Supplemental Material for "Thermodynamic Uncertainty Relation for Arbitrary Initial States"

Kangqiao Liu,[1,*] Zongping Gong,[1] and Masahito Ueda[1,2,3]
[1]Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[2]RIKEN Center for Emergent Matter Science, 2-1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan
[3]Institute for Physics of Intelligence, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
(Dated: August 3, 2020)

I. MINIMAL MODEL FOR EQUILIBRIUM STEADY STATES

In our model, the initial state is $\mathbf{P}(0) = [p, 1-p]^{\mathbb{T}}$ and two transition rates satisfy the local detailed balance condition $r(0,1) = e^{-\beta} r(1,0)$. For simplicity, we denote $r(0,1) = a$ and $r(1,0) = b$. Then the transition rate matrix reads

$$\mathcal{R} = \begin{bmatrix} -a & b \\ a & -b \end{bmatrix}. \tag{S1}$$

In terms of the transition rate matrix, the final state can be related to the initial state as

$$\mathbf{P}(T) = e^{\mathcal{R}T}\mathbf{P}(0) = [P_0(T), P_1(T)]^{\mathbb{T}}, \tag{S2}$$

where

$$P_0(T) = \frac{e^{-(a+b)T}}{a+b}[ap + b(-1+p+e^{(a+b)T})] \tag{S3}$$

and $P_1(T) = 1 - P_0(T)$. The ensemble-averaged accumulated current $\langle J\rangle$ and its variance Var$[J]$ can be calculated by means of full counting statistics [1–3]. By introducing a counting field χ in the rate matrix, the characteristic function should read

$$\mathcal{Z}(\chi) = [1,1]e^{\mathcal{R}(\chi)T}\mathbf{P}(0), \tag{S4}$$

where

$$\mathcal{R}(\chi) = \begin{bmatrix} -a & be^{-i\chi} \\ ae^{i\chi} & -b \end{bmatrix}. \tag{S5}$$

After a straightfoward calculation, the characteristic function is given by

$$\begin{aligned}\mathcal{Z}(\chi) =&\frac{1}{a+b}\big[(b+ae^{-(a+b)T})p + (a+be^{-(a+b)T})(1-p)\\ &+ be^{-(a+b)T}e^{-i\chi}(-1+e^{(a+b)T})(1-p)\\ &+ ae^{-(a+b)T}e^{i\chi}(-1+e^{(a+b)T})p\big].\end{aligned} \tag{S6}$$

The mean and variance of the current can be obtained from the first and second derivatives of $\ln\mathcal{Z}(\chi)$ with respect to $i\chi$ at $\chi = 0$ as

$$\langle J\rangle = \frac{b(-1+p)+ap}{a+b}(1-e^{-(a+b)T}), \tag{S7}$$

$$\mathrm{Var}[J] = \frac{e^{-2(a+b)T}}{(a+b)^2}\big[-[b(-1+p)+ap]^2 + e^{2(a+b)T}[ab+(a+b)^2p-(a+b)^2p^2] + e^{(a+b)T}[b(-a+b)-(a+b)(a+3b)p+2(a+b)^2p^2]\big]. \tag{S8}$$

The total entropy production σ involves two contributions. One is the Shannon entropy change in the system:

$$\sigma_{\mathrm{S}} = S(T) - S(0), \tag{S9}$$

where $S(t) = -[P_0(t)\ln P_0(t) + P_1(t)\ln P_1(t)]$. The other is the entropy production in the heat bath $\sigma_{\mathrm{B}} = -\beta\langle J\rangle$. Thus all terms involved in the TUR are analytically obtained. In the simulation, we set $p = 0.3$, $r(0,1) = 1$ and $r(1,0) = 2$.

In Fig. S1, the Q values of our inequality, the original TUR and the GTUR are presented. We find that while the conventional TUR and the GTUR break down, our bound remains valid, which is consistent with Fig. 2(b) in the main text.

* kqliu@cat.phys.s.u-tokyo.ac.jp

II. MINIMAL MODEL FOR NONEQUILIBRIUM STEADY STATES

In this model, the only modification from the previous example is that there are 4 transition rates satisfying 2 local detailed balance conditions: $r^h(1,0) = e^{\beta_h}r^h(0,1)$ and $r^c(1,0) = e^{\beta_c}r^c(0,1)$. We introduce the following shorthand notations: $r^h(0,1) = a_h$, $r^h(1,0) = b_h$, $r^c(0,1) = a_c$ and $r^c(1,0) = b_c$. The rate matrix is then given by

$$\mathcal{R}(\chi) = \begin{bmatrix} -a_c - a_h & b_c + b_h e^{-i\chi} \\ a_c + a_h e^{i\chi} & -b_c - b_h \end{bmatrix}. \tag{S10}$$

The instantaneous hot current is $j^h(t) = P_0(t)r^h(0,1) - P_1(t)r^h(1,0)$. We start from an initial state where $j^h(0)$ van-

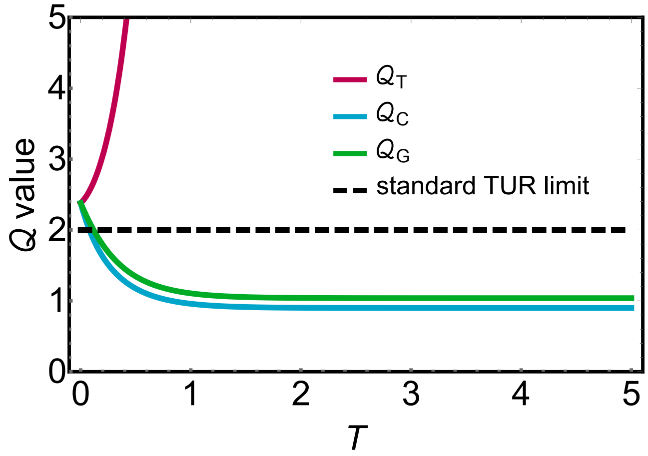

FIG. S1. $\mathcal{Q}$ values defined in the main text versus time T. The red solid curve is $\mathcal{Q}_{\rm T}$ in our TUR. The blue and green curves show the $\mathcal{Q}$ values of the conventional TUR and the GTUR, respectively. As in Fig. 2(b) in the main text, our inequality always holds, whereas the original TUR and the GTUR are violated.

ishes, i.e.,

$$\mathbf{P}(0) = \left[\frac{1}{1+e^{-\beta_h}}, \frac{e^{-\beta_h}}{1+e^{-\beta_h}}\right]^{\mathbb{T}} \equiv [\alpha, \beta]^{\mathbb{T}}. \qquad \text{(S11)}$$

The characteristic function can explicitly be calculated in terms of

$$D(\chi) = \sqrt{4mn + (a-b)^2}, \qquad \text{(S12)}$$

with $a = -a_c - a_h$, $m(\chi) = b_c + b_h e^{-i\chi}$, $n(\chi) = a_c + a_h e^{i\chi}$ and $b = -b_c - b_h$, as

$$\mathcal{Z}(\chi) = e^{\frac{a+b}{2}\tau}\{(\alpha+\beta)\cosh(\frac{D}{2}\tau) + \frac{1}{D}[(a-b+2n)\alpha - (a-b-2m)\beta]\sinh(\frac{D}{2}\tau)\}. \qquad \text{(S13)}$$

Similarly, the first and second derivatives of the characteristic function give analytic expressions of the mean and the variance. The entropy production of the bath is given by $\sigma_{\rm B} = \beta_h\langle J^h\rangle + \beta_c\langle J^c\rangle$, where the mean cold and hot currents can be calculated similarly to the single-bath case (see Sec. I). We set $\beta_h = 1$, $\beta_c = 1.5$, $r^h(0,1) = r^c(0,1) = 1$ in our simulation.

In Fig. S2, we see that while all TURs are satisfied, ours is tightest, which is consistent with Fig. 2(d) in the main text.

III. APPLICATION TO FEEDBACK CONTROL

Let us verify inequality (8) in the main text for the two-level minimal model. The only difference from the minimal model for an NESS is that there is a two-state meter that probes the system and performs feedback control. As a result of two control protocols and two channels, there are 8 transition rates which satisfy 4 local detailed balance relations. The current is again defined as the heat transferred from the system to the hot bath, and its fluctuation is calculated from full counting statistics.

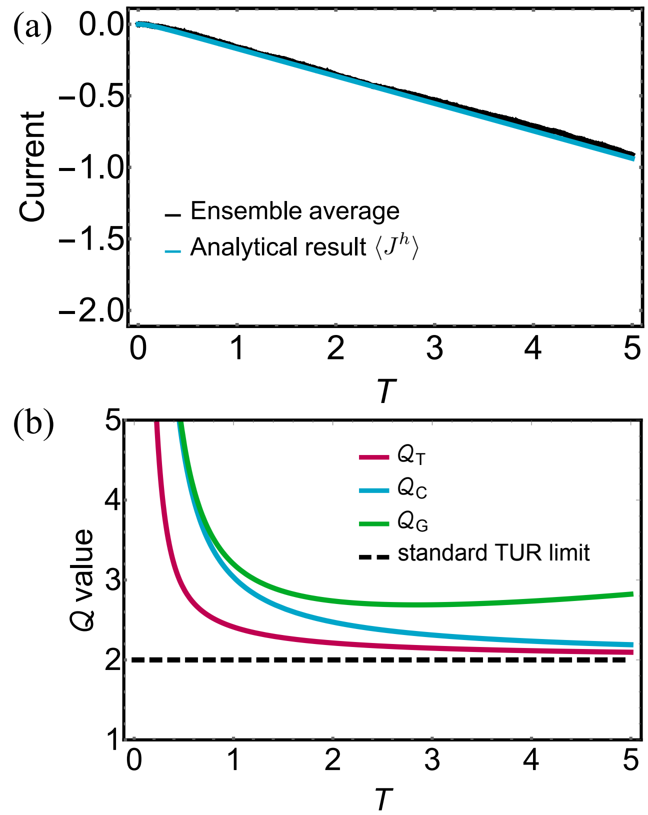

FIG. S2. (a) Time dependence of the current obtained from the average over 10000 trajectories of Monte Carlo simulations. Superimposed is the analytic result calculated from full counting statistics. An excellent agreement confirms the correctness of our calculation. (b) $\mathcal{Q}$ values for the conventional TUR (blue), the GTUR (green) and our TUR (red). We find that while all TURs are valid, ours is the tightest, which is consistent with Fig. 2(d) in the main text.

The meter is initially set to be in state 0. Let the probability of the system being found in state $u(d)$ be $p_u^{\rm ini}(p_d^{\rm ini})$. The meter and the system are assumed to be uncorrelated initially. Thus we have

$$\mathbf{p}_{\rm s}^{\rm ini} = [p_u^{\rm ini}, p_d^{\rm ini}]^{\mathbb{T}}, \qquad \text{(S14)}$$
$$\mathbf{p}_{\rm m}^{\rm ini} = [1, 0]^{\mathbb{T}}, \qquad \text{(S15)}$$
$$\mathbf{p}^{\rm ini} = [p_u^{\rm ini}, p_d^{\rm ini}, 0, 0]^{\mathbb{T}} = \mathcal{C}\mathbf{p}_{\rm s}^{\rm ini}, \qquad \text{(S16)}$$

where $\mathcal{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ is the projection operator for the composite state.

Now suppose that the state of the system is measured. If the system is found to be in state u, the conditional probability of the meter being found in state 0 (1) is $p_{0|u} = 1-p$ ($p_{1|u} = p$); if the system is found to be in state d, the conditional probability of the meter being found in state 0 (1) is $p_{0|d} = q$ ($p_{1|d} = 1-q$). Here p and q denote measurement errors. The post-measurement joint state is then given by

$$\mathbf{p}_M = [(1-p)p_u^{\rm ini}, qp_d^{\rm ini}, pp_u^{\rm ini}, (1-q)p_d^{\rm ini}]^{\mathbb{T}} = \mathcal{M}\mathbf{p}^{\rm ini} = \mathcal{M}\mathcal{C}\mathbf{p}_{\rm s}^{\rm ini}, \tag{S17}$$

where $\mathcal{M} = \begin{bmatrix} 1-p & 0 & 0 & 0 \\ 0 & q & 0 & 0 \\ p & 0 & 0 & 0 \\ 0 & 1-q & 0 & 0 \end{bmatrix}$ gives the effect of measurement.

Feedback control is performed according to the measurement outcome. If the meter is found to be in state 0, then we change the energy gap between u and d to $\Delta_0 = \Delta(1-w)$; if the meter is found to be in state 1, we change it to $\Delta_1 = \Delta(1+w)$. The transition rate matrix is then given by

$$\mathcal{R} = \mathcal{R}^h + \mathcal{R}^c = \begin{pmatrix} \mathcal{R}_0^h & \mathbf{0} \\ \mathbf{0} & \mathcal{R}_1^h \end{pmatrix} + \begin{pmatrix} \mathcal{R}_0^c & \mathbf{0} \\ \mathbf{0} & \mathcal{R}_1^c \end{pmatrix}, \tag{S18}$$

where $\mathcal{R}_0^h = \begin{pmatrix} -a_h & b_h \\ a_h & -b_h \end{pmatrix}$ and $\mathcal{R}_1^h = \begin{pmatrix} -c_h & d_h \\ c_h & -d_h \end{pmatrix}$, with $a_h = r^h(0u, 0d)$, $b_h = r^h(0d, 0u)$, $a_c = r^c(0u, 0d)$, $b_c = r^c(0d, 0u)$, $c_h = r^h(1u, 1d)$, $d_h = r^h(1d, 1u)$, $c_c = r^c(1u, 1d)$ and $d_c = r^c(1d, 1u)$. These 8 rates satisfy 4 local detailed balance conditions:

$$\frac{a_h}{b_h} = e^{\beta_h \Delta_0}, \ \frac{a_c}{b_c} = e^{\beta_c \Delta_0}, \ \frac{c_h}{d_h} = e^{\beta_h \Delta_1}, \ \frac{c_c}{d_c} = e^{\beta_c \Delta_1}. \tag{S19}$$

After the feedback control, the system evolves according to the master equation for an interaction time τ:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathcal{R}\mathbf{p}. \tag{S20}$$

At the end of one cycle, the marginal state of the system is

$$\mathbf{p}_{\rm s}(\tau) = \mathcal{P}^{\rm s}\mathbf{p}(\tau) = \mathcal{P}^{\rm s} e^{\mathcal{R}\tau}\mathcal{M}\mathcal{C}\mathbf{p}_{\rm s}^{\rm ini} \equiv \mathcal{T}\mathbf{p}_{\rm s}^{\rm ini}, \tag{S21}$$

where $\mathcal{P}^{\rm s} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ is the projector onto the marginal state of the system. Finally, we reset the state of the meter to the default state 0. After many cycles, the system will reach a stroboscopic steady state:

$$\mathcal{T}\mathbf{p}_{\rm s,sss}^{\rm ini} = \mathbf{p}_{\rm s,sss}^{\rm ini}. \tag{S22}$$

Thus the stroboscopic steady state is given by the eigenvector associated with eigenvalue 1 of matrix $\mathcal{T}$. It can be calculated explicitly. As can be seen from Eq. (S18), the sectors with $m = 0$ and 1 are completely isolated from each other, and full counting statistics can be calculated independently for each sector in a manner similar to what is done in the second example discussed in Sec. II (see Eq. (S13)). In our simulation, we set $r^h(0u, 0d) = 3$, $r^c(0u, 0d) = 2$, $r^h(1u, 1d) = 2$, $r^c(1u, 1d) = 1$, $\beta_c = 2$, $\beta_h = 1$, $p = q = 0$ and $\Delta = 1$.

As shown in Fig. S3(a), for different control parameters w, our bound always holds. We note that in this model, the current monotonically decreases, implying that $\mathcal{Q}_{\rm C} < \mathcal{Q}_{\rm T}$. Although for our chosen parameter set $\mathcal{Q}_{\rm C}$ does not surpass the

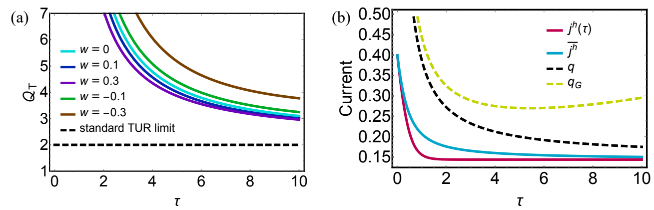

FIG. S3. (a) The solid curves show $\mathcal{Q}_{\rm T}$ for different values of the feedback control parameter w. The black dashed line shows the lower bound 2 for the conventional TUR limit. It can be seen that our bound is always satisfied. (b) Dependence of the current on the period τ of the feedback cycle for the control parameter $w = 0.3$. The red and blue solid curves show the final instantaneous current and the time-averaged one, respectively. The black and green dashed curves represent the current bounds associated with the conventional TUR and the GTUR. Both the final current and the time-averaged one decrease with τ as discussed in the main text.

bound 2, there could be a possible violation for a wider parameter range. In Fig. S3(b), we show the current as a function of the time period τ for the feedback control parameter $w = 0.3$. It can be seen that both the final current and the time-averaged one decrease with τ and that they are bounded from above by q and q_G. As stated in the main text, the decreasing current implies a possible violation in the conventional TUR. While the conventional TUR is not violated for the parameter range chosen here, it might break down for a wider range of the parameters.

IV. SOME DETAILS IN THE DERIVATION OF THE MAIN RESULT (6)

First, we derive Eq. (14) in the main text.

$$\begin{aligned} F(\theta) &= -\left\langle \frac{\partial^2}{\partial\theta^2} \ln \mathcal{P}_\theta[\omega] \right\rangle_\theta \\ &= \int_0^T dt \sum_{x\neq y,\nu} \Big[-P_\theta(x;t)\frac{\partial^2}{\partial\theta^2} r_\theta^\nu(x,y;t) \\ &\quad + P_\theta(x;t) r_\theta^\nu(x,y;t)\frac{\partial^2}{\partial\theta^2}\ln r_\theta^\nu(x,y;t)\Big] \\ &= \int_0^T dt \sum_{x\neq y,\nu} P_\theta(x;t) r_\theta^\nu(x,y;t)\left(\frac{\partial}{\partial\theta}\ln r_\theta^\nu(x,y;t)\right)^2. \end{aligned} \tag{S23}$$

Then, we prove that Eqs. (16) and (17) always have solutions. We first note that they take the following forms:

$$AX^2 + BY^2 = \frac{1}{2}(A-B)\ln\frac{A}{B}, \tag{S24}$$

$$AX - BY = A - B. \tag{S25}$$

We use Eq. (S25) to eliminate Y in Eq. (S24):

$$A(A+B)X^2 - 2A(A-B)X + (A-B)^2 - \frac{1}{2}(A-B)B\ln\frac{A}{B} = 0. \tag{S26}$$

The discriminant of this equation can be shown to be non-negative:

$$\Delta = 2AB(A+B)(A-B)\ln\frac{A}{B} - 4AB(A-B)^2 \\ \geq 4AB(A+B)(A-B)^2 - 4AB(A-B)^2 = 0, \quad \text{(S27)}$$

where the inequality $(a-b)\ln\frac{a}{b} \geq \frac{2(a-b)^2}{a+b}$ for $a, b > 0$ is used. Thus the equations always have a solution. Explicitly, it can be written as

$$\alpha^\nu_{xy} = X = \jmath^\nu_{xy} \pm \jmath^\nu_{xy}\sqrt{\frac{1-\jmath^\nu_{xy}}{1+\jmath^\nu_{xy}}\left(\frac{F^\nu_{xy}}{2\jmath^\nu_{xy}} - 1\right)}, \quad \text{(S28)}$$

where $\jmath^\nu_{xy}(t) \equiv \frac{j^\nu(x,y;t)}{t^\nu(x,y;t)}$ is the current normalized by the traffic rate $t^\nu(x,y;t) \equiv P(x;t)r^\nu(x,y) + P(y;t)r^\nu(y,x)$, and $F^\nu_{xy}(t) \equiv \ln\frac{P(x;t)r^\nu(x,y)}{P(y;t)r^\nu(y,x)}$ is the thermodynamic force at time t. In Eq. (S28), the time dependence is omitted for simplicity.

To prove Eq. (18), we have to prove another important relation. For small θ, we expand $\mathcal{R}_\theta(t)$ up to the first order in θ:

$$\mathcal{R}_\theta(t) = \mathcal{R} + \theta\mathcal{R}_1(t) + O(\theta^2). \quad \text{(S29)}$$

We use Eq. (17) to show that $\mathcal{R}_1(t)$ satisfies

$$[\mathcal{R}_1(t)\mathbf{P}(t)]_y = \sum_{x:x\neq y,\nu}\left[K^\nu_{xy}\alpha^\nu_{xy} - K^\nu_{yx}\alpha^\nu_{yx}\right] \\ = \sum_{x:x\neq y,\nu}(K^\nu_{xy} - K^\nu_{yx}) = [\mathcal{R}\mathbf{P}(t)]_y. \quad \text{(S30)}$$

We use this relation to prove Eq. (18). Instead of $\mathbf{P}_\theta(t)$, we consider its transformed quantity:

$$\tilde{\mathbf{P}}_\theta(t) \equiv e^{-\mathcal{R}t}\mathbf{P}_\theta(t). \quad \text{(S31)}$$

The time derivative of this quantity can be calculated as

$$\dot{\tilde{\mathbf{P}}}_\theta(t) = -\mathcal{R}\tilde{\mathbf{P}}_\theta(t) + e^{-\mathcal{R}t}\dot{\mathbf{P}}_\theta(t) \\ = -\mathcal{R}\tilde{\mathbf{P}}_\theta(t) + e^{-\mathcal{R}t}(\mathcal{R} + \theta\mathcal{R}_1(t))\mathbf{P}_\theta(t) \\ = \theta e^{-\mathcal{R}t}\mathcal{R}_1(t)e^{\mathcal{R}t}\tilde{\mathbf{P}}_\theta(t). \quad \text{(S32)}$$

Integrating this equation from 0 to t gives

$$\tilde{\mathbf{P}}_\theta(t) = \mathbf{P}(0) + \theta\int_0^t ds e^{-\mathcal{R}s}\mathcal{R}_1(s)\mathbf{P}_\theta(s) \\ = \mathbf{P}(0) + \theta\int_0^t ds e^{-\mathcal{R}s}\mathcal{R}_1(s)\mathbf{P}(s) + O(\theta^2). \quad \text{(S33)}$$

We use Eq. (S30) to obtain

$$\tilde{\mathbf{P}}_\theta(t) = \mathbf{P}(0) + \theta\int_0^t ds e^{-\mathcal{R}s}\mathcal{R}\mathbf{P}(s) + O(\theta^2) \\ = \mathbf{P}(0) + \theta t\mathcal{R}\mathbf{P}(0) + O(\theta^2). \quad \text{(S34)}$$

We thus arrive at

$$\mathbf{P}_\theta(t) = \mathbf{P}(t) + \theta t\dot{\mathbf{P}}(t) + O(\theta^2). \quad \text{(S35)}$$

V. ALTERNATIVE DERIVATION OF THE MAIN RESULT (6) FROM THE LARGE DEVIATION THEORY

In this alternative derivation, we employ the level 2.5 large deviation theory which is a slight modification of that used for periodically driven dynamics [4] and finite-time TUR [5]. As in Ref. [5], we consider a large ensemble of N copies of the dynamics. The joint probability distribution of empirical states $\tilde{P}(x;t)$ and empirical currents $\tilde{j}^\nu(x,y;t)$, which deviate from their typical values, should be exponentially small for large N as

$$P(\tilde{P}(t), \tilde{j}(t)) \asymp e^{-NI(\tilde{P}(t),\tilde{j}(t))}, \quad \text{(S36)}$$

where $\asymp$ denotes the asymptotic logarithmic equivalence [5]. The rate function is

$$I(\tilde{P}(t), \tilde{j}(t)) = \int_0^T dt L(\tilde{P}(t), \tilde{j}(t)) + D_{\rm KL}(\tilde{\mathbf{P}}(0)||\mathbf{P}(0)), \quad \text{(S37)}$$

where

$$D_{\rm KL}(\tilde{\mathbf{P}}(0)||\mathbf{P}(0)) \equiv \sum_x \tilde{P}(x;0)\ln\frac{\tilde{P}(x;0)}{P(x;0)} \quad \text{(S38)}$$

is the Kullback-Leibler divergence of the empirical initial distribution $\tilde{\mathbf{P}}(0)$ from the typical one $\mathbf{P}(0)$, and

$$L(\tilde{P}(t), \tilde{j}(t)) = \sum_{x>y,\nu}\Psi(\tilde{j}^\nu(x,y;t), j^\nu_{\tilde{P}}(x,y;t), a^\nu_{\tilde{P}}(x,y;t)), \quad \text{(S39)}$$

with

$$\Psi(j, \bar{j}, a) \equiv j\left(\text{arsinh}\frac{j}{a} - \text{arsinh}\frac{\bar{j}}{a}\right) - \sqrt{a^2 + j^2} - \sqrt{a^2 + \bar{j}^2}. \quad \text{(S40)}$$

Here the quantities in Eq. (S39) are defined as

$$\tilde{j}^\nu(x,y;t) \equiv \tilde{P}(x;t)\tilde{r}^\nu(x,y;t) - \tilde{P}(y;t)\tilde{r}^\nu(y,x;t), \quad \text{(S41)}$$
$$j^\nu_{\tilde{P}}(x,y;t) \equiv \tilde{P}(x;t)r^\nu(x,y) - \tilde{P}(y;t)r^\nu(y,x), \quad \text{(S42)}$$
$$a^\nu_{\tilde{P}}(x,y;t) \equiv 2\sqrt{\tilde{P}(x;t)\tilde{P}(y;t)r^\nu(x,y)r^\nu(y,x)}. \quad \text{(S43)}$$

It has been shown [4, 5] that the rate function in Eq. (S37) has a parabolic upper bound as

$$I(\tilde{P}(t),\tilde{j}(t)) \leq \int_0^T dt \sum_{x>y,\nu} \left[\frac{\tilde{j}^\nu(x,y;t)-j^\nu_{\tilde{P}}(x,y;t)}{2j^\nu_{\tilde{P}}(x,y;t)}\right]^2 \sigma^\nu_{\tilde{P}}(x,y;t) + D_{\rm KL}(\tilde{\mathbf{P}}(0)||\mathbf{P}(0)), \tag{S44}$$

where

$$\sigma^\nu_{\tilde{P}}(x,y;t) \equiv j^\nu_{\tilde{P}}(x,y;t) \ln \frac{\tilde{P}(x;t)r^\nu(x,y)}{\tilde{P}(y;t)r^\nu(y,x)}. \tag{S45}$$

The empirical time-integrated current is given by

$$J_d \equiv \int_0^T dt \sum_{x\neq y,\nu} \tilde{j}^\nu(x,y;t) d^\nu(x,y). \tag{S46}$$

To obtain the rate function of this empirical accumulated current J_d, the contraction principle should be employed under two constraints [4, 5]. One is the normalization condition for the empirical distribution $\sum_x \tilde{P}(x;t) = 1$. The other is a master equation obeyed by the empirical transition rates and the empirical states, i.e., $\dot{\tilde{P}}(x;t) = \sum_{y:y\neq x,\nu} \tilde{j}^\nu(y,x;t)$.

We let the empirical quantities be the same as the parametrized ones in the derivation via the Cramé-Rao inequality (see Eqs. (16) and (18) in the main text). They are written as

$$\tilde{r}^\nu(x,y;t) = r^\nu(x,y)e^{\theta\alpha^\nu_{xy}(t)}, \tag{S47}$$

$$\tilde{P}(x;t) = P(x;t) + \theta t\dot{P}(x;t) + O(\theta^2). \tag{S48}$$

This choice ensures that the above two constraints are satisfied. It can be shown that

$$\tilde{j}^\nu(x,y;t) = (1+\theta)j^\nu(x,y;t) + \theta t\dot{j}^\nu(x,y;t) + O(\theta^2), \tag{S49}$$

$$j^\nu_{\tilde{P}}(x,y;t) = j^\nu(x,y;t) + \theta t\dot{j}^\nu(x,y;t) + O(\theta^2), \tag{S50}$$

$$\sigma^\nu_{\tilde{P}}(x,y;t) = \sigma^\nu(x,y;t) + O(\theta), \tag{S51}$$

where

$$\sigma^\nu(x,y;t) \equiv j^\nu(x,y;t) \ln \frac{P(x;t)r^\nu(x,y)}{P(y;t)r^\nu(y,x)} \tag{S52}$$

is the typical entropy production rate for the transition between states x and y via channel ν. The relative entropy $D_{\rm KL}(\tilde{\mathbf{P}}(0)||\mathbf{P}(0))$ then vanishes. From the definition of the empirical accumulated current J_d Eq. (S46), θ is determined as

$$\theta = \frac{J_d - \langle J\rangle}{Tj(T)}. \tag{S53}$$

Finally, by contracting Eq. (S44), the bound on the rate function of J_d is given by

$$I(J_d) \leq \frac{1}{4}\left[\frac{J_d-\langle J\rangle}{Tj(T)}\right]^2 \sigma + O(\theta^3). \tag{S54}$$

Hence,

$$\mathrm{Var}[J] = \frac{1}{I''(\langle J\rangle)} \geq 2\frac{(Tj(T))^2}{\sigma}, \tag{S55}$$

which is the desired bound, i.e., Eq. (6) in the main text.

VI. DERIVATION OF THE DISCRETE-TIME TUR (11)

The parametrized path probability is given as

$$\mathcal{P}_\theta[\omega] = P_\theta(x_0)\prod_{i=1}^n A^{\nu_i}_\theta(x_i|x_{i-1};t_{i-1}), \tag{S56}$$

where the parametrized transition probability $A^{\nu_i}_\theta(x_i|x_{i-1};t_{i-1})$ depends on the time t_{i-1} and the time step Δt. By definition, the Fisher information involves an off-diagonal part and a diagonal one:

$$F(\theta) = F_{\text{off-dia}}(\theta) + F_{\text{dia}}(\theta), \tag{S57}$$

where

$$F_{\text{off-dia}}(\theta) \equiv \sum_{i=1}^n \sum_{x\neq y,\nu} P_\theta(x,t_{i-1})A^\nu_\theta(y|x;t_{i-1})\left[\frac{\partial}{\partial\theta}\ln A^\nu_\theta(y|x;t_{i-1})\right]^2, \tag{S58}$$

$$F_{\text{dia}}(\theta) \equiv \sum_{i=1}^n \sum_x P_\theta(x,t_{i-1})A_\theta(x|x;t_{i-1})\left[\frac{\partial}{\partial\theta}\ln A_\theta(x|x;t_{i-1})\right]^2. \tag{S59}$$

The total EP can be derived from a change in the Shannon entropy of the system per unit step. For a step between t_{i-1} and t_i, the EP in the heat reservoirs is given by

$$\Delta\sigma_{B_i} = \sum_{x,y,\nu} P(x,t_{i-1})A^\nu(y|x)\ln\frac{A^\nu(y|x)}{A^\nu(x|y)}. \tag{S60}$$

The Shannon entropy change in the system is

$$\Delta\sigma_{S_i} = \sum_x \left[-P(x,t_i)\ln P(x,t_i) + P(x,t_{i-1})\ln P(x,t_{i-1})\right]. \tag{S61}$$

By defining [6, 7], we have

$$\Delta P(x,t_{i-1}) \equiv P(x,t_i) - P(x,t_{i-1}), \tag{S62}$$

$$k^\nu(x|y) \equiv A^\nu(x|y) - \delta_{xy}, \tag{S63}$$

where δ_{xy} is the Kronecker delta, and a change in the Shannon entropy in one step can be calculated as

$$\Delta\sigma_{S_i} = \sum_{x,y,\nu} P(x,t_{i-1})A^\nu(y|x)\ln\frac{P(x,t_{i-1})}{P(y,t_i)}. \tag{S64}$$

Therefore, the total EP in one step reads

$$\Delta\sigma_i = \sum_{x,y,\nu} P(x,t_{i-1})A^\nu(y|x)\ln\frac{P(x,t_{i-1})A^\nu(y|x)}{P(y,t_i)A^\nu(x|y)}. \tag{S65}$$

For n steps, the total EP should be

$$\sigma = \sum_{i=1}^n\sum_{x,y,\nu} P(x,t_{i-1})A^\nu(y|x)\ln\frac{P(x,t_{i-1})A^\nu(y|x)}{P(y,t_i)A^\nu(x|y)}, \tag{S66}$$

As we can see, the probabilities in the logarithm are not equal-time, which is different from the continuous-time case. This feature prevents us from rewriting Eq. (S66) into a form of $(A-B)\ln\frac{A}{B}$ as in the continuous-time one. However, the key equation (16) in the main text requires such a form. We have to modify the total EP into the tilde EP as

$$\begin{aligned}\tilde{\sigma} &\equiv \sigma + \sum_{i=1}^n D_{\mathrm{KL}}(\mathbf{P}(t_i)||\mathbf{P}(t_{i-1}))\\ &= \sum_{i=1}^n\sum_{x,y,\nu} P(x,t_{i-1})A^\nu(y|x)\ln\frac{P(x,t_{i-1})A^\nu(y|x)}{P(y,t_{i-1})A^\nu(x|y)}.\end{aligned} \tag{S67}$$

By choosing the parametrized transition probabilities as

$$A^\nu_\theta(y|x;t_{i-1}) = A^\nu(y|x)e^{\theta\alpha^\nu(y|x;t_{i-1})}, \tag{S68}$$

similar assumptions as in Eqs. (16) and (17) can be made as well:

$$K^\nu_{xy}(\alpha^\nu_{xy})^2 + K^\nu_{yx}(\alpha^\nu_{yx})^2 = \frac{1}{2}(K^\nu_{xy}-K^\nu_{yx})\ln\frac{K^\nu_{xy}}{K^\nu_{yx}}, \tag{S69}$$

$$K^\nu_{xy}\alpha^\nu_{xy} - K^\nu_{yx}\alpha^\nu_{yx} = K^\nu_{xy} - K^\nu_{yx}, \tag{S70}$$

where $K^\nu_{xy}(t_{i-1}) \equiv P(x,t_{i-1})A^\nu(y|x)$ and $\alpha^\nu_{xy}(t_{i-1}) \equiv \alpha^\nu(y|x;t_{i-1})$. It is clear that Eq. (S69) gives

$$F_{\text{off-dia}}(0) = \frac{\tilde{\sigma}}{2}. \tag{S71}$$

The diagonal term of the Fisher information is shown to be bounded from above as

$$\begin{aligned}F_{\text{dia}}(0) &= \sum_{i=1}^n\sum_x P(x)\frac{[\sum_{y:y\neq x,\nu}A^\nu(y|x)\alpha^\nu_{xy}(t_{i-1})]^2}{1-\sum_{y:y\neq x,\nu}A^\nu(y|x)}\\ &\le \sum_{i=1}^n\sum_{x\neq y,\nu} P(x)A^\nu(y|x)(\alpha^\nu_{xy}(t_{i-1}))^2\frac{1-A(x|x)}{A(x|x)}\\ &\le \sum_{i=1}^n\frac{1-\min_x A(x|x)}{\min_x A(x|x)}\sum_{x\neq y,\nu}P(x)A^\nu(y|x)(\alpha^\nu_{xy}(t_{i-1}))^2\\ &\equiv \frac{1-a}{a}\frac{\tilde{\sigma}}{2},\end{aligned} \tag{S72}$$

where Titu's lemma $\frac{(\sum_i u_i)^2}{\sum_i v_i} \le \sum_i\frac{u_i^2}{v_i}$ has been used in the first inequality, and in the second inequality we take the minimal diagonal entry of $\mathcal{A}$, namely the minimal staying probability, denoted as $a \equiv \min_x A(x|x)$ which satisfies $0<a<1$. Finally, we obtain an upper bound on the Fisher information:

$$F(0) \le \frac{1}{a}\frac{\tilde{\sigma}}{2}. \tag{S73}$$

For the current, similarly to the continuous-time case, we expand the parametrized transition probability matrix up to the first order in θ as

$$\mathcal{A}_\theta(t_{i-1}) = \mathcal{A} + \theta\mathcal{A}_1(t_{i-1}) + \mathcal{O}(\theta^2). \tag{S74}$$

It can be shown that $\mathcal{A}_1(t_{i-1})$ satisfies the following important property similar to Eq. (S30):

$$\begin{aligned}[\mathcal{A}_1(t_{i-1})\mathbf{P}(t_{i-1})]_y &= \sum_{x,\nu}P(x,t_{i-1})A^\nu(y|x)\alpha^\nu_{xy}(t_{i-1})\\ &= \sum_{x:x\neq y,\nu}(K^\nu_{xy}\alpha^\nu_{xy} - K^\nu_{yx}\alpha^\nu_{yx})\\ &= \sum_{x:x\neq y,\nu}(K^\nu_{xy} - K^\nu_{yx})\\ &= [\mathbf{P}(t_i) - \mathbf{P}(t_{i-1})]_y,\end{aligned} \tag{S75}$$

where Eq. (S70) is used in the third equality. From Eq. (S75), the parametrized state probabilities can be calculated by iteration as

$$\begin{aligned}\mathbf{P}_\theta(t_i) &= \mathcal{A}_\theta(t_{i-1})\mathbf{P}_\theta(t_{i-1})\\ &= [\mathcal{A}+\theta\mathcal{A}_1(t_{i-1})]\mathbf{P}_\theta(t_{i-1}) + \mathcal{O}(\theta^2)\\ &= [\mathcal{A}+\theta\mathcal{A}_1(t_{i-1})][\mathcal{A}+\theta\mathcal{A}_1(t_{i-2})]\mathbf{P}_\theta(t_{i-2}) + \mathcal{O}(\theta^2)\\ &\cdots\\ &= [\mathcal{A}+\theta\mathcal{A}_1(t_{i-1})]\cdots[\mathcal{A}+\theta\mathcal{A}_1(0)]\mathbf{P}_\theta(0) + \mathcal{O}(\theta^2)\\ &= \{\mathcal{A}^i + \theta[\mathcal{A}_1(t_{i-1})\mathcal{A}^{i-1} + \cdots + \mathcal{A}^{i-1}\mathcal{A}_1(0)]\}\mathbf{P}(0)\\ &\quad + \mathcal{O}(\theta^2)\\ &= \mathbf{P}(t_i) + i\theta[\mathbf{P}(t_i) - \mathbf{P}(t_{i-1})] + \mathcal{O}(\theta^2).\end{aligned} \tag{S76}$$

Hence

$$\left.\frac{\partial}{\partial\theta}P_\theta(x,t_i)\right|_{\theta=0} = i[P(x,t_i) - P(x,t_{i-1})]. \tag{S77}$$

It is now possible to calculate $\psi'(0)$ in the Cramér-Rao inequality:

$$\psi'(0) = \sum_{i=1}^{n} \sum_{x\neq y,\nu} \left[P_\theta(x,t_{i-1}) \frac{\partial}{\partial\theta} A_\theta^\nu(y|x;t_{i-1}) + A_\theta^\nu(y|x;t_{i-1}) \frac{\partial}{\partial\theta} P_\theta(x,t_{i-1}) \right] d^\nu(y|x) \bigg|_{\theta=0}$$
$$= \sum_{i=1}^{n} \sum_{x\neq y,\nu} \{P(x,t_{i-1}) + (i-1)[P(x,t_{i-1}) - P(x,t_{i-2})]\} A^\nu(y|x) d^\nu(y|x)$$
$$= \sum_{i=1}^{n} \sum_{x\neq y,\nu} iP(x,t_{i-1}) A^\nu(y|x) d^\nu(y|x) - \sum_{i=1}^{n-1} \sum_{x\neq y,\nu} iP(x,t_{i-1}) A^\nu(y|x) d^\nu(y|x)$$
$$= n \sum_{x\neq y,\nu} P(x,t_{n-1}) A^\nu(y|x) d^\nu(y|x) \equiv nj(t_{n-1}), \tag{S78}$$

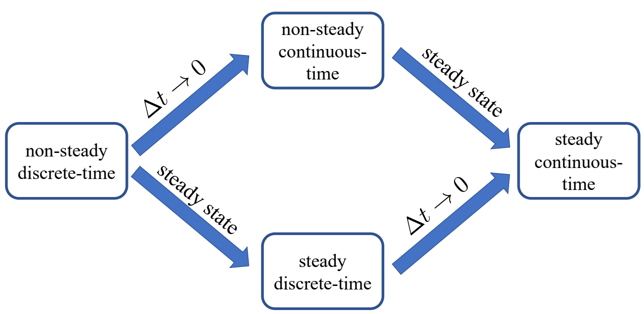

FIG. S4. Hierarchical relations among the four TURs. The non-steady-state discrete-time TUR (11) reduces to the non-steady-state continuous-time TUR (6) in the continuous-time limit, namely in the limit where the time step vanishes: $\Delta t \to 0$. The non-steady state discrete-time TUR reduces to the steady-state discrete-time TUR (S85) for the steady initial state. The non-steady continuous-time TUR and the steady-state discrete-time TUR reduce to the original steady-state continuous-time TUR, namely the conventional TUR (1), for steady initial states and the continuous-time limit, respectively.

where the current at the final step is defined as

$$j(t_{n-1}) \equiv \sum_{x\neq y,\nu} P(x,t_{n-1}) A^\nu(y|x) d^\nu(y|x). \tag{S79}$$

Combining inequality (S73) and Eq. (S78), we obtain the desired TUR (11).

VII. RELATIONS AMONG THE FOUR TURS

There are hierarchical relations among all the bounds obtained in Fig. S4 in which the non-steady-state discrete-time TUR (11) is the most general one. All other three TURs can be deduced from it by considering proper limits, and the conventional TUR (1) can be reduced from any of the other bounds.

Specifically, we prove that the TUR (6) for continuous-time dynamics can be reduced from the discrete-time TUR (11) by taking the continuous-time limit, namely $\Delta t \to 0$. In this limit, we denote $P(x,t_i) = P(x,t_{i-1}) + q(x,t_{i-1})\Delta t$. According to the normalization condition, we have $\sum_x q(x,t_{i-1}) = 0$, the Kullback-Leibler (KL) divergence in the tilde EP (S67) is $\mathcal{O}(\Delta t^2)$:

$$D_{\mathrm{KL}}(\mathbf{P}(t_i)||\mathbf{P}(t_{i-1})) \equiv \sum_x P(x,t_i) \ln \frac{P(x,t_i)}{P(x,t_{i-1})}$$
$$= \sum_x q(x,t_{i-1})\Delta t + \sum_x \frac{q(x,t_{i-1})^2}{P(x,t_{i-1})} \Delta t^2$$
$$= \mathcal{O}(\Delta t^2). \tag{S80}$$

The sum of the KL divergences is thus $\mathcal{O}(\Delta t)$:

$$\sum_{i=1}^{n} D_{\mathrm{KL}}(\mathbf{P}(t_i)||\mathbf{P}(t_{i-1})) = n\mathcal{O}(\Delta t^2) = \mathcal{O}(\Delta t). \tag{S81}$$

Because the total EP σ is $\mathcal{O}(1)$, the tilde EP $\tilde{\sigma}$ is approximately the total EP σ. The transition probabilities can be expressed in terms of transition rates and time step as

$$A^\nu(y|x) = r^\nu(x,y)\Delta t, \tag{S82}$$
$$A(x|x) = 1 - \sum_{y:y\neq x,\nu} A^\nu(y|x) = 1 - \sum_{y:y\neq x,\nu} r^\nu(x,y)\Delta t. \tag{S83}$$

The minimal staying probability is hence approximately given by $a = 1 - \mathcal{O}(\Delta t)$. Eventually, we obtain

$$\frac{\tilde{\sigma}}{a} \to \sigma. \tag{S84}$$

The steady-state discrete-time TUR can simply be obtained from the bound (11) for the steady initial state as:

$$\frac{\mathrm{Var}[J]}{\langle J \rangle^2} \frac{\sigma}{a} \geq 2. \tag{S85}$$

There exists a generalized TUR-like TUR valid for discrete-time Markov chains for an NESS [8]:

$$\frac{\mathrm{Var}[J]}{\langle J \rangle^2} (e^{\Delta\sigma_i} - 1) \geq 2. \tag{S86}$$

Compared with this result, our steady-state discrete-time TUR (S85) is linear in the total EP and we have a new linear bound (11) valid for arbitrary initial states. Thus we have greatly generalized the existing results in the discrete-time regime.

[S1] M. J. M. de Jong, Phys. Rev. B **54**, 8144 (1996).
[S2] D. A. Bagrets and Y. V. Nazarov, Phys. Rev. B **67**, 085316 (2003).
[S3] J. Ren, P. Hänggi, and B. Li, Phys. Rev. Lett. **104**, 170601 (2010).
[S4] A. C. Barato, R. Chetrite, A. Faggionato, and D. Gabrielli, New J. Phys. **20**, 103023 (2018).
[S5] J. M. Horowitz and T. R. Gingrich, Phys. Rev. E **96**, 020103 (2017).
[S6] T. Morimoto, Journal of the Physical Society of Japan **18**, 328 (1963).
[S7] J. Lee, Phys. Rev. E **97**, 032110 (2018).
[S8] K. Proesmans and C. V. den Broeck, Europhys. Lett. **119**, 20001 (2017).