Bayesian estimation of the internal structure of proteins from single-molecule measurements

Makito Miyazaki[1,a)] and Takahiro Harada[2,b)]

[1]*Department of Physics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*
[2]*Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan*

In single-molecule protein experiments, the observable variables are restricted within a small fraction of the entire degrees of freedom. Therefore, to investigate the physical nature of proteins in detail, we always need to estimate the hidden internal structure referring only to the accessible degrees of freedom. We formulate this problem on the basis of Bayesian inference, which can be applied to various complex systems. In the ideal case, we find that in general the framework actually works. Although careful numerical studies confirm that our method outperforms the conventional method by up to two orders of magnitude, we find a striking phenomenon: a loss-of-precision transition occurs abruptly when the design of the observation system is inappropriate. The basic features of the proposed method are illustrated using a simple but nontrivial model. © *2011 American Institute of Physics*. [doi:10.1063/1.3516587]

I. INTRODUCTION

Continual experimental innovations in recent years have produced high-resolution data on the dynamics of complex systems such as proteins[1–7] and cells.[8,9] However, even when provided with the most sophisticated experimental techniques, one can access only a limited numbers of the degrees of freedom out of the entire complex system. For instance, even in single molecular motors, one of the best-studied systems, the motion of a large probe particle coupled to the motor is typically the only observable[4–7] (see Fig. 1). Thus, we always face the following problem: estimation of the hidden internal structure from measurements on only the visible part of the system.

Despite the generality and the importance of problems of this kind, a general and practical framework applicable to them is still out of reach. In many recent experimental studies, heuristic methods were used to analyze time-series data,[10] the relevance of which has not been theoretically established. In addition, several theoretical studies have attempted to extract the topology of the (Markovian) state space from single-molecule time series.[11–13] Although these theories aim to obtain information on the structure of the state space, a framework that enables the estimation of the physical parameters of systems with hidden internal structure has not yet been proposed.

This article presents a framework for inferring the physical parameters of systems in the presence of hidden degrees of freedom. We base our approach on a standard Bayesian argument, and thus, it applies to a wide spectrum of systems. By using this framework, we obtain general evidence that the error of the estimate decreases in proportion to the observation time, and in the limit of infinite observation time, the system parameters can be estimated exactly. We also examine the practical utility of this method for finite observation time and in the presence of measurement noise. By comparing with the estimates from spectrum fitting, we confirm that the proposed method is generally more accurate than a conventional method by at most two orders of magnitude. In addition, we find that a critical line exists in the original parameter space below which the precision of the estimate is abruptly lost.

II. GENERAL FRAMEWORK

We first describe the mathematical framework of our method. Suppose that a system consists of multiple degrees of freedom denoted by $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_m)$. We assume that the motion of the members of the vector $\mathbf{y}$ can be observed with a sufficiently high temporal and spatial resolution, while the motion of the members of the vector $\mathbf{x}$ is completely hidden. A set of parameters $\mathbf{\Pi}^* = (\Pi_1^*, \Pi_2^*, \ldots, \Pi_p^*)$ characterizes the system, and an asterisk denotes true values of the system parameters. Here, our task is to estimate the values of the system parameters $\mathbf{\Pi}^*$ without referring to the motion of $\mathbf{x}$ (the hidden part); thus our estimates are based on measurements of the motion of $\mathbf{y}$.

If the system is simple enough, we can determine the motion of $\mathbf{x}$ from the motion of $\mathbf{y}$ by assuming an appropriate model and solving it. Once we determine the motion of the entire system we use a variety of fitting algorithms to estimate $\mathbf{\Pi}^*$. However, such a procedure requires a case-by-case treatment and expertise.

To overcome such problems, we adopt the framework of Bayesian inference,[14–16] which not only enables us to formulate the problem in a unified way but also facilitates calculation of the marginal likelihood by various approximation techniques. To apply the Bayesian approach, we need to assume an appropriate model of the system. If the molecular structure

---

a)Author to whom correspondence should be addressed. Department of Physics, Graduate School of Science, the University of Tokyo. Electronic mail: miyazaki@chem.scphys.kyoto-u.ac.jp.
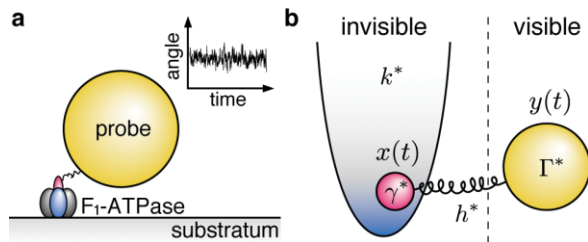b)Electronic mail: harada@phys.s.u-tokyo.ac.jp.

FIG. 1. Example of a system with hidden degrees of freedom. (a) Side view of the single-molecule observation system for a rotary molecular motor, F_1-ATPase (Refs. 5 and 19). Attaching a probe particle to the central subunit γ (magenta) allows us to visualize the stepwise rotation of the central subunit, which is surrounded by cylinder subunits $\alpha_3\beta_3$ (gray and blue). Since the connection is elastic, the motion of the probe particle does not precisely follow that of the central subunit. To simplify the problem, we consider the case where ATP is absent, and thus, the central subunit fluctuates around a certain resting angle. Here, we want to estimate the physical parameters of the protein, such as the interaction potential between the central subunit and the cylinder subunits, and the friction coefficient for the rotation of the central subunit by measuring the fluctuating motion of the probe particle alone. Moreover, this is the usual case in single-molecule experiments, the elasticity of the linker is unknown, and because the probe particle is very close to the substratum, it is also difficult to predict the friction coefficient of the probe particle from Stokes' law (Ref. 20). (b) Model for the system depicted in (a). A small bead trapped in a harmonic potential corresponds to the central subunit γ, and a large bead connected to it via a spring corresponds to the probe particle. As in the usual approximation for this rotary molecular motor, we assume that spatial degrees of freedom other than rotation along the central axis of the molecule are irrelevant. For simplicity, we also assume that the central position of the harmonic potential is already given by the histogram of probe positions.

of the protein being examined has been solved, we can construct an elastic network model based on the structure. Note that at this moment, we need not know the physical parameters of the model such as elastic moduli and friction coefficients. Instead, estimation of these parameters is the task we consider here.

Once we determine the structure of the mechanical model we can write a formal expression of the probability for a trajectory of the system from time $t = 0$ to $t = \tau$, denoted by the degrees of freedom mentioned previously as $[\mathbf{x}, \mathbf{y}]$. Let $P([\mathbf{x}, \mathbf{y}]|\mathbf{\Pi})$ express this path probability, which indicates the probability density of a certain trajectory $[\mathbf{x}, \mathbf{y}]$ under a given parameter value $\mathbf{\Pi}$. To apply Bayes' theorem, we need to calculate the marginal likelihood, defined as

$$P([\mathbf{y}]|\mathbf{\Pi}) \equiv \int \mathcal{D}\mathbf{x} P([\mathbf{x}, \mathbf{y}]|\mathbf{\Pi}), \tag{1}$$

where the integral is taken over all possible trajectories of $\mathbf{x}$. The marginal likelihood [Eq. (1)] along with Bayes' theorem yields the expression for the posterior probability of the parameters

$$P(\mathbf{\Pi}|[\mathbf{y}]) = \frac{P([\mathbf{y}]|\mathbf{\Pi})P(\mathbf{\Pi})}{P([\mathbf{y}])}, \tag{2}$$

where $P(\mathbf{\Pi})$ is the prior probability for the system parameters. In accordance with the Bayesian inference, the estimate of the parameter denoted by $\hat{\mathbf{\Pi}}$ can be obtained by maximizing the posterior probability $P(\mathbf{\Pi}|[\mathbf{y}])$ with respect to $\mathbf{\Pi}$. Here, we adopt the noninformative prior $P(\mathbf{\Pi}) = $ constant over a large parameter region; this enables us to replace the maximization of $P(\mathbf{\Pi}|[\mathbf{y}])$ with the minimization of the "Hamiltonian" defined by using the marginal likelihood

$$\mathcal{H}_\tau(\mathbf{\Pi})[\mathbf{y}] \equiv -\tau^{-1} \ln P([\mathbf{y}]|\mathbf{\Pi}). \tag{3}$$

Once the system parameters have been estimated, the Bayesian framework also allows us to estimate the motion of the hidden degrees of freedom as the maximum a posteriori (MAP) estimator. The MAP estimator is obtained by maximizing $P([\mathbf{x}]|[\mathbf{y}]; \hat{\mathbf{\Pi}})$ with respect to $[\mathbf{x}]$.[17] Here, we use Bayes' theorem again

$$P([\mathbf{x}]|[\mathbf{y}]; \hat{\mathbf{\Pi}}) = \frac{P([\mathbf{x}, \mathbf{y}]|\hat{\mathbf{\Pi}})}{P([\mathbf{y}]|\hat{\mathbf{\Pi}})}, \tag{4}$$

we can replace the maximization of $P([\mathbf{x}]|[\mathbf{y}]; \hat{\mathbf{\Pi}})$ with that of the path probability $P([\mathbf{x}, \mathbf{y}]|\hat{\mathbf{\Pi}})$. Therefore, we can obtain the MAP estimator from the following equation:

$$\frac{\delta}{\delta \mathbf{x}(t)} \ln P([\mathbf{x}, \mathbf{y}]|\hat{\mathbf{\Pi}}) = 0. \tag{5}$$

The MAP estimator $[\hat{\mathbf{x}}]$ represents the most heavily weighted trajectory of $\mathbf{x}(t)$. As Bialek *et al.* pointed out,[18] it corresponds to the classical trajectory in the language of field theory. (Regarding the correctness of $[\hat{\mathbf{x}}]$, see the discussion below.)

The strategy mentioned above is quite standard in the framework of the Bayesian inference. The question we address here is: can we precisely estimate the system parameters within this framework? Although the answer depends on the nature of the system of interest, we can obtain a simple and general consequence for the ideal case with an infinitely large observation time $\tau \to \infty$. In this case, we can prove that $\mathcal{H}_\infty(\mathbf{\Pi})$ has a minimum exactly at $\mathbf{\Pi}^*$ (see Appendix A). Therefore, if the value of the Hamiltonian evaluated at this point is smaller than any other minima, we can obtain the correct estimate. Furthermore, if we assume that this point is the global minimum, we can then derive the general consequence that the error of the estimates is proportional to τ^{-1} for finite τ (see Appendix A). Therefore, if the above assumption is correct, the systematic error of the estimate decreases asymptotically proportional to τ, and in the limit of $\tau \to \infty$, we can obtain the correct estimate.

However, the above assumption is nontrivial, and thus we must confirm whether other minima of $\mathcal{H}_\tau(\mathbf{\Pi})[\mathbf{y}]$ exist in the parameter space. In addition, because τ can be large but finite in practice, we must carefully examine how the finiteness of the data affects the estimate. Furthermore, in actual experiments, contamination with measurement noise is unavoidable. In the next section, we study a simple but nontrivial example to check whether the present framework actually works.

III. EXAMPLE: BROWNIAN DOUBLE PENDULUM

A. Model

To verify the practical utility of our framework, we examine a simple model shown in Fig. 1(b). In this model, we consider two Brownian particles connected via a spring, and one of the particles is trapped in a harmonic potential. If we regard the trapped particle as a central subunit of the rotary molecular motor F_1-ATPase, and the other particle as the probe particle attached to the central subunit, this serves as a very crude

model of a typical experimental setup for F_1-ATPase (Ref. 19) (see Fig. 1).

By letting $x(t)$ and $y(t)$ denote the angles of the central subunit and the probe, respectively, the dynamics of the entire system can be modeled by the Langevin equations[21]

$$\gamma \dot{x}(t) = -kx(t) - h(x(t) - y(t)) + \xi(t), \tag{6}$$

$$\Gamma \dot{y}(t) = -h(y(t) - x(t)) + \eta(t), \tag{7}$$

where $\xi(t)$ and $\eta(t)$ are the zero-mean white Gaussian noise with variances $2\gamma k_B T$ and $2\Gamma k_B T$, respectively, and they are uncorrelated with each other. k_B is the Boltzmann constant and T is the temperature of the system. Four physical parameters characterize this model: the stiffnesses of the harmonic potential k^* and the connecting spring h^*, and the friction coefficients of the trapped bead γ^* and the probe particle Γ^*.

We suppose the position of the trapped bead $x(t)$ is completely invisible, whereas we can monitor the position of the probe particle $y(t)$ at sufficiently high temporal and spatial resolution. Then, we must simultaneously estimate the four physical parameters $\mathbf{\Pi}^* \equiv \{k^*, h^*, \gamma^*, \Gamma^*\}$ that only refer to the motion of the probe $y(t)$. We assume that we already know T because we need it to specify the energy scale of the system.

Note that the present choice of the model is solely for the sake of simplicity. We study the simplest case here to obtain basic insights into the effectiveness of the proposed method. However, the proposed framework itself is applicable to more complicated systems that have more visible and hidden degrees of freedom as well as nonlinear interactions among the components. We discuss such applications in our final remarks.

B. Path probability

The path probability for the entire system is decomposed into the initial and transition probabilities,

$$P([x, y]|\mathbf{\Pi}) = P_{\text{init}}(x_0, y_0|\mathbf{\Pi}) P_{\text{tr}}((x_0, y_0) \to [x, y]|\mathbf{\Pi}), \tag{8}$$

where x_0 and y_0 express the position of the central subunit and the probe, respectively, at $t = 0$. For the initial probability, the canonical distribution under the given temperature is naturally chosen as

$$P_{\text{init}}(x, y|\mathbf{\Pi}) = \frac{\sqrt{kh}}{2\pi k_B T} \exp[-\beta H(x, y; \mathbf{\Pi})], \tag{9}$$

$$H(x, y; \mathbf{\Pi}) \equiv \frac{k}{2}x^2 + \frac{h}{2}(y - x)^2, \tag{10}$$

where $\beta^{-1} \equiv k_B T$.

Next, we consider the transition probability. Let the data be collected at discrete time intervals Δt. Thus, the number of time points is expressed as $N + 1$ with $N \equiv \tau/\Delta t$. We express the value of variables at each time point by $x_i \equiv x(i\Delta t)$ and $y_i \equiv y(i\Delta t)$. Following these notations, the transition probability can be expressed in terms of the

Onsager–Machlup path probability[22,23]

$$P_{\text{tr}}((x_0, y_0) \to [x, y]|\mathbf{\Pi})$$

$$= \left(\frac{\sqrt{\gamma\Gamma}}{4\pi k_B T \Delta t}\right)^N \exp[-\beta S([x, y]; \mathbf{\Pi})], \tag{11}$$

where the action functional is expressed as

$$S([x, y]; \mathbf{\Pi}) \equiv \frac{1}{4\gamma} \sum_{i=0}^{N-1} \Delta t \left[\gamma \frac{x_{i+1} - x_i}{\Delta t} + k\frac{x_{i+1} + x_i}{2}\right.$$

$$\left. + h\left(\frac{x_{i+1} + x_i}{2} - \frac{y_{i+1} + y_i}{2}\right)\right]^2$$

$$- \frac{k_B T(k + h)\tau}{2\gamma} + \frac{1}{4\Gamma} \sum_{i=0}^{N-1} \Delta t \left[\Gamma \frac{y_{i+1} - y_i}{\Delta t}\right.$$

$$\left. + h\left(\frac{y_{i+1} + y_i}{2} - \frac{x_{i+1} + x_i}{2}\right)\right]^2 - \frac{k_B T h\tau}{2\Gamma}. \tag{12}$$

Note that the variables are evaluated at the intermediate time points, corresponding to the Stratonovich-type interpretation.[24,25]

C. Calculating the Hamiltonian by the Wentzel–Kramers–Brillouin (WKB) method

According to the definition of the marginal likelihood [Eq. (1)], we integrate out the trajectory of $x(t)$ in Eq. (8). The Wentzel–Kramers–Brillouin method allows this calculation to be done straightforwardly.[26] Owing to the linearity of the present model, we can obtain the exact expression for $P([y]|\mathbf{\Pi})$ as follows.

The first step in the WKB method is to determine the classical trajectory, or, in the language of Bayesian statistics, the MAP estimator $[\hat{x}]$. Because our final task is to estimate $\mathbf{\Pi}$, we solve Eq. (5) here using temporarily settled values of $\mathbf{\Pi}$ to obtain the temporal MAP estimator. The temporal MAP estimator depends on the observed trajectory $[y]$ and the choice of the temporal parameter set $\mathbf{\Pi}$. Therefore, we specify this dependency as $[\hat{x}]([y]; \mathbf{\Pi})$. For the present model, we can obtain a unique and exact solution of $[\hat{x}]([y]; \mathbf{\Pi})$.[27]

The second step is to calculate the path integral. For convenience, we rewrite the path probability [Eqs. (8)–(12)] as follows:

$$P([x, y]|\mathbf{\Pi}) = \frac{\sqrt{kh}}{2\pi k_B T} \left(\frac{\sqrt{\gamma\Gamma}}{4\pi k_B T \Delta t}\right)^N$$

$$\times \exp[-\beta F([x, y]; \mathbf{\Pi})], \tag{13}$$

$$F([x, y]; \mathbf{\Pi}) \equiv H(x_0, y_0; \mathbf{\Pi}) + S([x, y]; \mathbf{\Pi}). \tag{14}$$

According to the WKB method, we can use the MAP estimator $[\hat{x}]([y]; \mathbf{\Pi})$ to rewrite $F([x, y]; \mathbf{\Pi})$ as

$$F([x, y]; \mathbf{\Pi}) = F([\hat{x}([y]; \mathbf{\Pi}), y]; \mathbf{\Pi})$$

$$+ \frac{1}{2} \sum_{i,j=0}^{N} (x_i - \hat{x}_i([y]; \mathbf{\Pi})) \Lambda_{ij}$$
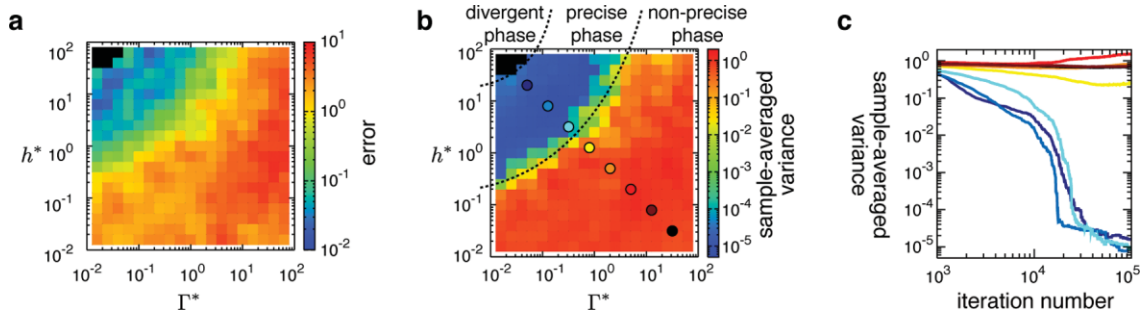
$$\times (x_j - \hat{x}_j([y]; \mathbf{\Pi})), \tag{15}$$

FIG. 2. Results of the numerical experiments. The system of units is fixed as $k^* = \gamma^* = k_B T = 1$. It shows the Γ^* and h^* dependence of the (a) systematic error and (b) sample-averaged variance of $\mathbf{\Pi}^{\ddagger}$. Because the estimate should be a positive number, these quantities are defined on a logarithmic scale as $err. \equiv \|\langle \log \hat{\mathbf{\Pi}}_j \rangle_s - \log \mathbf{\Pi}^* \|$ and $s.a.var. \equiv \langle \langle \| \log \mathbf{\Pi}_{i,j}^{\ddagger} - \langle \log \mathbf{\Pi}_{i,j}^{\ddagger} \rangle_r \|^2 \rangle_r \rangle_s$, where $\|\cdots\|$ represents the Euclidean norm, $\langle \cdots \rangle_r \equiv \frac{1}{r}\Sigma_{i=1}^r \cdots \rangle_r$ represents the replica average, and $\langle \cdots \rangle_s \equiv \frac{1}{s}\Sigma_{j=1}^s \cdots$ represents the sample average (i: index number of initial setting of $\mathbf{\Pi}^{\ddagger}$, j: index number of [y]). The black regions in the upper left corners of (a) and (b) represent areas where the estimates always diverge in the optimization process. We also evaluate the estimate at a normal scale and obtain almost the same features as at a logarithmic scale. (c) Relaxation time courses of the sample-averaged variance in the optimization process. These data are taken from data points in (b). The color of dots in (b) corresponds to the color of lines in (c). In all cases, the time resolution and the length of the time series are $\Delta t = 10^{-4}$ and $\tau = 10$, respectively. The iteration number of the optimization process is fixed at 10^5, and the total replica number and the total sample number are $r = 10$ and $s = 3$, respectively.

where the $(N + 1) \times (N + 1)$-square matrix $\mathbf{\Lambda}$ is defined by

$$\Lambda_{ij} \equiv \left. \frac{\partial^2 F([x, y]; \mathbf{\Pi})}{\partial x_i \partial x_j} \right|_{[x] = [\hat{x}]([y]; \mathbf{\Pi})}. \quad (16)$$

Note that Eq. (15) is exact because $F([x, y]; \mathbf{\Pi})$ is quadratic with respect to $[\hat{x}]$. Therefore, the WKB method gives the exact result for the present model. We easily evaluate the remaining path integral by the technique of Wiegel[26, 28, 29] with slight modifications.[30]

As a result, the Hamiltonian is expressed as

$$\mathcal{H}_\tau(\mathbf{\Pi})[y] = \frac{s([\hat{x}, y]; \mathbf{\Pi})}{k_B T} + \frac{\sqrt{\phi^2 + \chi^2}}{2}$$
$$- \frac{\ln(\Gamma/4\pi k_B T \Delta t)}{2\Delta t} + \frac{c(\hat{x}_0, y_0; \mathbf{\Pi})}{\tau}, \quad (17)$$

where $\phi^2 \equiv (k + h)^2/\gamma^2$ and $\chi^2 \equiv h^2/\gamma \Gamma$. $s([\hat{x}, y]; \mathbf{\Pi})$ is the time-normalized action functional defined as $s([\hat{x}, y]; \mathbf{\Pi}) \equiv \tau^{-1} S([\hat{x}, y]; \mathbf{\Pi})$, and

$$c(\hat{x}_0, y_0; \mathbf{\Pi}) \equiv -\ln\left(\sqrt{\frac{kh}{2\pi k_B T(k + h)}} \frac{2\sqrt{\phi + \sqrt{\phi^2 + \chi^2}}}{\phi + \sqrt{\phi^2 + \chi^2}}\right)$$
$$+ \frac{H(\hat{x}_0, y_0; \mathbf{\Pi})}{k_B T}. \quad (18)$$

Since $c(\hat{x}_0, y_0; \mathbf{\Pi})$ does not depend on τ, the last term in Eq. (17) becomes irrelevant when τ is sufficiently large. In this manner, we obtain the expression for the Hamiltonian.

IV. RESULTS OF THE NUMERICAL EXPERIMENTS

A. Numerical results for finite τ cases

Now, we numerically evaluate the model described above. First, we numerically integrate the model equations (6) and (7) from $t = 0$ to $t = \tau$ under the condition of $\mathbf{\Pi}^*$, and we obtain the true trajectory set [x, y]. Here, we assume that [y]

is the only observable part of the system, and our task is to estimate $\mathbf{\Pi}$ from [y]. Next, we temporarily set $\mathbf{\Pi}$ at random. This serves as the starting point to search for a better estimation of $\mathbf{\Pi}$. Then, on the basis of the obtained trajectory [y] alone, we seek a better $\mathbf{\Pi}$ in accordance with the minimization of $\mathcal{H}_\tau(\mathbf{\Pi})[y]$.

We apply different initial settings of $\mathbf{\Pi}$ to a single datum of [y] at random and optimize them independently. We name the index number of these different settings as the replica number i and we denote each of these optimized parameters as $\mathbf{\Pi}_i^{\ddagger}$. Then, we select the $\mathbf{\Pi}_i^{\ddagger}$ that possesses the minimal value of $\mathcal{H}_\tau$ as the estimate denoted as $\hat{\mathbf{\Pi}}$. In addition, to investigate the effect of the variety in [y], we also examine different [y]. We name the index number of [y] as the sample number j and we specify this dependency as $\mathbf{\Pi}_{i,j}^{\ddagger}$ and $\hat{\mathbf{\Pi}}_j$. For more details on this numerical experiment, see the Appendix B.

Figure 2(a) displays the systematic error of the estimate. As this figure shows, when Γ^* increases and h^* decreases, the error gradually increases. Therefore, to obtain a good estimate of the system parameters, the probe should be small and tightly coupled to the hidden part. This is consistent with our intuition.

Next, we investigated the global structure of $\mathcal{H}_\tau(\mathbf{\Pi})$ in the parameter space. If the landscape of $\mathcal{H}_\tau(\mathbf{\Pi})$ is smooth and $\mathcal{H}_\tau(\mathbf{\Pi})$ has a steep valley around $\mathbf{\Pi}^*$, we can easily obtain the precise estimate. In contrast, if the landscape is rough and a lot of local minima exist, we have difficulty obtaining the precise estimate. To clarify the roughness of $\mathcal{H}_\tau(\mathbf{\Pi})$ in the parameter space, we employ the sample-averaged variance of $\mathbf{\Pi}^{\ddagger}$ (see the caption of Fig. 2) instead of normal variance. That is because the global minimum point of $\mathcal{H}_\tau(\mathbf{\Pi})$ should vary between different [y], and thus when we evaluate normal variance, which has a large finite value in both cases. Instead, the sample-averaged variance yields almost zero only in the former case. Therefore, we can ascertain the roughness of $\mathcal{H}_\tau(\mathbf{\Pi})$ from the sample-averaged variance.
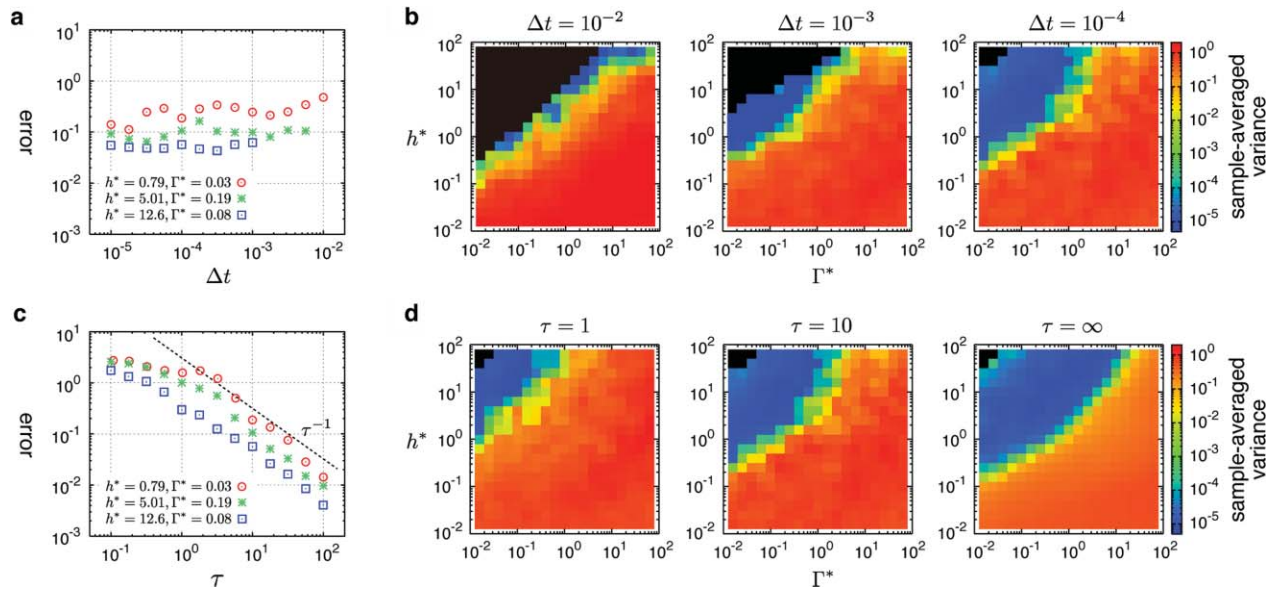
FIG. 3. Dependence on Δt and τ. The plots in (a) and (c) display typical examples of the dependence of the error on Δt and τ, respectively, in the precise phase [except $\tau \leq 10^0$ for $h^* = 0.79$, $\Gamma^* = 0.03$ and $h^* = 5.01$, $\Gamma^* = 0.19$ in (c)]. We do not plot several data points in (a) where $\Delta t = 10^{-2}$ for $h^* = 5.01$, $\Gamma^* = 0.19$ and $\Delta t > 10^{-3}$ for $h^* = 12.6$, $\Gamma^* = 0.08$ because all these estimates diverge (divergent phase). The plots in (b) and (d) display the dependence of the sample-averaged variance on Δt and τ, respectively. For the dependence on Δt [(a) and (b)], τ is fixed at 10^1, and Δt is varied as shown in these figures. Similarly, for the dependence on τ [(c) and (d)], Δt is fixed at 10^{-4} and τ is varied. In the limit of $\tau \to \infty$ in (d), the action density in Eq. (12) is explicitly calculated from the fact that the temporal average in the expression of the action can be replaced with the ensemble average. Then, by taking the limit $\tau \to \infty$ of Eq. (17), we obtain $\mathcal{H}_\infty(\Pi)$. By using $\mathcal{H}_\infty(\Pi)$, we can evaluate the estimates in the same manner as at finite τ (Ref. 30). In all cases, the optimization protocol is the same as in Fig. 2. The total replica number r and the total sample number s for these plots are as follows. For (a) and (c), we adopt $r = 1$ for all data points because as Fig. 2(c) shows, the variability of the estimates obtained from the same [y] should be small in the precise phase. The s-values for (a) and (c) are: $s = 100$ for $\Delta t \geq 10^{-4}$ and $\tau \leq 10^1$, and $s = 30$ for $\Delta t < 10^{-4}$ and $\tau > 10^1$. For (b), we adopt $r = 10$ for all cases and s values of: $s = 25$ for $\Delta t = 10^{-2}$, $s = 11$ for $\Delta t = 10^{-3}$, and $s = 3$ for $\Delta t = 10^{-4}$. For (d), we adopt $r = 10$ for all cases and s values of: $s = 9$ for $\tau = 1$ and $s = 3$ for $\tau = 10$.

Figure 2(b) shows our sample-averaged variance result. When Γ^* increases and h^* decreases, the sample-averaged variance increases not gradually but abruptly by more than four orders of magnitude, at a certain curve. Figure 2(c) displays typical relaxation courses of the sample-averaged variance in the optimization process. In Fig. 2(b), in the region where Γ^* is small and h^* is large, the sample-averaged variance decreases step-by-step and almost converges to a tiny value. We also confirm that in this region, the error converges. In contrast, in the region where Γ^* is large and h^* is small, the sample-averaged variance shows no visible sign of convergence.

Our careful investigation of the relaxation property confirms that we can classify the parameter space into three distinct regions [see Fig. 2(b)]. When Γ^* is small and h^* is large, $\hat{\Pi}$ converges almost uniformly ($\Pi_i^\ddagger$ is independent of i). From the result in Fig. 2(a), we can obtain a precise estimate in this region (precise phase). On the other hand, below the critical line, the effect of noise originating from [y] can no longer be ignored, and the landscape of $\mathcal{H}_\tau$ is roughened. Therefore, the minimization process of $\mathcal{H}_\tau$ greatly affects $\hat{\Pi}$ and the estimate is unreliable in this region (nonprecise phase). In the small region in the upper left corner of Fig. 2(c), the estimates always diverged (divergent phase). The next section gives the reasons for this behavior.

B. Dependence on Δt and τ

We next address the dependence on Δt and τ. For experiments, it is particularly important to clarify how the precise phase shifts and how the systematic error of the estimate decreases with increasing time resolution and observation time. The results are depicted in Fig. 3.

As Figs. 3(b) and 3(d) show, the critical line between the divergent and precise phases seems to depend only on Δt. This is quite natural because the critical line is directly coupled to the relaxation time scale of the probe: $\tau_r \sim \Gamma^*/h^*$. To obtain the estimate, Δt should be smaller than τ_r; otherwise, we cannot follow the motion of the probe. In contrast, the critical line between the precise and nonprecise phases depends only on τ. This line should also be related to the relaxation time scale of the probe τ_r because it partly specifies the characteristic duration needed to collect enough information on the system.

However, note that even in the limit of $\tau \to \infty$, the critical line still seems to exist. In this limit, the fluctuations inherent in [y] must vanish. Therefore, the result should be interpreted as follows. The landscape of $\mathcal{H}_\infty(\Pi)$ becomes flattened around the bottom when $\{h^*, \Gamma^*\}$ crosses over a certain threshold. In this case, the optimization process continues to seek a better estimate, but cannot find a small dip in the flattened bottom. Thus, the estimates vary. As a matter of course,

the nature of this line differs from that of finite τ (Fig. 4), this result implies that once we choose inappropriate values for the probe size and the stiffness of the linker, it abruptly becomes hard to obtain a precise estimate even if we collect data for an almost infinite amount of time with super-high temporal and spatial resolution.

Figures 3(a) and 3(c) display the dependence of the systematic error on Δt and τ, respectively. These plots clearly suggest that the error of the estimate depends on τ, while interestingly, it is almost independent of Δt. All of the data are not shown, but we have confirmed that these properties hold at various $\{h^*, \Gamma^*\}$. As Fig. 3(c) shows, when τ increases, the error gradually decreases and the slope approaches τ^{-1}. In addition, we numerically verify that $\mathcal{H}_\infty(\mathbf{\Pi})$ possesses a minimum at exactly $\mathbf{\Pi} = \mathbf{\Pi}^*$. This behavior is consistent with the theoretical prediction.

C. Effect of measurement noise

In practice, as well as the finiteness of the data length, we cannot avoid measurement noise. we examine the effect of the addition of the white Gaussian noise to $[y]$. Figure 5 shows the result. When the power of the noise decreases below a characteristic value, the systematic error of the estimate becomes almost independent of the noise. We also confirm that the sample-averaged variance in this case is almost zero. In contrast, if the power of the noise is larger than the threshold value, all of the estimates diverge. In summary, when we can obtain estimates and confirm that the estimates uniformly converge, we can obtain precise estimates in the presence of measurement noise.

D. Estimating the motion of the hidden part

Note that the present method does not require a correct estimate of the trajectory of the hidden part in order to precisely estimate the system parameters. Indeed, the MAP estimator $[\hat{x}]$ does not coincide with the true trajectory even after the system parameters have been correctly estimated. Figure 6 displays examples of $[\hat{x}]$ calculated from the true parameter values. As we can easily imagine, if we intend to obtain a good estimate of the trajectory of the hidden part, the probe should be tightly coupled to the hidden part.

Although this numerical result is quite natural, we can derive the same result analytically. For instance, to obtain the precise estimate of the motion of the hidden part, the MAP estimator $\hat{x}(t)$ should satisfy the following relation:

$$\langle (\hat{x}(t) - x(t))^2 \rangle \ll 1, \tag{19}$$

where $\langle \cdots \rangle \equiv \lim_{\tau \to \infty} \tau^{-1} \int_{-\tau/2}^{\tau/2} dt \cdots$. By calculating $\langle x(t)^2 \rangle$, $\langle \hat{x}(t)^2 \rangle$, and $\langle x(t)\hat{x}(t) \rangle$,[30] we finally obtain

$$\langle (\hat{x}(t) - x(t))^2 \rangle = \frac{1}{\beta h^*} \left[\frac{2}{\kappa^*} - \frac{1}{\sqrt{\kappa^{*2} + g^*}} \right], \tag{20}$$

where $\kappa^* \equiv 1 + k^*/h^*$ and $g^* \equiv \gamma^*/\Gamma^*$. Equation (19) is satisfied if $k^* \ll h^*$, which is consistent with the numerical result.
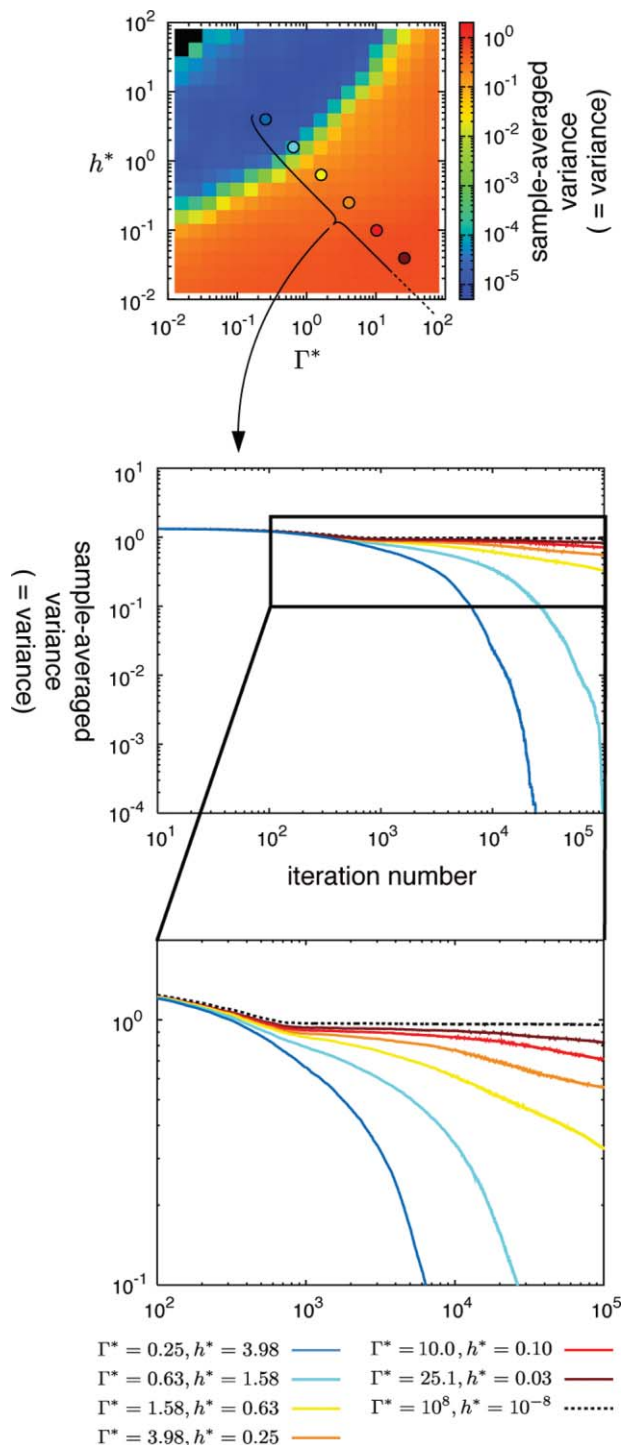

FIG. 4. Loss-of-precision transition for an infinite observation interval. Relaxation courses of the variance in the optimization process at various $\{h^*, \Gamma^*\}$. (The sample-averaged variance is consistent with the normal variance in this limit because the variety of the sample $[y]$ vanishes in the limit of $\tau \to \infty$.) The colors of the lines in the bottom figure correspond to the colors of the dots in the top figure. As these figures show, the characteristic time scale of the optimal parameter search increases when Γ^* increases and h^* decreases. However, this is different from the case of finite τ [see Fig. 2(c)]. The numerical result obtained from $\Gamma^* \gg 1$, $h^* \ll 1$ (broken line in the bottom figure) indicates that the relaxation time scale of the optimization process gradually increases and diverges to ∞ in the limit of $\Gamma^*/\gamma^* \to \infty$ and $h^*/k^* \to 0$. This seems natural because the effect of the noise arising from $[y]$ becomes irrelevant and the landscape of $\mathcal{H}_\tau(\mathbf{\Pi})$ becomes smooth in the limit of $\tau \to \infty$, thus the critical line of the loss-of-precision transition should disappear. In all the cases, $\Delta t = 10^{-4}$ and $r = 200$.
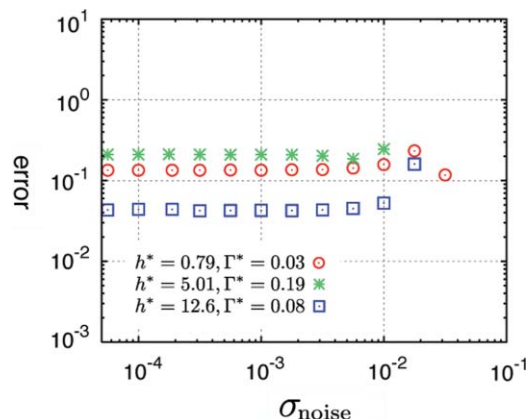
FIG. 5. Effect of contamination by measurement noise. The plot shows the dependence of the systematic error on the standard deviation (SD) of the noise, σ_{noise}. We select all $\{h^*, \Gamma^*\}$ from the precise phase with no measurement noise. We do not plot several data points where $\sigma_{\text{noise}} \geq 5.62 \times 10^{-1}$ for $h^* = 0.79$, $\Gamma^* = 0.03$, $\sigma_{\text{noise}} \geq 1.78 \times 10^{-1}$ for $h^* = 5.01$, $\Gamma^* = 0.19$, and $\sigma_{\text{noise}} \geq 3.16 \times 10^{-1}$ for $h^* = 12.6$, $\Gamma^* = 0.08$ because all these estimates diverge. In all cases, $\Delta t = 10^{-4}$, $\tau = 10$, $r = 10$, and $s = 3$. The system of units is the same as Fig. 2, therefore, the SD of the fluctuation of $x(t)$ is 1.

E. Comparison with the conventional method

Since the present model is simple enough, we can directly calculate the power spectrum of $[y]$ from the original model.[30] Therefore, following a conventional procedure, we can also estimate the system parameters by fitting the analytical solution to the observed spectrum of $[y]$. Figure 7 compares the performance of the present method with that of spectrum fitting. It clearly shows that the present method performs better in almost all cases. Indeed, in the region that overlaps the precise phase [see Fig. 2(c)], the present method is more accurate than the conventional method by 1–2 orders of magnitude.

V. CONCLUSIONS AND DISCUSSION

In the present paper, we studied the fundamental aspects of the problem of parameter estimation in the presence of hidden degrees of freedom. First, on general grounds, we formulated the problem within the framework of the Bayesian inference. This formulation provides a systematic procedure for estimating the system parameters by minimizing the Hamiltonian when an appropriate model has been input. As-
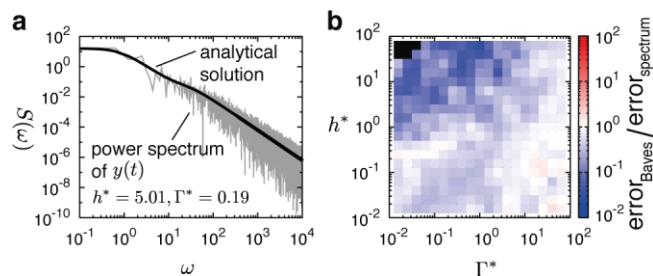

FIG. 7. Comparison with the conventional method. (a) Example of the power spectrum of $y(t)$ and the analytical solution at $\mathbf{\Pi} = \mathbf{\Pi}^*$. System parameters in this plot are $h^* = 5.01$, $\Gamma^* = 0.19$. (b) Comparison of the error in proposed and conventional methods. Error of the present method divided by error of the spectrum fitting is plotted. In the blue region, the error of the present method is smaller than that of the spectrum fitting. This is not our main concern, but we find that the dependence of the estimates from spectrum fitting on $\{h^*, \Gamma^*\}$ has a structure similar to that of the proposed method. We also compare the error of the estimate at a normal scale and find features similar to those visible at a logarithmic scale. Optimization protocol, Δt, τ, r, and s are the same as in Fig. 2.

suming that the original system has an ergodic property and that there is no contamination of the measurement noise, we proved that the distance between the true value of the system parameters ($\mathbf{\Pi}^*$) and the minimum point of the Hamiltonian ($\hat{\mathbf{\Pi}}$) decreases in proportion to the observation time length (τ), and the Hamiltonian exhibits a minimum at exactly the correct value of the system parameters in the limit of infinitely large data length ($\tau \to \infty$). It should be noted that even though the derivation procedure is different, a similar asymptotic behavior of the maximum likelihood estimator has been known.[31,32]

However, a careful numerical study of the finite-time case revealed the presence of a striking phenomenon: a loss-of-precision transition occurred at a certain curve in the space of the original system parameters. In the present case, when the size of the probe increases and the stiffness of the coupling spring decreases to a certain curve in the parameter space, the precision of the estimation is abruptly lost. In this phase, the roughness originating from noise in the time series of the data $[y]$ dominates the landscape of $\mathcal{H}_\tau$ and numerous minima appear around the bottom of the Hamiltonian. Thus, the inferred values of the system parameters vary over orders of magnitude. As the observation time increases, the noise inherent in $[y]$ becomes negligible, and the primary structure of $\mathcal{H}_\tau$
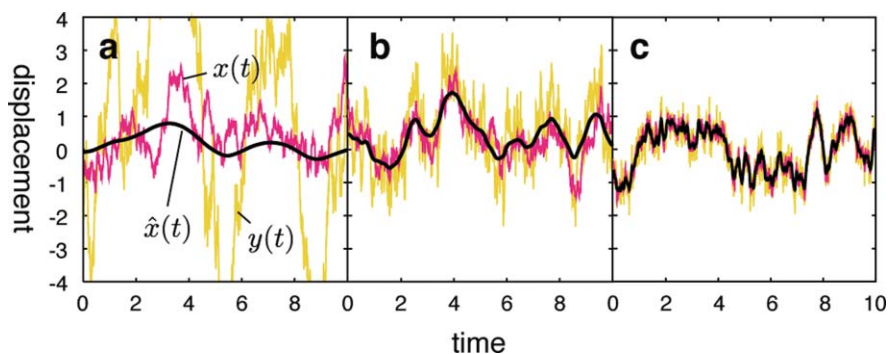

FIG. 6. Examples of the estimated trajectory of the hidden part. We plot the trajectory of the visible part $[y]$, true trajectory of the hidden part $[x]$, and MAP estimator of the hidden part $[\hat{x}]$ obtained from the correct system parameters. In these plots, Γ^* is fixed at 0.12 and h^* varies as (a) 0.12, (b) 1.25, and (c) 12.5. In all cases, the system of units is the same as Fig. 2, and $\Delta t = 10^{-4}$.

appears. In this limit, the landscape of $\mathcal{H}_\tau$ around the bottom becomes almost completely flattened below the critical curve. This is why the critical curve shifts as the observation time increases, but it seems to exist even in the limit of infinite observation time.

The unexpected presence of such a transition suggests a limit on the choice of experimental conditions. In the model studied here, the transition occurs when the size of the probe increases or the stiffness of the linker coupled to the motor decreases. This implies that if inappropriate values are chosen for these parameters, the entire system might enter the nonprecise phase, which would make the precise and accurate estimation of the system parameters hopeless.

In contrast, if the system is in the precise phase, the present method provides an accurate and precise estimate of the system parameters in the presence of hidden degrees of freedom, the error of which decreases in proportion to the observation time. Compared to the conventional method of spectrum fitting, if the appropriate model for the system has been selected, the present method yields an estimate that is up to two orders of magnitude more accurate. This is partly because the present method naturally incorporates the dynamic structure of the system. In addition, considering an actual application to single-molecule experiments, we clarified the effect of the contamination of measurement noise. We found that the noise does not affect the preciseness of our estimate if the power of the noise is small enough. The Bayesian inference thus provides a powerful framework for interpreting experimental results.

For this paper, we chose a simple yet nontrivial example to clarify the basic features of the proposed method. However, because the Onsager–Machlup functional integral approach has broad applications to over-damped Langevin equations,[22,26] the proposed method can be applied to a variety of complex models. For instance, we can model the motion of particles in three dimensional space by increasing the number of degrees of freedom. In addition, we can increase the number of particles to describe the structure of proteins more precisely within the framework of elastic network models. We can also incorporate nonlinear interactions, including metastable potentials,[33] into the model. Although, in these cases exact calculations of the marginal probability become difficult, we expect that the WKB method would give a reasonable approximation of the marginal likelihood under suitable conditions.[26,33] We could also incorporate numerical methods such as the Markov-chain Monte Carlo method.[15,16] Further studies in this direction will enables us to examine more complex systems.

In addition to the parameter estimation, the choice of the model itself is an important problem. For this problem, if we have several candidates for the model, we can adopt either an empirical or a hierarchical Bayesian approach to determine the most likely model.[15,16] Suppose we have several models to test, then, by minimizing the Hamiltonian for each model, and by comparing the minimum values of the Hamiltonian or the marginalized Hamiltonian with respect to the model parameters, we could determine the most likely model.

Once we determine the model, we need to choose an appropriate condition for the experimental system. Although we examined only one model in this study, the presence of the loss-of-precision transition in the simplest model naturally suggests that such a transition is a universal feature among various complex models. Therefore, by evaluating the sample-averaged variance of the estimates, we can verify whether the present experimental condition is in the precise phase or not, and we can find an appropriate condition by testing various conditions. When we confirmed that the system is in the precise phase, we can adopt one of the many available optimization algorithms, including the expectation–maximization (EM) algorithm,[34] to improve the searching efficiencies. To obtain more precise estimates, the bias correction techniques[32] should also be helpful because we know general asymptotic dependence of the estimates on the data duration.

In actual experiments, even if we carefully optimize the experimental condition, we expect that the Hamiltonian possesses several minima because of the slow protein dynamics, drift of the measurement system, and so on. Therefore, in practice, we face the problem of how to guarantee that the minimum we locate corresponds to the correct estimate. One strategy is to change not all but only several values of the system parameters experimentally,[35] and to confirm whether only those parameter values are altered as expected. For cases in which the measurement noise is too large, the hierarchical Bayesian approach might be applicable. First, we would estimate the most probable trajectory of the observable degrees of freedom from noisy data. Second, by marginalizing these variables around the most probable trajectory, we could calculate the (marginalized) Hamiltonian. Then, in a similar manner to the example studied here, we might obtain reasonable estimates of the system parameters.

In summary, the proposed method facilitates the precise quantification of the physical parameters of the systems with the hidden degrees of freedom. Since this method enables us to extract information about a molecule from a single time series, it can be used to calibrate the attached probe on a molecule-to-molecule basis. We hope that the development of sophisticated methods of data analysis together with precise measurement techniques will bring us novel insights into the mechanics of single proteins.

APPENDIX A: GENERAL CONSEQUENCES OF THE PROPOSED METHOD

If the original system possesses an ergodic property, the long-term limit of Eq. (3) can be replaced with the ensemble average as follows:

$$\mathcal{H}_\infty(\mathbf{\Pi}) = -\lim_{\tau \to \infty} \frac{1}{\tau} \int \mathcal{D}y \, P([y]|\mathbf{\Pi}^*) \ln P([y]|\mathbf{\Pi})$$

$$\equiv -\langle \ln P([y]|\mathbf{\Pi}) \rangle_{\mathbf{\Pi}^*}. \quad (A1)$$

Then, by the Bayes' theorem [Eq. (2)] and the assumption that $P(\mathbf{\Pi})$ is a noninformative prior, we can confirm that $\mathcal{H}_\infty(\mathbf{\Pi})$ indeed exhibits a minimum at $\mathbf{\Pi} = \mathbf{\Pi}^*$ as follows. First,

$$\left. \frac{\partial \mathcal{H}_\infty(\mathbf{\Pi})}{\partial \mathbf{\Pi}} \right|_{\mathbf{\Pi}=\mathbf{\Pi}^*} = -\lim_{\tau \to \infty} \frac{1}{\tau} \frac{\partial}{\partial \mathbf{\Pi}^*} \int \mathcal{D}y \, P([y]|\mathbf{\Pi}^*)$$

$$= \mathbf{0}. \quad (A2)$$

Next, the components of the Hessian matrix of $\mathcal{H}(\mathbf{\Pi})$ evaluated at $\mathbf{\Pi} = \mathbf{\Pi}^*$, which is denoted by H, are calculated as

$$H_{ij} \equiv \left. \frac{\partial^2 \mathcal{H}_\infty(\mathbf{\Pi})}{\partial \Pi_i \partial \Pi_j} \right|_{\mathbf{\Pi}=\mathbf{\Pi}^*}$$

$$= \left\langle \frac{\partial \ln P([y]|\mathbf{\Pi}^*)}{\partial \Pi_i^*} \frac{\partial \ln P([y]|\mathbf{\Pi}^*)}{\partial \Pi_j^*} \right\rangle_{\mathbf{\Pi}^*}. \quad (A3)$$

Since this expression has the form of a correlation matrix, we can confirm that the Hessian is a non-negative definite. We thus find that $\mathbf{\Pi} = \mathbf{\Pi}^*$ gives a minimum of the Hamiltonian.

Furthermore, a perturbative calculation about this minimum in the series of $\varepsilon \equiv \tau_0/\tau$ (τ_0 is the characteristic timescale of the system we examine) yields a prediction for how the systematic error will decrease with increasing τ. First, we note that the Hamiltonian can be expanded into a series of ε:

$$\mathcal{H}_\tau(\mathbf{\Pi})[y] \equiv -\tau^{-1} \ln P([y]|\mathbf{\Pi})$$

$$= \phi_0(\mathbf{\Pi}) + \varepsilon \phi_1(\mathbf{\Pi}) + O(\varepsilon^2). \quad (A4)$$

Note that the zeroth-order term is identical to $\mathcal{H}_\infty(\mathbf{\Pi})$ because

$$\phi_0(\mathbf{\Pi}) \equiv \lim_{\tau \to \infty} \mathcal{H}_\tau(\mathbf{\Pi}) = \mathcal{H}_\infty(\mathbf{\Pi}). \quad (A5)$$

Next, we expand the solution of the estimate as

$$\hat{\mathbf{\Pi}} = \hat{\mathbf{\Pi}}_0 + \varepsilon \hat{\mathbf{\Pi}}_1 + O(\varepsilon^2), \quad (A6)$$

which should satisfy

$$\left. \frac{\partial \mathcal{H}_\tau(\mathbf{\Pi})}{\partial \mathbf{\Pi}} \right|_{\mathbf{\Pi}=\hat{\mathbf{\Pi}}} = \mathbf{0}. \quad (A7)$$

By substituting Eq. (A6) into Eq. (A7), a standard perturbative calculation leads to the following set of equations:

$$\left. \frac{\partial \phi_0(\mathbf{\Pi})}{\partial \mathbf{\Pi}} \right|_{\mathbf{\Pi}=\hat{\mathbf{\Pi}}_0} = \mathbf{0}, \quad (A8)$$

$$\left. \frac{\partial^2 \phi_0(\mathbf{\Pi})}{\partial \mathbf{\Pi}^2} \right|_{\mathbf{\Pi}=\hat{\mathbf{\Pi}}_0} \cdot \hat{\mathbf{\Pi}}_1 + \left. \frac{\partial \phi_1(\mathbf{\Pi})}{\partial \mathbf{\Pi}} \right|_{\mathbf{\Pi}=\hat{\mathbf{\Pi}}_0} = \mathbf{0}, \quad (A9)$$

and so on. Recalling Eq. (A5) and the result that the Hamiltonian satisfies Eq. (A2), we find $\hat{\mathbf{\Pi}}_0 = \mathbf{\Pi}^*$ from Eq. (A8). Therefore, Eq. (A9) yields

$$\hat{\mathbf{\Pi}}_1 = -\left[H^{-1} \cdot \left. \frac{\partial \phi_1(\mathbf{\Pi})}{\partial \mathbf{\Pi}} \right|_{\mathbf{\Pi}=\mathbf{\Pi}^*} \right], \quad (A10)$$

under our assumption that this Hessian is a positive definite rather than a non-negative definite.

In summary, from Eqs. (A6) and (A10), we obtain the expression of the estimate in the case of finite τ as

$$\hat{\mathbf{\Pi}} = \mathbf{\Pi}^* - \varepsilon \left[H^{-1} \cdot \left. \frac{\partial \phi_1(\mathbf{\Pi})}{\partial \mathbf{\Pi}} \right|_{\mathbf{\Pi}=\mathbf{\Pi}^*} \right] + O(\varepsilon^2), \quad (A11)$$

where $\phi_1(\mathbf{\Pi})$ can be defined as

$$\phi_1(\mathbf{\Pi}) \equiv \lim_{\varepsilon \to 0} [\mathcal{H}_\tau(\mathbf{\Pi})[y] - \mathcal{H}_\infty(\mathbf{\Pi})]/\varepsilon. \quad (A12)$$

Since the first-order term in Eq. (A11) depends only on $\mathbf{\Pi}^*$ and not on τ, the error of the estimate decreases in proportion to τ. Although we defined the error using a logarithmic scale in the main text, we can easily prove from Eq. (A11) that it has the same features as in a normal scale.

APPENDIX B: NUMERICAL EXPERIMENT

We adopt the zero-temperature Metropolis method for minimizing $\mathcal{H}_\tau$. Because the system parameter should be a positive number, we optimize $\mathcal{H}_\tau$ in a logarithmic scale. The procedure we use follows. (A) We randomly generate an initial condition of the system parameter $\mathbf{\Pi}^0$ as $\log \Pi_i^0 = \log \Pi_i^* + U_i$, and we evaluate $\mathcal{H}_\tau$ using $\mathbf{\Pi}^0$. Here, i denotes the index number of the system parameter and U_i represents uniform random variables with range $(-1, 1)$. (B) The system parameter changes randomly as $\log \Pi_i^{s+1} = \log \Pi_i^s + G_i$, where s means the present number of iteration steps, and G_i represents the zero-mean Gaussian random variables with standard deviation of 0.01. (C) We evaluate $\mathcal{H}_\tau$ using $\mathbf{\Pi}^{s+1}$ and if $\mathcal{H}_\tau$ decreased by this step, we store $\mathbf{\Pi}^{s+1}$, otherwise we restore $\mathbf{\Pi}^s$. We repeat (B) and (C) for 10^5 steps. In some cases, $\| \log \hat{\mathbf{\Pi}} - \log \mathbf{\Pi}^* \|$ becomes considerably large. We remove such an irrelevant result from the analysis when $\| \log \hat{\mathbf{\Pi}} - \log \mathbf{\Pi}^* \| > 3$. We also test the larger value of variance in G_i and find the same features in the results. Only the optimization efficiency decreases in those cases.

In the least-square fitting of the power spectrum of $[y]$, we only replace $\mathcal{H}_\tau$ in the source code of the Bayesian method with the fitting function for the power spectrum, and then execute the estimation program under the same settings. To avoid the effect of the aliasing of the power spectrum on the estimation results, we do not use one order of magnitude of the high-frequency component in the fitting.

[1] T. Nishizaka, K. Oiwa, H. Noji, S. Kimura, E. Muneyuki, M. Yoshida, and K. Kinosita, Jr., Nat. Struct. Mol. Biol. **11**, 142 (2004).
[2] T. Ishii, Y. Murayama, A. Katano, K. Maki, K. Kuwajima, and M. Sano, Biochem. Biophys. Res. Commun. **375**, 586 (2008).
[3] J. P. Junker, F. Ziegler, and M. Rief, Science **323**, 633 (2009).
[4] K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block, Nature (London) **365**, 721 (1993).
[5] H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita, Jr., Nature (London) **386**, 299 (1997).

[6] K. Kitamura, M. Tokunaga, A. H. Iwane, and T. Yanagida, Nature (London) **397**, 129 (1999).

[7] K. Shiroguchi and K. Kinosita, Jr., Science **316**, 1208 (2007).

[8] K. Hu, L. Ji, K. T. Applegate, G. Danuser, and C. M. Waterman-Storer, Science **315**, 111 (2007).

[9] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).

[10] J. W. J. Kerssemakers, E. L. Munteanu, L. Laan, T. L. Noetzel, M. E. Janson, and M. Dogterom, Nature (London) **442**, 709 (2006).

[11] O. Flomenbom, J. Klafter, and A. Szabo, Biophys. J. **88**, 3780 (2005).

[12] A. Baba and T. Komatsuzaki, Proc. Natl. Acad. Sci. U.S.A. **104**, 19297 (2007).

[13] C. B. Li, H. Yang, and T. Komatsuzaki, Proc. Natl. Acad. Sci. U.S.A. **105**, 536 (2008).

[14] H. Akaike, "Likelihood and the Bayes Procedure," in *Bayesian Statistics*, edited by J. M. Bernardo, M. H. D. Groot, D. V. Lindley, and A. F. M. Smith (University Press, Valencia, 1980), pp. 143–166.

[15] C. P. Robert, *The Bayesian Choice: A Decision-Theoretic Motivation* (Springer-Verlag, New York, 1994).

[16] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer-Verlag, New York, 2006).

[17] We naturally assume that the prior probability distribution of $[x]$ is uniform. In this case, the MAP estimator is identical to a maximum likelihood estimator (MLE).

[18] W. Bialek, C. G. Callan, and S. P. Strong, Phys. Rev. Lett. **77**, 4693 (1996).

[19] R. Yasuda, H. Noji, M. Yoshida, K. Kinosita, Jr., and H. Itoh, Nature (London) **410**, 898 (2001).

[20] H. C. Berg, in *Random Walks in Biology* (Princeton University, Princeton, NJ, 1993), Chap. 4

[21] K. Sekimoto, *Stochastic Energetics* (Springer-Verlag, Berlin Heidelberg, 2010), Chap. 1.

[22] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953); *ibid.* **91**, 1512 (1953).

[23] R. Graham, Z. Physik B **26**, 281 (1976).

[24] R. L. Stratonovich, SIAM J. Control **4**, 362 (1966).

[25] R. Mortensen, J. Stat. Phys. **1**, 271 (1969).

[26] K. L. C. Hunt and J. Ross, J. Chem. Phys. **75**, 976 (1981).

[27] Using the concrete expression for the path probability [Eqs. (8)–(12)], Eq. (5) can be explicitly written as a tridiagonal system of linear equations. Therefore, the exact solution of the MAP estimator, $[\hat{x}](y; \boldsymbol{\Pi})$, can be easily obtained.

[28] F. W. Wiegel, Physica **33**, 734 (1967).

[29] F. W. Wiegel, Physica **37**, 105 (1967).

[30] See supplementary material at http://dx.doi.org/10.1063/1.3516587 for the detailed derivation of the Hamiltonian, the analytical solution of the power spectrum, the time correlation of $x(t)$ and $\hat{x}(t)$, and the Hamiltonian in the infinite interval limit.

[31] D. R. Cox and E. J. Snell, J. R. Stat. Soc. Ser. B (Methodol.) **30**, 248 (1968).

[32] D. R. Cox and D. V. Hinkley, in *Theoretical Statistics* (Chapman and Hall, London (1974)), Chap. 8.

[33] B. Caroli, C. Caroli, and B. Roulet, J. Stat. Phys. **26**, 83 (1981).

[34] A. Dempster, N. Laird, and D. Rubin, J. R. Stat. Soc. Ser. B **39**, 1 (1977).

[35] In the case of the single-molecule experiment of F_1-ATPase, the value of Γ^* is easily changed through controlling viscosity of the solution. h^* can be controlled by the length of the linker. γ^* and k^* also will be changed by specific ligands such as ADP or ATP analogs, or the mutation.