Supplemental Materials for *Entropy production of nano systems with timescale separation*

Shou-Wen Wang
Beijing Computational Science Research Center, Beijing, 100094, China and Department of Engineering Physics, Tsinghua University, Beijing, 100086, China

Kyogo Kawaguchi
Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

Shin-ichi Sasa
Department of Physics, Kyoto University, Kyoto 606-8502, Japan

Lei-Han Tang
Beijing Computational Science Research Center, Beijing, 100094, China and Department of Physics and Institute of Computational and Theoretical Studies, Hong Kong Baptist University, Hong Kong, China

COMPUTATION OF THE CORRELATION AND RESPONSE SPECTRA FROM DATA

To determine the velocity correlation spectrum, we do the following: (a) sample the unperturbed trajectory $x(t)$ at a temporal resolution δt for a duration T_{sp} ($\gg \tau_s$); (b) compute the Discrete Fourier Transform (DFT) $\tilde{x}(\omega)$ of $x(t)$, from which the velocity correlation spectrum $\omega^2|\tilde{x}(\omega)|^2$ is constructed; (c) following the proposal in [1], perform local average of the spectrum to suppress noise (known as data compression).

To determine the velocity response spectrum at selected frequencies $\{\omega_k\}$ simultaneously, we do the following: (a) sample the trajectory $x(t)$ perturbed by a linear combination of periodic driving forces $h = \sum_k h_k \exp(i\omega_k t)$, at a temporal resolution δt for a duration T_{sp} ($\gg \tau_s$); (b) compute the DFT $\tilde{x}(\omega)$ of the perturbed trajectory $x(t)$; (c) at each ω_k, calculate the response $\tilde{R}_x(\omega_k) = \tilde{x}(\omega_k)/h_k$, and thus the velocity response $\tilde{R}_{\dot{x}}(\omega_k) = -i\omega_k \tilde{R}_x(\omega_k)$.

Below we examine in some detail fluctuations of the spectra constructed through the above procedure when a single long trajectory is used in computation. This will allow us to devise suitable averaging procedures to suppress noise in the data without significantly hampering its information content.

Correlation spectrum

The discussion below follows main ideas of the power spectrum analysis presented in Ref. [1].

General considerations

For a general stochastic trajectory $x(t)$ sampled at resolution δt and for duration $T_{sp} = N_{sp}\delta t$, we introduce the following DFT,

$$\tilde{x}(\omega_k) = \frac{\sqrt{\delta t}}{\sqrt{N_{sp}}} \sum_{j=1}^{N_{sp}} x(j\delta t) \exp(-i\omega_k j\delta t), \tag{S1}$$

where $\omega_k = 2\pi k/T_{sp}, k = 0, \ldots, N_{sp}-1$. As illustrated by the harmonic potential switching model below, each Fourier amplitude $\tilde{x}(\omega_k)$ is a random variable whose value, just like the trajectory $x(t)$, changes from realization to realization. Therefore the velocity spectral function

$$g(\omega) = \omega^2|\tilde{x}(\omega)|^2, \tag{S2}$$

obtained from a single trajectory is a strongly fluctuating quantity. To estimate the ensemble averaged spectrum

$$\tilde{C}_{\dot{x}}(\omega) = \langle g(\omega) \rangle, \tag{S3}$$

one may invoke the smoothness of the function by averaging $g(\omega_k)$ over nearby frequencies in a suitable window of width $\Delta\omega$. Since the spacing between accessible frequencies is $\delta\omega = 2\pi/T_{sp}$, relative error of the window-averaged $g(\omega)$ should decrease as $(\Delta\omega/\delta\omega)^{-1/2} = (T_{sp}\Delta\omega/2\pi)^{-1/2}$. On the other hand, $\Delta\omega$ should not be chosen too large to cause significant information loss. For the latter, we may consider the Taylor expansion of $\tilde{C}_{\dot{x}}(\omega)$ around a chosen $\bar{\omega}$,

$$\tilde{C}_{\dot{x}}(\omega) = \tilde{C}_{\dot{x}}(\bar{\omega}) + \frac{d\tilde{C}_{\dot{x}}}{d\omega}\Big|_{\bar{\omega}}(\omega-\bar{\omega}) + \frac{1}{2}\frac{d^2\tilde{C}_{\dot{x}}}{d\omega^2}\Big|_{\bar{\omega}}(\omega-\bar{\omega})^2 + \dots$$

Averaging the above expression over a frequency window of width $\Delta\omega$ centered at $\bar{\omega}$ yields,

$$\overline{\tilde{C}_{\dot{x}}(\omega)} \simeq \tilde{C}_{\dot{x}}(\bar{\omega}) + \frac{1}{24}\frac{d^2\tilde{C}_{\dot{x}}}{d\omega^2}\Big|_{\bar{\omega}}(\Delta\omega)^2.$$

Equating the two terms on the right-hand-side of the expression above, we obtain,

$$\Delta\omega_{\rm M} = \sqrt{24}\left|\frac{\tilde{C}_{\dot{x}}(\bar{\omega})}{\partial^2_{\bar{\omega}}\tilde{C}_{\dot{x}}(\bar{\omega})}\right|^{1/2}. \tag{S4}$$

This sets an upper bound on the window size for averaging. In the part of the spectrum where the curvature of $\tilde{C}_{\dot{x}}(\omega)$ is small, a large window is desirable so as to reduce statistical error in the original data.

Harmonic potential switching model

We now illustrate the above general ideas with an explicit example, the harmonic potential switching model. The bead displacement $x(t)$ from the mid-point $L/2$ satisfies the Langevin equation,

$$\gamma\dot{x} = -kx + \xi(t) + \eta(t), \tag{S5}$$

where $\xi(t)$ and $\eta(t)$ are switching and thermal noise, respectively. Integrating Eq. (S5) over the time interval δt, we obtain,

$$x(t+\delta t) = \exp(-\delta t/\tau_s)x(t) + \gamma^{-1}\int_t^{t+\delta t} dt' \exp\big(\frac{t-t'}{\tau_s}\big)[\xi(t') + \eta(t')], \tag{S6}$$

where $\tau_s = \gamma/k$ is the relaxation time constant of bead displacement.

With the help of Eq. (S6), and ignoring boundary terms at the beginning and end of the time series, we obtain the following equation for the DFT $\tilde{x}(\omega)$ when the sampling time $\delta t \ll \tau_s, 2\pi/\omega$,

$$i\gamma\omega\tilde{x}(\omega) \simeq -k\tilde{x}(\omega) + \tilde{\xi}(\omega) + \tilde{\eta}(\omega). \tag{S7}$$

Here, $\tilde{\xi}(\omega)$ and $\tilde{\eta}(\omega)$ are the usual Fourier transforms of the active noise $\xi(t)$ and thermal noise $\eta(t)$, respectively, independent of δt. Therefore the equation satisfied by $\tilde{x}(\omega)$ is nearly identical to its continuous counterpart, provided the sampling time interval δt is much shorter than both τ_s and $2\pi/\omega$ (i.e., the part of the spectrum with $\omega < 2\pi/\delta t$).

Following Eq. (S7), the velocity correlation spectrum from a single trajectory is given by,

$$g(\omega) = \omega^2|\tilde{x}(\omega)|^2 \simeq \frac{1}{k^2}\frac{\omega^2}{1+(\omega\tau_s)^2}\left[|\tilde{\xi}(\omega)|^2 + 2{\rm Re}\left(\tilde{\xi}(\omega)\tilde{\eta}^*(\omega)\right) + |\tilde{\eta}(\omega)|^2\right], \tag{S8}$$

where $*$ denotes complex conjugate. Since the two noises are uncorrelated from each other, the cross term vanishes upon ensemble average that yields the desired correlation spectrum $\tilde{C}_{\dot{x}}(\omega)$.

The mean of $\tilde{\xi}(\omega)$ and $\tilde{\eta}(\omega)$ are zero while their variances are given by $\langle|\xi(\omega)|^2\rangle = \epsilon\gamma k(L/2)^2/[1+(\omega\tau_f/2)^2]$ and $\langle|\eta(\omega)|^2\rangle = 2\gamma T$, respectively. On the high frequency end of the spectrum, i.e., $\omega \gg \tau_s^{-1}$, $\tilde{C}_{\dot{x}}(\omega)$ varies only on the scale of τ_f^{-1} which is much greater than τ_s^{-1} and even greater than $\delta\omega = 2\pi/T_{sp}$. Therefore $\Delta\omega$ can be chosen to be very large to reduce the noise in the bare correlation spectrum (S8). On the other hand, for frequencies that are comparable or even smaller than τ_s^{-1}, $\tilde{C}_{\dot{x}}(\omega)$ varies appreciably so that the window size is limited by τ_s^{-1}. This is the main reason behind the noisy low frequency spectrum shown in Fig. 1(d) in the Main Text.

Fluctuation of the response spectrum

Now, we consider applying a periodic driving force $h_0 \exp(-i\omega_0 t)$ to the system. Similarly to Eq. (S7), the equation at $\omega = \omega_0$ satisfies,

$$i\gamma\omega_0\tilde{x}(\omega_0) \simeq -k\tilde{x}(\omega_0) + \tilde{\xi}(\omega_0) + \tilde{\eta}(\omega_0) + \sqrt{T_{sp}}h_0. \tag{S9}$$

Here, the prefactor $\sqrt{T_{sp}}$ appears due to the rescaling scheme in Eq. (S1). The velocity response spectrum calculated for a single perturbed trajectory is given by,

$$f(\omega_0) \equiv -i\omega_0 \frac{\tilde{x}(\omega_0)}{h_0\sqrt{T_{sp}}} = -\frac{i\omega_0}{i\gamma\omega_0 + k}\left(1 + \frac{\tilde{\xi}(\omega_0)}{h_0\sqrt{T_{sp}}} + \frac{\tilde{\eta}(\omega_0)}{h_0\sqrt{T_{sp}}}\right), \tag{S10}$$

which contains fluctuations. Its ensemble average converges to the correct result

$$\langle f(\omega_0)\rangle = \tilde{R}_{\dot{x}}(\omega_0). \tag{S11}$$

Its standard deviation is given by

$$\sqrt{\langle |f(\omega_0) - \tilde{R}_{\dot{x}}(\omega_0)|^2\rangle} = \frac{1}{h_0\sqrt{T_{sp}}}\tilde{C}_{\dot{x}}(\omega_0). \tag{S12}$$

Equation (S12) suggests that fluctuations in the velocity response spectrum can be reduced by increasing h_0, thus the signal/noise ratio, and also by using a longer trajectory.

However, h_0 must be small enough to ensure that we are measuring linear response. Besides, T_{sp} may also be limited in experiments. In such a situation, one may consider to perform averaging over nearby frequencies to reduce noise in the response spectrum (S10). For example, we may apply a perturbation with multiple frequencies inside a window of width $\Delta\omega$, i.e., $h_0 \sum_{k=-n_r/2}^{n_r/2-1} \exp(i[\omega_0 + k\delta\omega]t)$, where $\delta\omega \gg 2\pi/T_{sp}$ to avoid possible interference effects. Here $n_r = \Delta\omega/\delta\omega$ is the number of frequencies considered. Averaging over the response at these frequencies will further reduce the error (S12) by a factor $1/\sqrt{n_r}$.

Error bars on the reconstructed spectra

Fig. S1 shows error bars on the reconstructed spectra presented in Fig. 1 of the Main Text. They have been estimated from variations in the data within each window used for averaging. For the parameters chosen, our procedure yields very good results on the high frequency end, but less satisfactory results on the low frequency side close to $\omega \simeq \tau_s^{-1}$. Fluctuations in the latter case are mainly due to insufficient averaging when computing the correlation spectrum. Although we have not attempted to optimize this part of the data analysis, it is conceivable that data smoothening procedures apply to the autocorrelation function $C_x(t)$ can lead to much improved results.

Data points in Fig. S2(a) show the numerically determined FRR violation spectrum from a trajectory 10 times longer than the one used to obtain the data points in Fig. S1(b). Fig. S2(b) shows the data in the same time window but with a much reduced sampling rate $0.1\tau_f^{-1}$. The low frequency part of the violation spectrum is faithfully reproduced.

VALIDITY OF THE HARADA-SASA EQUALITY IN MARKOV SYSTEMS

In this section, we show that the HS equality holds approximately for a discrete Markov jump system with a ladder network. Our argument is a generalization of the work in [2], where the HS equality in a one-dimensional discretized Langevin system was studied. The model we study here, which describes the chemical state dynamics of the chemotaxis-related membrane chemoreceptor in E.coli [3], is illustrated in FIG. S3(a). The chemical state of this receptor is a combination of the methylation level m $(= 0, 1, 2, 3, 4)$, and the activity a $(= 0, 1)$. α is a small parameter that tunes the irreversibility of the methylation/demethylation dynamics, and we only consider the range $\alpha \leq \exp(1)$. The transition from the active state $a = 1$ to the inactive state $a = 0$ at methylation level m is denoted

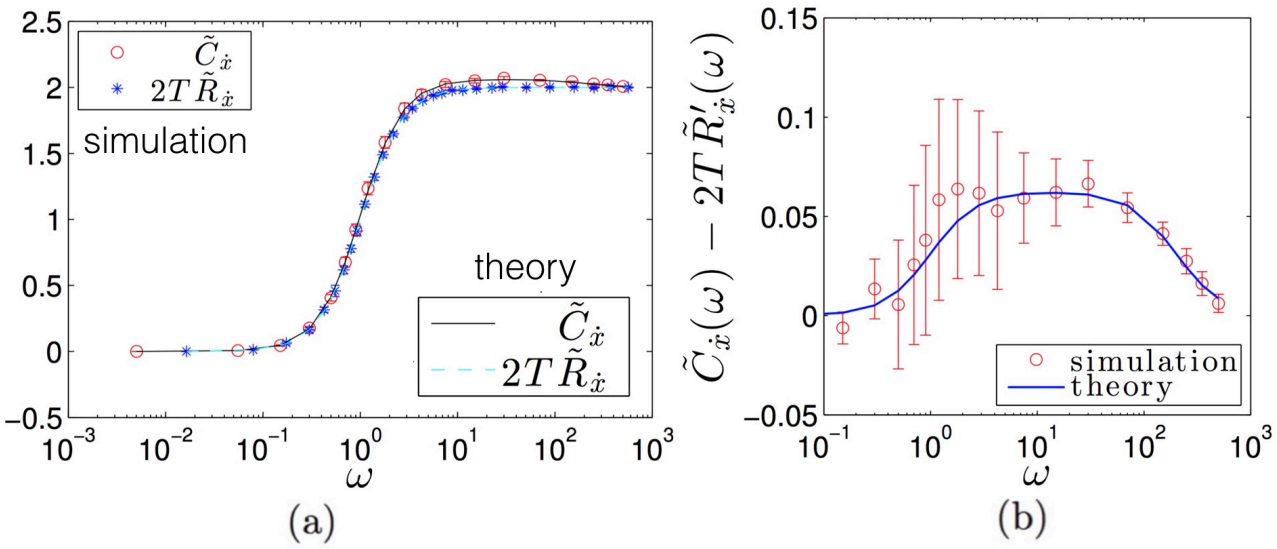
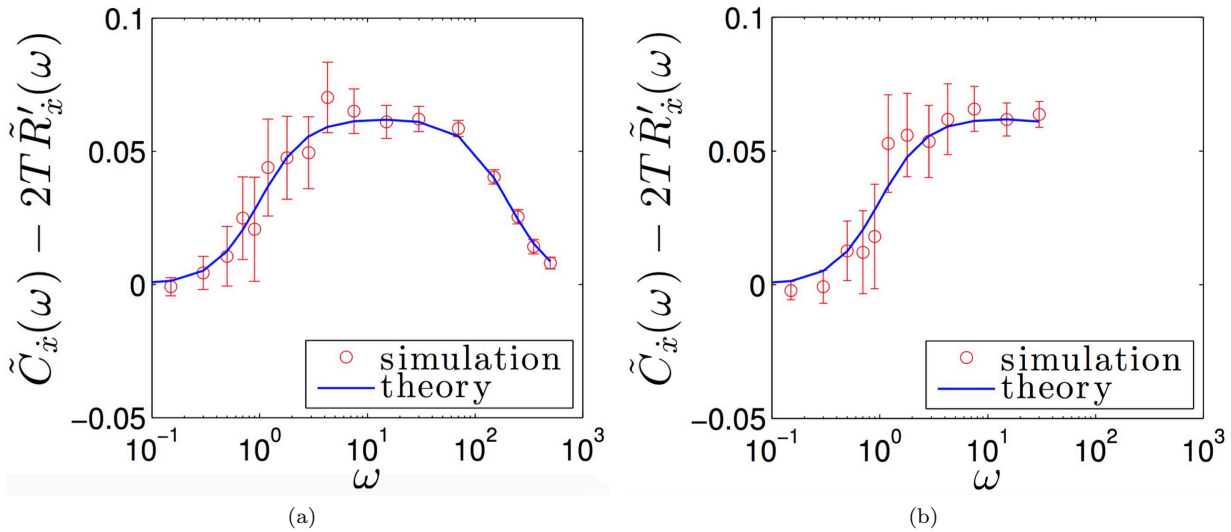


FIG. S1: (a) The velocity correlation and response spectrum reconstructed from a simulated trajectory, the same data as those presented in FIG.(1)(c) of the Main Text but with error bars. (b) The corresponding FRR violation spectrum, plotted with error bars. Parameters: $\gamma = k = 1$, $L = 5$, $\tau_f = 0.01$, $T_{sp} = 10^4$, $\delta t = \tau_f/2$, and $h_0 = 0.5$.



(a) (b)

FIG. S2: The FRR violation spectrum constructed by using a longer simulation time $T_{sp} = 10^5\tau_s$ for the potential switching model. (a) and (b) correspond to the case with temporal resolution $\delta t = \tau_f/2$ and $\delta t = 10\tau_f$, respectively. Other parameters are the same as FIG. S1.

as $w_1(m)$, while the reverse transition rate is denoted as $w_0(m)$. The rates satisfy

$$w_0(m) = \frac{1}{\tau_f}\exp\left(-\frac{\Delta E(m)}{2T}\right),$$
$$w_1(m) = \frac{1}{\tau_f}\exp\left(\frac{\Delta E(m)}{2T}\right),$$

where $\Delta E(m) = e_0(m_* - m)$. Here, we set $e_0 = 2$, $m_* = 2$, $\tau_f = 0.01$ and $r = 1$ in the numerical study. Therefore, dynamics of m is relative slow and the timescale separation index $\epsilon \approx 0.01$. For more details of this model, refer to [4]. Here, we study the validity of the HS equality when applied to the slow variable m.

Suppose that a transition from state n to m occurs with a rate w_n^m. Then this jump produces entropy $\Delta S_n^m =$

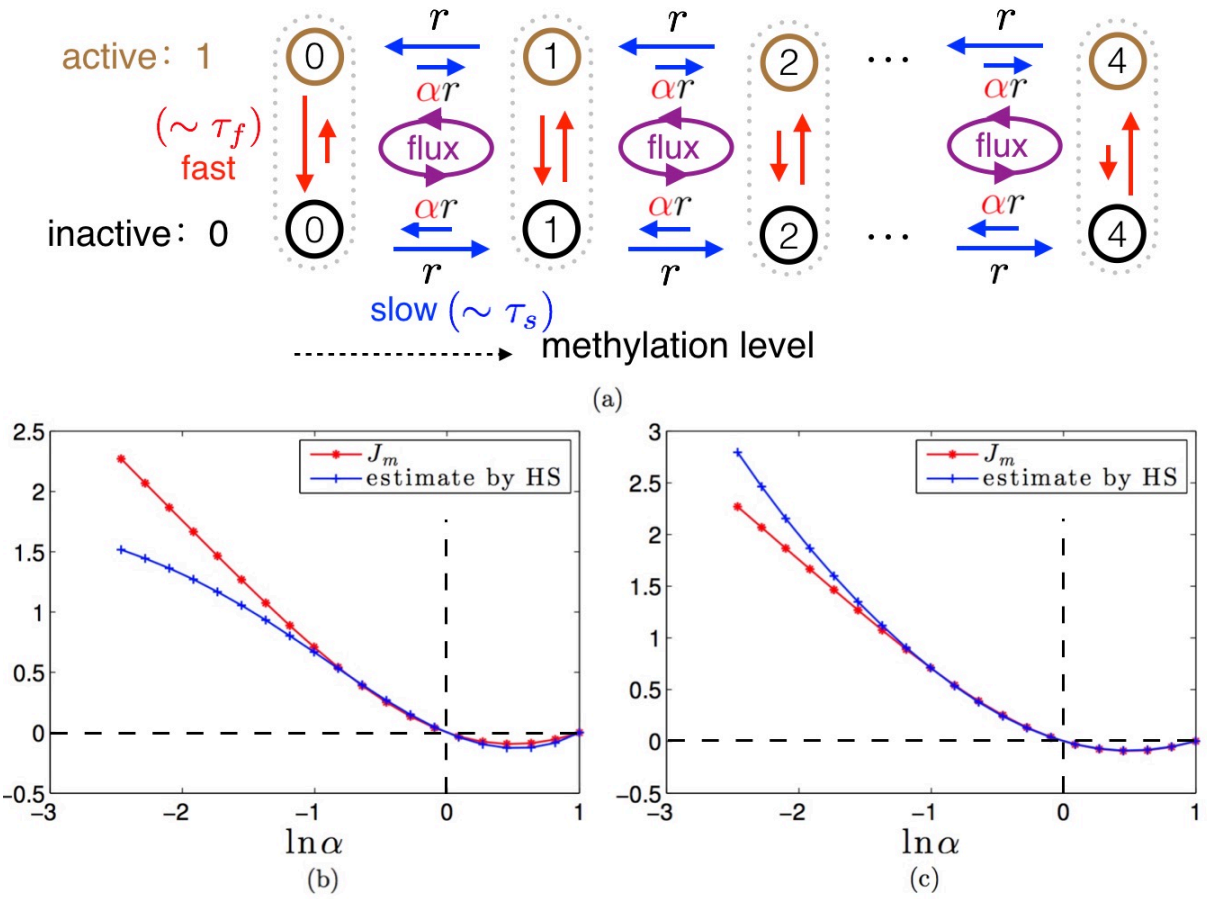


FIG. S3: (a) The sensory adaptation network for E.coli. (b) The comparison between actual dissipation of the methylation dynamics, i.e., J_m, and that estimated from the HS equality, where we use $\gamma_{eff} = \lim_{\omega\to\infty} 1/\tilde{R}_{\dot{m}}(\omega)$. (c) Similar as (b) except that we use $\gamma_{eff} = T/[r\sqrt{\alpha}]$ here.

$\ln[w_n^m/w_m^n]$ in the surrounding media, and the average dissipation rate of this transition is given by

$$[P_n^s w_n^m - P_m^s w_m^n] \ln \frac{w_n^m}{w_m^n},$$

where $[P_n^s w_n^m - P_m^s w_m^n]$ is the net flux for this type of transition. The average dissipation rate due to the change of m, denoted as J_m, is obtained by summing over all dissipations that involves changes of m, which is given by

$$J_m = \sum_m \Big(\left[P_{1,m+1}^s r - P_{1,m}^s \alpha r\right] \ln \frac{1}{\alpha} + \left[P_{0,m+1}^s \alpha r - P_{0,m}^s r\right] \ln \alpha \Big). \tag{S13}$$

where $P_{a,m}^s$ is the stationary distribution at state (a, m).

Now, we consider how to estimate this dissipation rate from the FRR violation spectrum. From the temporal trajectory $m(t)$, we can compute its velocity correlation spectrum $\tilde{C}_{\dot{m}}(\omega)$ and response spectrum $\tilde{R}_{\dot{m}}(\omega)$. Then, we are supposed to estimate the dissipation rate by

$$\gamma_{eff} \left(\langle \dot{m} \rangle + \int_{-\infty}^{\infty} \left[\tilde{C}_{\dot{m}}(\omega) - 2T\tilde{R}'_{\dot{m}}(\omega) \right] \frac{d\omega}{2\pi} \right). \tag{S14}$$

The key component in this estimation is the effective friction coefficient γ_{eff}, which is in general missing for a Markov process. Below, we provide two possible definitions, both of which are rooted in the generalization of friction coefficient defined Langevin systems.

In the over-damped Langevin system, we know that the friction coefficient is the inverse of the velocity response spectrum in the high frequency limit. Therefore, we may define this effective friction coefficient

$$\gamma_{eff} = \lim_{\omega\to\infty} 1/\tilde{R}_{\dot{m}}(\omega), \tag{S15}$$

which is directly measurable.

Alternatively, if we discretize a continuous over-damped Langevin system with spatial unit Δx, its transition rates take the form

$$\frac{T}{\gamma \Delta x^2}\exp(\Delta S/2), \tag{S16}$$

where ΔS is to be understood as the entropy produced in the medium in this single jump. The corresponding backward jump produces medium entropy $-\Delta S$ due to time-reversal asymmetry. This is consistent with the fact that more probable transitions always tend to produce positive entropy in the medium, which is the microscopic origin for the irreversibility of the second law. In the case of methylation dynamics in FIG. S3, we may identify ΔS as the medium entropy produced by the transition with rate r, and rewrite the methylation/demethylation rates in the form of Eq. (S16), i.e., $r = \frac{T}{\gamma_{eff}}\exp(\Delta S/2)$ and $\alpha r = \frac{T}{\gamma_{eff}}\exp(-\Delta S/2)$, where the discretization unit is 1 here. Then, we can identify the effective friction coefficient

$$\gamma_{eff} = \frac{T}{r\sqrt{\alpha}}. \tag{S17}$$

Using the estimator Eq. (S14) and effective friction coefficient defined in Eq. (S15), we obtain the approximate dissipation rate through m dynamics under various α, and compare it with the exact dissipation rate J_m, as shown in FIG. S3(b). It quantitatively captures the dependence of the dissipation rate on α in the range $\alpha \in [e^{-1}, \ e^1]$, where $|\Delta S| = |\ln\alpha| < 1$. Around $\ln\alpha \approx 0$, where the two dash lines intersect, the HS equality becomes almost exact. Alternatively, we may use the effective friction coefficient definition in Eq. (S17), as shown in FIG. S3(c), which is almost exact in the range $\alpha \in [e^{-1}, \ e^1]$. Therefore, both of these two definitions are correct in the limit of small entropy production per step, i.e., $|\Delta S| = |\ln\alpha| \ll 1$.

To understand why the HS estimate fails for large ΔS, below we identify the higher order correction of this estimation. The HS equality is essentially to use the kinetic information to deduce the energetic information, or irreversibility of the dynamics. From the kinetic side, the biased transition rate $(w_+ - w_-)$, i.e., the difference between forward and backward rate, provides the hope to extract the dissipation of this single jump, i.e., $\ln[w_+/w_-]$. This is essentially reduced to the following approximation

$$\begin{aligned} w_+ - w_- &= \frac{T}{\gamma_{eff}}\exp(\Delta S/2) - \frac{T}{\gamma_{eff}}\exp(-\Delta S/2) \\ &= \frac{T}{\gamma_{eff}}\Big[\Delta S + \frac{1}{24}(\Delta S)^3 + o(\Delta S^3)\Big] \\ &= \frac{T}{\gamma_{eff}}\Big[\ln[w_+/w_-] + \frac{1}{24}(\Delta S)^3 + o(\Delta S^3)\Big] \end{aligned} \tag{S18}$$

Therefore, the HS equality is a leading order approximation, which becomes exact for Langevin dynamics. Since the correction is in the third order of ΔS, the HS equality can still be reasonably good for many discrete Markov systems, as we have demonstrated in this sensory adaptation network. We will discuss this more deeply in our coming paper.

DERIVATION OF EQUATION (10) IN THE MAIN TEXT

Now, we study a special type of observable $Q_{p_k} = Q_p$ that only depends on the coarse-grained state, and therefore evolves on a slow timescale τ_s. We ask how its violation spectrum may reveal the information of the microscopic dynamics that is hidden from our observation.

Intuitively, its dynamics seems to be equally well described in both the complete state space by matrix M and the coarse-grained state space by $\widehat{M}$. The projection coefficients for the original system is $\alpha_j \equiv \sum_{p_k} Q_p x_j(p_k)$, $\beta_j \equiv \sum_{p_k} Q_p y_j(p_k) P^s_{p_k}$, and $\phi_j \equiv \sum_{p_k} B_{p_k} y_j(p_k)$, where $B_{p_k} \equiv \sum_{q_l}[w^{p_k}_{q_l} P^s_{q_l} + w^{q_l}_{p_k} P^s_{p_k}](Q_p - Q_q)/2T$. On the other hand, in the coarse-grained system, we may also introduce the projection coefficients $\widehat{\alpha}_j \equiv \sum_p Q_p \widehat{x}_j(p)$, $\widehat{\beta}_j \equiv \sum_p Q_p \widehat{y}_j(p)\widehat{P}^s_p$, and $\widehat{\phi}_j \equiv \sum_p \widehat{B}_p \widehat{y}_j(p)$, where $\widehat{B}_p \equiv \sum_q [\widehat{w}^p_q \widehat{P}^s_q + \widehat{w}^q_p \widehat{P}^s_p](Q_p - Q_q)/2T$. Here, $\widehat{w}^q_p$ is the transition rate of the coarse-grained dynamics, which is given by $\widehat{w}^q_p \equiv \sum_{k,l} w^{q_l}_{p_k} P^s(k|p)$. Note that $p \neq q$ here. Besides, we define λ_j as $\sum_m M_{nm} x_m = -\lambda_j x_n$ and $\sum_m y_m M_{mn} = -\lambda_j y_n$, which affects the sign below.

To connect the two levels of description, we use the following relations concerning eigenmodes ($j \leq K$) at the two levels,

$$x_j(p_k) = \widehat{x}_j(p)P^s(k|p) + O(\epsilon), \tag{S19a}$$
$$y_j(p_k) = \widehat{y}_j(p) + O(\epsilon), \tag{S19b}$$

with $\lambda_j = \widehat{\lambda}_j + O(\epsilon)$. A special but important case of Eq. (S19) is

$$P^s_{p_k} = \widehat{P}^s_p P^s(k|p) + O(\epsilon). \tag{S20}$$

Applying these relations to the projection coefficients, we obtain the following results for the slow modes ($j \leq K$)

$$\alpha_j = \widehat{\alpha}_j + O(\epsilon), \quad \beta_j = \widehat{\beta}_j + O(\epsilon), \quad \phi_j = \widehat{\phi}_j + O(\epsilon). \tag{S21}$$

The sum rule of the effective system $\widehat{M}$ demands that

$$\sum_{j=2}^{K} \widehat{\alpha}_j(\widehat{\beta}_j\widehat{\lambda}_j - T\widehat{\phi}_j) = 0. \tag{S22}$$

For fast modes ($j > K$), perturbative analysis shows that

$$x_j(p_k) = \delta_{p,q}\bar{x}^q_j(k) + O(\epsilon), \tag{S23a}$$
$$y_j(p_k) = \delta_{p,q}\bar{y}^q_j(k) + O(\epsilon), \tag{S23b}$$

with $\lambda_j = \epsilon^{-1}\lambda^q + O(1)$. Here, $\bar{x}^q_j(k)$ and $\bar{y}^q_j(k)$ are non-stationary eigenmodes of the matrix $\epsilon^{-1}M^q$ describing transitions within the coarse-grained state q, with $\epsilon^{-1}\lambda^q$ being the corresponding eigenvalue. These non-stationary eigenmodes ($\lambda^q \neq 0$) satisfy

$$\sum_k \bar{x}^q_j(k) = 0, \quad \sum_k \bar{y}^q_j(k)P^s(k|q) = 0, \tag{S24}$$

which results from the orthogonal relations with the stationary eigenmodes ($\lambda^q = 0$). With these relations, we can prove that the fast modes ($j > K$) satisfy

$$\alpha_j = O(\epsilon), \quad \beta_j = O(\epsilon), \quad \phi_j = O(1). \tag{S25}$$

With the above preparation, we now derive Eq.(10) in the Main Text. First, we note that there is a big gap between eigenvalues of the slow modes, which are of order $\tau_s^{-1} \sim 1$, and those of the fast modes, which are of order $\tau_f^{-1} \sim \epsilon^{-1}$. Therefore, in the frequency region $\omega \ll \tau_f^{-1}$, only the slow modes contribute to correlation and response spectrum according to Eq.(6) in the Main Text, i.e.,

$$\tilde{C}_{\dot{Q}}(\omega) = \sum_{j=2}^{K} 2\alpha_j\beta_j\lambda_j\Big[1 - \frac{1}{1+(\omega/\lambda_j)^2}\Big] + \epsilon^3 O([\omega\tau_s]^2), \tag{S26a}$$

$$\tilde{R}_{\dot{Q}}(\omega) = \sum_{j=2}^{K} \alpha_j\phi_j\Big[1 - \frac{1-i(\omega/\lambda_j)}{1+(\omega/\lambda_j)^2}\Big] + \epsilon^3 O([\omega\tau_s]^2), \tag{S26b}$$

where the correction comes from the fast modes. According to Eq. (S25), this correction is of order $\epsilon(\omega\tau_f)^2$, or equivalently, $\epsilon^3(\omega\tau_s)^2$. Then, the violation spectrum for $\omega \ll \tau_f^{-1}$ can be written as

$$\tilde{C}_{\dot{Q}}(\omega) - 2T\tilde{R}'_{\dot{Q}}(\omega) = 2\sum_{j=2}^{K} \alpha_j \frac{T\phi_j - \beta_j\lambda_j}{1+(\omega/\lambda_j)^2} + 2\sum_{j=2}^{K} \alpha_j(\beta_j\lambda_j - T\phi_j) + \epsilon^3 O([\omega\tau_s]^2). \tag{S27}$$

Second, we apply the relations Eq. (S21) to this violation spectrum, and obtain for $\omega \ll \tau_f^{-1}$

$$\tilde{C}_{\dot{Q}}(\omega) - 2T\tilde{R}'_{\dot{Q}}(\omega) = 2\sum_{j=2}^{K} \widehat{\alpha}_j \frac{T\widehat{\phi}_j - \widehat{\beta}_j\widehat{\lambda}_j}{1+(\omega/\widehat{\lambda}_j)^2} + \epsilon V_s(\omega), \tag{S28}$$

which is exactly Eq.(10) in the Main Text, with $\epsilon V_s(\omega)$ a residual contribution defined by Eq. (S28). $V_s(\omega)$ vanishes for $\omega \gg \tau_f^{-1}$ due to FRR in the high frequency limit. For $\omega \ll \tau_f^{-1}$, we have

$$V_s(\omega) = \frac{2}{\epsilon}\sum_{j=2}^{K}\alpha_j(\beta_j\lambda_j - T\phi_j) + \frac{2}{\epsilon}\sum_{j=2}^{K}\alpha_j\frac{T\phi_j - \beta_j\lambda_j}{1+(\omega/\lambda_j)^2} - \frac{2}{\epsilon}\sum_{j=2}^{K}\widehat{\alpha}_j\frac{T\widehat{\phi}_j - \widehat{\beta}_j\widehat{\lambda}_j}{1+(\omega/\widehat{\lambda}_j)^2} + \epsilon^2 O([\omega\tau_s]^2). \tag{S29}$$

The diverging contribution from the first term vanishes due to the approximate relations Eq. (S21) and the sum rule Eq. (S22), and the other two diverging contributions from the second and the third term cancel each other due to Eq. (S21). Therefore, $V_s(\omega)$ is a well-defined frequency-dependent function in the timescale separation limit $\epsilon \to 0$. Below, we analyze its frequency dependence.

In the intermediate frequency region $\tau_s^{-1} \ll \omega \ll \tau_f^{-1}$, both the second and the third term in Eq. (S29) vanish, and we obtain

$$V_s(\omega) = \frac{2}{\epsilon}\sum_{j=2}^{K}\alpha_j(\beta_j\lambda_j - T\phi_j) \xrightarrow{\epsilon\to 0} const \tag{S30}$$

due to Eq. (S21) and the sum rule Eq. (S22). For $\omega \ll \lambda_2 \sim \tau_s^{-1}$, the third term in Eq. (S29) becomes $2\epsilon^{-1}\sum_{j=2}^{K}\widehat{\alpha}_j(T\widehat{\phi}_j - \widehat{\beta}_j\widehat{\lambda}_j)$, which vanishes due to the sum rule Eq. (S22). Besides, the second term becomes $2\epsilon^{-1}\sum_{j=2}^{K}\alpha_j(T\phi_j - \beta_j\lambda_j)$ in this frequency region, which cancels the first term. Therefore, $V_s(\omega) = 0$ for $\omega \ll \tau_s^{-1}$. To conclude, $V_s(\omega)$ vanishes both in the high ($\omega \gg \tau_f^{-1}$) and low ($\omega \ll \tau_s^{-1}$) frequency region, and it becomes a plateau in the intermediate frequency region.

[1] K. Berg-Sørensen and H. Flyvbjerg, Rev. Sci. Instrum. **75**, 594–612 (2004).
[2] E. Lippiello, M. Baiesi and A. Sarracino, Phys. Rev. Lett. **112**, 140602 (2014).
[3] Y. Tu, Annu. Rev. Biophys. **42**, 337 (2013).
[4] S.-W. Wang, Y. Lan, and L.-H. Tang, J. Stat. Mech. Theor. Exp. **2015**, P07025 (2015).