}\mathbf{J}_T)^{1/2}$, where t denotes transposition. By appropriately changing the coordinate systems $(x_1, \ldots, x_{2d_{\mathrm{S}}-1})$ and $(x'_1, \ldots, x'_{2d_{\mathrm{S}}-1})$, we may regard $\mathbf{J}_T$ as a diagonal matrix with diagonal entries $M_1, \ldots, M_{2d_{\mathrm{S}}-1}$, where (B14) can be rewritten as

$$\Phi_T\big(x_1 + \delta x_1, \ldots, x_{2d_{\mathrm{S}}-1} + \delta x_{2d_{\mathrm{S}}-1}\big)$$
$$\approx \big(x'_1 + M_1\delta x_1, \ldots, x'_{2d_{\mathrm{S}}-1} + M_{2d_{\mathrm{S}}-1}\delta x_{2d_{\mathrm{S}}-1}\big). \tag{B16}$$

That is, the $j$th component is rescaled by $M_j$ through the time evolution.

As $T$ goes to infinity, the singular values $M_j = M_j(\gamma_0, T)$ are governed by the Lyapunov exponents $\Lambda_1 \leqslant \cdots \leqslant \Lambda_{2d_{\mathrm{S}}-1}$ as

$$M_j(\gamma_0, T) \sim e^{\Lambda_j T}. \tag{B17}$$

According to Oseledets's theorem [25], in a volume-preserving dynamical system, the limit

$$\mathbf{L} = \lim_{T \to \infty} (\mathbf{J}_T^{\mathrm{t}}\mathbf{J}_T)^{1/2T} \tag{B18}$$

exists for almost all $\gamma_0$. The eigenvalues of $\mathbf{L} = \mathbf{L}(\gamma_0)$ are independent of the initial state $\gamma_0$, and the logarithms thereof give the Lyapunov exponents.

We may choose the coordinate system $(x_1, \ldots, x_{2d_{\mathrm{S}}-1})$ such that $\mathbf{L}$ becomes diagonal, with which the linear approximation (B16) holds for $M_j \sim e^{\Lambda_j T}$. Although the coordinate system $(x_1, \ldots, x_{2d_{\mathrm{S}}-1})$ is determined locally on $\gamma_0$, these local coordinates can be integrated to form one coordinate system [26]. In this coordinate system, the approximation in (B16) is consistent with the degree of convergence of the limit (B18).

We also note that a similar analysis on the backward time evolution $\Phi_T^{-1} = \Phi_{-T}$ leads to the reversed Lyapunov exponents $-\Lambda_{2d_{\mathrm{S}}-1} \leqslant \cdots \leqslant -\Lambda_1$. The time-reversal invariance of the Hamiltonian system requires $\Lambda_j = -\Lambda_{2d_{\mathrm{S}}-j}$, and therefore the spectrum of Lyapunov exponents satisfies the following conditions:

$$-|\Lambda_1| \leqslant \cdots \leqslant -|\Lambda_{d_{\mathrm{S}}-1}| \leqslant |\Lambda_{d_{\mathrm{S}}}| \leqslant |\Lambda_{d_{\mathrm{S}}-1}| \leqslant \cdots \leqslant |\Lambda_1| \tag{B19}$$

and $\Lambda_{d_{\mathrm{S}}} = 0$.

Let us now consider the time-reversal test, where $\boldsymbol{\gamma}_T$ is perturbed into $\boldsymbol{\gamma}'_T = (x'_1 + \delta x'_1, \ldots, x'_{2d_{\mathrm{S}}-1} + \delta x'_{2d_{\mathrm{S}}-1})$. The disturbances $\delta x'_1, \ldots, \delta x'_{2d_{\mathrm{S}}-1}$ are independently subject to Gaussian distributions of zero mean and the standard deviation $l$. Let the spread of $S_0$, which constitutes the support of the initial state, be $w$ in every direction. To estimate the Rényi-0 divergence, we roughly approximate $S_0$ by a hyper-rectangular region as

$$S_0 \sim \big\{(y_1, \ldots, y_{2d_{\mathrm{S}}-1})\big| a_j \leqslant y_j \leqslant b_j\big\}. \tag{B20}$$

Here, the interval $[a_j, b_j]$ contains $x_j$, and its width $b_j - a_j$ is comparable to $w$. With the linear approximation in (B16), the support $S_T$ of the time-evolved state $\rho_T$ assumes the form of

$$S_T \sim \big\{(y'_1, \ldots, y'_{2d_{\mathrm{S}}-1})\big| M_j a_j \leqslant y'_j \leqslant M_j b_j\big\}. \tag{B21}$$

Therefore the probability of $\gamma'_T$ lying in $S_T$ can be estimated as

$$\mathbf{P}[\rho_T(\gamma'_T) > 0 \mid \gamma_0]$$
$$\approx \prod_{j=1}^{2d_{\mathrm{S}}-1} \frac{1}{\sqrt{2\pi}l} \int_{M_j a_j}^{M_j b_j} \exp\left[-\frac{(y'_j - x'_j)^2}{2l^2}\right] dy'_j. \tag{B22}$$

The behavior of integrals in (B22) depends on the ratio $M_j w/l$. If $M_j w/l \gtrsim 1$, this integral becomes of the order of unity, since the interval $[M_j a_j, M_j b_j]$ contains a significant part of the Gaussian peak. When $M_j w/l \ll 1$, the integral
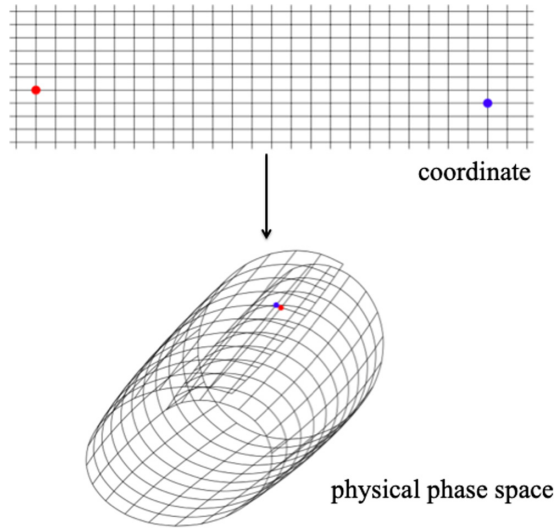
FIG. 6. Schematic illustration of a coordinate in phase space. While the blue and red points are located far from each other in this coordinate representation, they are close to each other in the physical phase space.

contains only a small fraction of the Gaussian peak and therefore remains of the order of $M_j w/l$. From these observations, we obtain

$$\frac{1}{\sqrt{2\pi}l} \int_{M_j a_j}^{M_j b_j} \exp\left[-\frac{(y_j' - x_j')^2}{2l^2}\right] dy_j' \sim \min\{1, M_j w/l\}. \tag{B23}$$

Therefore the conditional probability in (B21) is estimated to be

$$\mathbf{P}[\rho_T(\gamma_T') > 0 \mid \gamma_0] \sim \prod_{j=1}^{2d-1} \min\{1, M_j w/l\} \tag{B24}$$

$$= (w/l)^\alpha M_1 \cdots M_\alpha, \tag{B25}$$

where we denote by $\alpha$ the greatest integer that satisfies the scale separation $M_\alpha w/l \ll 1$. Although this criterion for the

scale separation may be somewhat ambiguous for finite $T$, the ambiguity disappears as $T$ goes to infinity.

Now we employ the Lyapunov estimates in (B17) to obtain

$$\mathbf{P}[\rho_T(\gamma_T') > 0 \mid \gamma_0] \sim (w/l)^\alpha \exp[(\Lambda_1 + \cdots + \Lambda_\alpha)T]. \tag{B26}$$

With sufficiently large $T$, the condition for the scale separation of $\alpha$ is independent of the initial state $\gamma_0$ and can be written as

$$\Lambda_\alpha < \frac{1}{T} \ln \frac{l}{w} < \Lambda_{\alpha+1}. \tag{B27}$$

Therefore we may remove the condition on $\gamma_0$ from the probability and finally obtain

$$D_0 = -\ln \mathbf{P}[\rho_T(\gamma_T') > 0] \tag{B28}$$

$$\approx \alpha \ln l - \alpha \ln w - (\Lambda_1 + \cdots + \Lambda_\alpha)T. \tag{B29}$$

Hence we have a linear relation between $D_0$ and $\ln l$ (and also $T$) in the regime of Eq. (B27). In particular, when $T$ tends to infinity, the middle term of inequality (B27) becomes zero, and therefore $\alpha$ is the number of negative Lyapunov exponents. When $l < w$, this number is generally $d_S - 1$ in a time-reversal-invariant system with Lyapunov exponents (B19), unless some exponents other than $\Lambda_d$ happen to be zero. Then, Eq. (B29) reduces to

$$D_0 \approx (d_S - 1)\ln(l/w) + hT, \tag{B30}$$

where $h$ is the sum of the positive Lyapunov exponents and coincides with the Kolmogorov-Sinai entropy according to the Pesin formula.

The linear relation (B29) breaks down for extremely large $T$ such that the Poincaré recurrence occurs. In the argument given in this section, the long-time evolution affects the mapping from the phase space to the coordinate system $(x_1, \ldots, x_{2d_S-1})$. Although the perturbed state $\gamma_T'$ appears different from the unperturbed state $\gamma_T$ in the coordinate space, it is possible that they are close in physical phase space as illustrated in Fig. 6. In this case, we wrongly judge that $\gamma_T'$ is not in the support of $\rho_T$ since it is located far from $\gamma_T$ in the coordinate space and make an incorrect approximation of $D_0$. This type of error can emerge when the spread of $S_T$ is at least larger than the system size, but numerical results show that Eq. (B29) holds beyond this time scale.

[1] L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen, Sitzungsber. Kais. Akad. Wiss. Wien, Math.-Naturwiss. Kl. Abt. II **66**, 275 (1872).

[2] J. Loschmidt, Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft, Sitzungsber. Kais. Akad. Wiss. Wien, Math.-Naturwiss. Kl. Abt. II **73**, 128 (1876).

[3] L. Boltzmann, Bemerkungen über einige Probleme der mechanischen Wärmetheorie, Sitzungsber. Kais. Akad. Wiss. Wien, Math.-Naturwiss. Kl. Abt. II **75**, 62 (1877).

[4] J. L. Lebowitz, Boltzmann's entropy and time's arrow, Phys. Today **46**( 9), 32 (1993).

[5] J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **57**, 617 (1985).

[6] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge University Press, New York, 1998).

[7] C. Jarzynski, Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale, Annu. Rev. Condens. Matter Phys. **2**, 329 (2011).

[8] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

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

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

[11] T. Hatano and S.-i. Sasa, Steady-State Thermodynamics of Langevin Systems, Phys. Rev. Lett. **86**, 3463 (2001).

[12] U. Seifert, Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem, Phys. Rev. Lett. **95**, 040602 (2005).

[13] M. Esposito and C. Van den Broeck, Three Detailed Fluctuation Theorems, Phys. Rev. Lett. **104**, 090601 (2010).

[14] D. L. Shepelyansky, Some statistical properties of simple classically stochastic quantum systems, Phys. D (Amsterdam) **8**, 208 (1983).

[15] S. Adachi, M. Toda, and K. Ikeda, Quantum-Classical Correspondence in Many-Dimensional Quantum Chaos, Phys. Rev. Lett. **61**, 659 (1988).

[16] R. A. Jalabert and H. M. Pastawski, Environment-Independent Decoherence Rate in Classically Chaotic Systems, Phys. Rev. Lett. **86**, 2490 (2001).

[17] T. van Erven and P. Harremoës, Rényi divergence and Kullback-Leibler divergence, IEEE Trans. Inf. Theory **60**, 3797 (2014).

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

[19] P. R. Halmos, *Measure Theory* (Springer, New York, 1974).

[20] L. A. Bunimovich, On the ergodic properties of nowhere dispersing billiards, Commun. Math. Phys. **65**, 295 (1979).

[21] G. M. Zaslavsky, *Hamiltonian Chaos & Fractional Dynamics* (Oxford University Press, New York, 2005).

[22] V. Latora and M. Baranger, Kolmogorov-Sinai Entropy Rate versus Physical Entropy, Phys. Rev. Lett. **82**, 520 (1999).

[23] P. Gaspard, Entropy production in open volume-preserving systems, J. Stat. Phys. **88**, 1215 (1997).

[24] H. G. E. Hentschel and I. Procaccia, The infinite number of generalized dimensions of fractals and strange attractors, Phys. D (Amsterdam) **8**, 435 (1983).

[25] V. I. Oseledets, A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems, Trans. Moscow Math. Soc. **19**, 197 (1968).

[26] A. Hammerlindl, Integrability and Lyapunov exponents, J. Mod. Dyn. **5**, 107 (2011).