HOKKAIDO UNIVERSITY

Title	Aggregated Markov Model Using Time Series of Single Molecule Dwell Times with Minimum Excessive Information
Author(s)	Li, Chun-Biu; Komatsuzaki, Tamiki
Citation	Physical Review Letters, 111(5), 58301 https://doi.org/10.1103/PhysRevLett.111.058301
Issue Date	2013-08-01
Doc URL	https://hdl.handle.net/2115/53360
Rights	©2013 American Physical Society
Type	journal article
File Information	PhysRevLett.111.058301.pdf



Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Aggregated Markov Model Using Time Series of Single Molecule Dwell Times with Minimum Excessive Information

Chun-Biu Li and Tamiki Komatsuzaki*

Molecule and Life Nonlinear Sciences Laboratory, Research Institute for Electronic Science, Hokkaido University, Kita 20 Nishi 10, Kita-ku, Sapporo 001-0020, Japan

(Received 6 December 2012; published 1 August 2013)

Statistics of the dwell times, the stationary state distributions (SSDs), are often studied to infer the underlying kinetics from a single molecule finite-level time series. However, it is well known that the underlying kinetic scheme, a hidden Markov model (HMM), cannot be identified uniquely from the SSDs because some features of the underlying HMM are hidden by finite-level measurements. Here, we quantify the amount of excessive information in a given HMM that is not warranted by the measured SSDs and extract the HMM with minimum excessive information as the most objective representation of the data. The method is applied to a single molecule enzymatic turnover experiment, and the origin of dynamic disorder is discussed in terms of the network properties of the HMM.

DOI: 10.1103/PhysRevLett.111.058301 PACS numbers: 82.37.−j, 02.50.Ga, 02.50.Tt, 05.45.Tp

A discrete finite-level time series is often observed in single molecule experiments to extract the dynamical and kinetic information on the working mechanisms of a single (or multiple) biomolecule(s) [1–7]. The underlying dynamics and kinetics from the data are usually represented by a finite-state hidden Markov model (HMM) whose state and transition properties are inferred from the dwell-time statistics of the levels. A common feature of these measurements is that the system can visit several different states (e.g., on the energy landscape [8–11]) while the observed time series stays on the same level. This means that some states are aggregated and some transitions are masked by the measurement. It is well known [12,13] that there can exist more than one HMM to reproduce all the observed stationary state distributions (SSDs) of the dwell times; i.e., the underlying HMM is, in general, *unidentifiable* from the dwell-time time series.

Let us consider a two-level time series with two symbols (a and b). The full statistical information of the time series is specified by the set of SSDs, which has the general form of $f_{\alpha_1\alpha_2\cdots}(t_{\alpha_1}, t_{\alpha_2}, \ldots)$ (with $\alpha_1, \alpha_2, \ldots = a, b$ and $\alpha_i \neq \alpha_{i+1}$) denoting the probability of the system visiting successively the symbol α_1 for t_{α_1} steps, the symbol α_2 for t_{α_2} steps, and so on. It has been shown [12] that the full set of SSDs does not contain enough information to uniquely determine the underlying HMM. HMMs which may have different state and transition properties are said to be *equivalent* if they reproduce the same SSDs. The transition matrices of the equivalent HMMs with the same number of states (N_a and N_b) for the two levels (a and b) are related by similarity (Kienker) transformations [12,13] whose matrix elements take continuous values. Therefore, there are, in general, infinitely many equivalent HMMs for a given set of measured SSDs.

There have been a few attempts to resolve the problem of unidentifiability by imposing additional constraints on the HMM, e.g., the maximum likelihood estimation of an aggregated Markov model with a presumed network topology [14,15], the uncoupled model [12,16] in which transitions between states with the same symbol are forbidden, and the manifest interconductance rank form [17] in which the transition matrices among states with different symbols are diagonalized. However, the additional constraints imposed may result in some properties in the inferred HMMs which cannot be warranted by the data. Here, we take a different approach from the information-theoretic viewpoint by formulating a measure, termed "excessive information," to quantify the amount of extra information in a given HMM that cannot be warranted by the observation. By minimizing the number of states and the excessive information with respect to all equivalent HMMs to reproduce the same measured SSDs, we identify the HMM that contains the least amount of *a priori* assumptions as the simplest and the most objective representation of the data.

Without loss of generality, we formulate the measure of excessive information for two-symbol (a and b) time series. Suppose one obtains from a measurement the set of joint dwell-time SSDs $[f_{ab}(t_a, t_b)$ and $f_{ba}(t_b, t_a)]$ which sufficiently represent all statistical information to construct HMMs [18]; i.e., no additional information is contained in the higher order SSDs and correlations. Let $\vec{\boldsymbol{\alpha}} = \{\vec{\alpha}_1^L, \vec{\alpha}_2^L, \ldots\}$ be the set of all symbolic sequences $\vec{\alpha}_i^L$ with length L which are different realizations of the SSDs, e.g., $\vec{\alpha}_i^L = (\cdots baabbba \cdots)$ with length L. Now let us consider a particular HMM from the set of all equivalent HMMs reproducing the same SSDs as those from the measurement and having the same number of states N_a and N_b. We label its states as $(a_1, \ldots, a_{N_a})$ and $(b_1, \ldots, b_{N_b})$, and denote $\vec{\mathbf{S}} = \{\vec{S}_1^L, \vec{S}_2^L, \ldots\}$ as the set of all possible state sequences with length L that can be generated by this HMM, e.g., $\vec{S}_i^L = (\cdots b_1 a_1 a_1 b_2 b_1 b_1 a_2 \cdots)$

 © 2013 American Physical Society

corresponding to $\vec{\alpha}_i^L$. Note here that any state sequence realization $\vec{S}^L$ corresponds to only one symbolic sequence $\vec{\alpha}^L$. Conversely, given a symbolic sequence, there can exist more than one corresponding state sequence since any given symbol can associate with more than one aggregated state. For simplicity, we omit the superscript L and indices i in $\vec{\alpha}_i^L$ and $\vec{S}_i^L$ hereinafter.

An information-theoretic measure, which quantifies the average uncertainty in identifying the state sequence $\vec{S}$ when the symbolic sequence $\vec{\alpha}$ is known, is given by the conditional entropy [19]

$$H(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}}) = H(\vec{\mathbf{S}}) - H(\vec{\boldsymbol{\alpha}}) = \sum_{\vec{\alpha}} P(\vec{\alpha}) \Big[-\sum_{\vec{S}} P(\vec{S}|\vec{\alpha}) \log_2 P(\vec{S}|\vec{\alpha}) \Big] \geq 0, \quad (1)$$

where $P(\vec{\alpha})$, $P(\vec{S})$, and $P(\vec{S}|\vec{\alpha})$ are the probabilities and conditional probability for $\vec{\alpha}$ and $\vec{S}$. $H(\vec{\mathbf{S}}) = -\sum_{\vec{S}} P(\vec{S}) \log_2 P(\vec{S}) \geq 0$ and $H(\vec{\boldsymbol{\alpha}}) = -\sum_{\vec{\alpha}} P(\vec{\alpha}) \log_2 P(\vec{\alpha}) \geq 0$ are Shannon entropies which measure the amount of uncertainty in identifying a particular $\vec{S}$ for the given HMM and a particular $\vec{\alpha}$ for the measured SSDs, respectively.

The conditional entropy $H(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$ vanishes if and only if solely one specific state sequence is permitted by the given HMM for each element $\vec{\alpha}$ in the set $\vec{\boldsymbol{\alpha}}$ while all other corresponding state sequences are forbidden, i.e., $P(\vec{S}|\vec{\alpha}) = 0$ or 1. This is the maximally biased case in which the HMM considered imposes extra information to completely favor one unique state sequence over all the others. Contrarily, the value of $H(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$ increases when $P(\vec{S}|\vec{\alpha})$ becomes more uniform with respect to $\vec{S}$ for any given $\vec{\alpha}$. This corresponds to the more unbiased case in which the HMM considered imposes less extra information to favor specific state sequences over the others. Therefore, we interpret the negative of the conditional entropy $I_E(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}}) \equiv -H(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$ as the amount of *excessive information* contained in a given HMM. One can also see this explicitly from Eq. (1) that $I_E(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}}) = [-H(\vec{\mathbf{S}})] - [-H(\vec{\boldsymbol{\alpha}})]$, stating that $I_E(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$ is equal to the difference between the amount of information contained in $\vec{\mathbf{S}}$ and that of $\vec{\boldsymbol{\alpha}}$. Thus, minimizing $I_E(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$ [or maximizing $H(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$] with respect to different equivalent HMMs (i.e., those reproducing the measured SSDs with the same N_a and N_b) corresponds to minimizing the extra information contained in the HMM that is not present in the measured SSDs.

Once the symbolic sequences are obtained from experiment, $P(\vec{\alpha})$ is fixed and $H(\vec{\boldsymbol{\alpha}})$ does not depend on which equivalent HMM is chosen. Thus, Eq. (1) tells us that maximizing $H(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$ with respect to different equivalent HMMs is the same as maximizing the Caliber (or path entropy) [20–23] $H(\vec{\mathbf{S}})$ of the state sequences. Since equivalent HMMs have the same SSDs, this corresponds to maximizing the Caliber [24,25] with the SSDs $f_{ab}(t_a, t_b)$ and $f_{ba}(t_b, t_a)$ as constraints. In practice, it is more expedient to maximize $H(\vec{\mathbf{S}})$ instead of $H(\vec{\mathbf{S}}|\vec{\boldsymbol{\alpha}})$ since the maximization of $H(\vec{\mathbf{S}})$ does not need to refer to the symbolic sequences $\vec{\boldsymbol{\alpha}}$ once the set of equivalent HMMs is declared.

Further simplification can be performed by decomposing $H(\vec{\mathbf{S}})$ using the Markovian property of the HMMs into

$$H(\vec{\mathbf{S}}) = H(\mathbf{S}) + (L-1) H(\mathbf{S}'|\mathbf{S}), \quad (2)$$

where $H(\mathbf{S}) = -\sum_{i=1}^{N_a+N_b} P(S_i) \log_2 P(S_i)$ is the Shannon entropy of the states, and

$$H(\mathbf{S}'|\mathbf{S}) = \sum_{j=1}^{N_a+N_b} P(S_j) \Big[-\sum_{i=1}^{N_a+N_b} P(S_i'|S_j) \log_2 P(S_i'|S_j) \Big] \quad (3)$$

is the conditional entropy of the state-to-state transitions in the HMM [26] with $P(S_i'|S_j)$ denoting the transition probability from S_j to S_i'. We consider the minimization of the excessive information for arbitrarily long (L large) state sequences, and one can see that the first term on the right-hand side of Eq. (2) becomes negligible. Therefore, maximizing $H(\vec{\mathbf{S}})$ is the same as maximizing $H(\mathbf{S}'|\mathbf{S})$. Moreover, it has been shown recently [27] that $H(\mathbf{S}'|\mathbf{S})$ is a natural information-theoretic measure for the complexity of a topographical feature, namely, the diversity in the state-to-state transitions, of Markovian networks. In the context of dynamical systems theory, one also notices that $H(\mathbf{S}'|\mathbf{S})$ is the Kolmogorov-Sinai entropy [28] of the HMM, which characterizes the dynamical randomness of a Markov process [29]. Therefore, the HMM with the minimum excessive information reproducing all the measured SSDs represents the model that generates the most "stochastic" state sequence with a minimal complexity of pattern (or syntax).

Figure 1 shows an example to outline our procedures in identifying the most unbiased HMM from the measured SSDs, and the details are given in the Supplemental Material [30]. Let us consider a two-symbol time series generated by an underlying true HMM with two aggregated states for each symbol [Fig. 1(a)]. Suppose that we have the SSDs [$f_{ab}(t_a, t_b)$ and $f_{ba}(t_b, t_a)$] constructed from the dwell-time time series. Because of its simplicity, we first extract the uncoupled model with minimum required number of states to reproduce the measured SSDs, which will serve as an initial HMM for the maximization of $H(\mathbf{S}'|\mathbf{S})$ [Eq. (3)] to obtain the most unbiased HMM. It can be easily shown that the minimum number of aggregated states is equal to the minimum number of exponentials needed to describe the dwell-time distributions [$f_a(t_a)$ and $f_b(t_b)$] [31]. In this work, we employed a multiexponential curving-fitting method, the so-called Padé-Laplace method [32], which does not require any

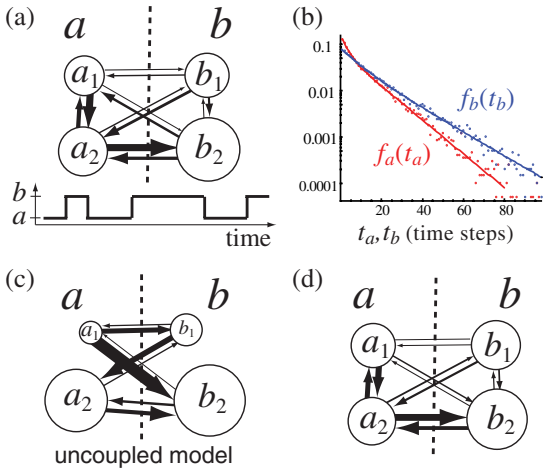


FIG. 1 (color online). An example of how to obtain the most unbiased HMM from a time series. (a) A two-symbol time series generated by the underlying true HMM. (b) The dwell-time distributions $f_a(t_a)$ (red dots) and $f_b(t_b)$ (blue dots) and the multiexponential fitting determined by the Padé-Laplace method (solid lines). (c) The uncoupled model served as an initial guess for the maximization of $H(\mathbf{S}'|\mathbf{S})$ in Eq. (3) with respect to the equivalent HMMs. (d) The unbiased HMM containing only the information from the SSDs. Note that all HMMs in this figure produce the same SSDs, but different HMMs contain different amounts of additional information. In the HMM, the state and transition probabilities are represented by the circle area and the arrow thickness, respectively. For clarity, self-transitions of the states are not drawn.

a priori assumption on the number of components, to determine the minimum number of exponentials in $f_a(t_a)$ and $f_b(t_b)$. After determining the number of exponentials, the actual fitting of the exponent values is performed by least-square methods.

Next we extract the transition matrix elements of the uncoupled model by using the singular value decomposition of $f_{ab}(t_a, t_b)$ and $f_{ba}(t_b, t_a)$ (see the Supplemental Material [30]). Figure 1(c) shows the uncoupled model extracted from the SSDs. Finally, we obtain the equivalent HMM [which relates to the uncoupled model by a similarity (Kienker) transformation] that maximizes $H(\mathbf{S}'|\mathbf{S})$ by using the Newton-Raphson method [33]. This equals to maximizing $H(\mathbf{S}'|\mathbf{S})$ constrained by the SSDs $f_{ab}(t_a, t_b)$ and $f_{ba}(t_b, t_a)$.

Figure 1(d) shows the resulting unbiased HMM constrained only by the SSDs which resembles the underlying HMM. The small discrepancy demonstrates that in general the underlying HMM cannot be fully identified from the measured SSDs. We also note from Figs. 1(c) and 1(d) that both the state and the transition probabilities of the uncoupled model are more uneven compared to the unbiased HMM. This is because the uncoupled model is constrained not only by the SSDs but also by the extra condition that no transition is allowed among states with the same symbol. Contrarily, no such extra condition exists for the unbiased HMM, and thus its state and transition probabilities can be assigned as evenly as possible.

We apply our method to the single molecule enzymatic turnover experiment [2] on *Escherichia coli* β-galactosidase with a fluorogenic substrate, resorufin-β-D-galactopyranoside. The real time observation of a single enzyme during successive catalytic reactions has revealed the multiscale fluctuation in the turnover rate, which has been interpreted [2] as originating from transitions among a variety of conformations with different catalytic activities. Figures 2(a)–2(d) show schematically the profile of the catalytic cycle and the corresponding intensity trace probed by the experiment. The distributions of the enzymatic turnover time for different substrate concentrations $[S]$ are shown in Fig. 2(e). It is evident from the Padé-Laplace method that kinetics for $[S]$ ($\leq 50\,\mu$M) much smaller than the Michaelis constant $K_M \simeq 380\ \mu$M can be well described by a single exponential, implying that the diffusion process [from Fig. 2(a) to Fig. 2(b)] is the rate-limiting step in the catalytic cycle. The distribution becomes multiexponential at $[S] = 100\ \mu$M. It has been argued [2] that increasing $[S]$ makes the rate-limiting step shift from the diffusion process to the catalytic reaction [from Fig. 2(b) to Fig. 2(c)] in which slow transitions between conformations of the enzyme-substrate system give rise to the fluctuation of the enzymatic rates, a phenomenon that has been termed as "dynamic disorder" [2,3,34]. In addition to the multiexponential distribution, dynamic disorder is also characterized by the existence of long memory in the turnover kinetics. Figure 2(f) shows the turnover-time autocorrelation $C(k) = \sum_i[(t_i - \bar{t}) \times (t_{i+k} - \bar{t})]/\bar{t}^2$ with k the event lag and $\bar{t}$ the time average of the turnover time, at $[S] = 50$ and $100\ \mu$M. A correlation is absent for $[S] \leq 50\ \mu$M when the diffusion process is the rate-limiting step. However, memory lasting for about six turnovers (~ 50 ms) appears at $[S] = 100\ \mu$M that can be attributed to the enzyme conformational fluctuation with a long correlation ($\sim 10^{-3} - 1$ s), as reported previously at the single molecule level [9,35].

The HMMs with the minimum excessive information which reproduces all measured SSDs are drawn in Fig. 2(b) for $[S] = 50$ and $100\ \mu$M [36]. A single state is sufficient to represent the catalytic cycle dominated completely by the diffusion process for $[S] \leq 50\ \mu$M. However, at least three states (S_1, S_2, and S_3) are required to explain the turnover-time distribution in Fig. 2(e). The average residential times for S_1, S_2, and S_3 are about 12, 3.4, and 1.4 ms, respectively. By comparing these time scales with that of the state at 50 μM, found to be $\sim$14 ms, we interpret that the state S_1 is still dominated by the diffusion process, whereas S_2 and S_3 with shorter time scales start to capture contributions from the catalytic steps. To single out the characteristic features of different catalytic states free from the effect of diffusion, one can consider other

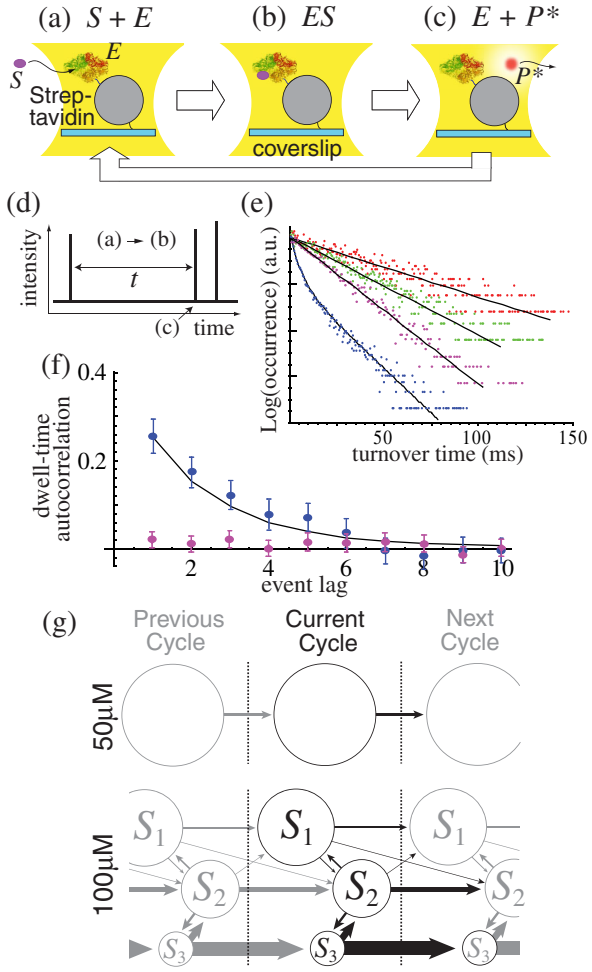


FIG. 2 (color online). The construction of the most unbiased HMM from the single molecule turnover experiment. A single enzyme (E) is immobilized on a coverslip and immersed in a buffer with constant $[S]$ in order to facilitate a long-time observation in the confocal detection volume (yellow regions). (a) The diffusion stage in which a freely moving nonfluorescent substrate "searches" the catalytic sites of the enzyme. This process is mainly governed by Brownian dynamics with a single time scale. (b) The enzyme-substrate complex is formed. (c) The bound substrate is converted into the fluorescent product P^*, then dissociates from the enzyme and quickly diffuses out from the confocal detection volume. (d) The intensity trace in which photon bursts (< 0.5 ms) appear only when a P^* is formed, and the "dark" dwell time t gives the catalytic turnover time. (e) Histograms of the turnover time (dots) in log-linear scale for different $[S]$ (from the top: red, 10 μM; green, 20 μM; purple, 50 μM; blue, 100 μM). (f) The dwell-time autocorrelation calculated from the experimental data [dots with error bars—purple (bottom), 50 μM; blue (top), 100 μM] indicates the existence of memory at 100 μM. (g) The most unbiased HMMs constructed at $[S] = 50$ μM (upper scheme) and $[S] = 100$ μM (lower scheme). Network notations are the same as those in Fig. 1. States at the current cycle and transitions originating from them are shown by dark color, whereas those from the previous and next cycles are shown in gray. The solid lines in (e) and (f) denote the exponential fitting by the Padé-Laplace method and the prediction from the constructed HMMs in (g), respectively.

experiments with $[S] \gg K_M$ in which the catalytic reaction becomes rate limiting.

The predictions of the autocorrelations from the constructed most unbiased HMMs [Fig. 2(f)] agree well with the measured ones. Figure 2(g) also shows how the memory observed at $[S] = 100$ μM is characterized by the network properties of the HMM: the transitions from the current cycle to the next one are mainly mediated by transitions among the same states (e.g., from S_2 in the current cycle to S_2 in the next one). As the three states have distinct residential times, this indicates that a short (long) dwell is more likely to be followed by a short (long) dwell which gives rise to the correlation in the turnover times.

In summary, the HMM from the set of equivalent HMMs that reproduce all observed SSDs with minimum excessive information is regarded as the most objective Markov modeling of the data, and any other equivalent HMM necessarily carries extra constraints unwarranted by the measurement. It is also shown that the minimization of excessive information results in maximizing the Caliber of the state sequences constrained by the measured SSDs [e.g., $f_{ab}(t_a, t_b)$ and $f_{ba}(t_b, t_a)$ in the case of two-level time series]. Our method complements but has a similar spirit to the canonical form of reduced dimensions [37], which discriminates HMMs by referring to the topological features of the canonical form, and the non-Markov memory kernel model [38], which can handle systems with heavy-tail statistics. The main distinctive feature of our approach is that it provides a concrete measure of excessive information which elucidates quantitatively the amount of information in the HMMs that cannot be warranted by the original data. This enables us to scrutinize which observable provides more information to construct a HMM closer to the underlying true HMM. Moreover, the non-Markovian properties of the reduced dimensions form and the non-Markov memory kernel model are reflected in the nonexponential dwell-time distribution and memory kernel of the states, whereas in our case they are manifest in the topographical properties of the constructed HMM whose state-to-state transitions are Markovian. As a result, a detailed comparison of the most unbiased HMMs associated with different observables allows us to scrutinize the origin of various dynamical features free from any unwarranted information in terms of the topographical features of the constructed networks [27,39].

We acknowledge H. Teramoto and H. Ueno for their valuable comments. We also thank X. S. Xie for providing the time series of the single molecule enzymatic turnover experiment. This work was partially supported by JSPS, JST/CREST, HFSP, a Grant-in-Aid for Research on Priority Areas "Innovative nano-science," and "Spying minority in biological phenomena," MEXT.

*tamiki@es.hokudai.ac.jp

[1] S. I. Sukharev, W. J. Sigurdson, C. Kung, and F. Sachs, J. Gen. Physiol. **113**, 525 (1999).

[2] B. P. English, W. Min, A. M. van Oijen, K. T. Lee, G. Luo, H. Sun, B. J. Cherayil, S. C. Kou, and X. S. Xie, Nat. Chem. Biol. **2**, 87 (2006).

[3] M. B. J. Roeffaers, G. D. Cremer, H. Uji-i, B. Muls, B. F. Sels, P. A. Jacobs, F. C. D. Schryver, D. E. D. Vos, and J. Hofkens, Proc. Natl. Acad. Sci. U.S.A. **104**, 12 603 (2007).

[4] F. Cichos, C. von Borczyskowski, and M. Orrit, Curr. Opin. Colloid Interface Sci. **12**, 272 (2007).

[5] R. D. Vale, T. Funatsu, D. W. Pierce, L. Romberg, Y. Harada, and T. Yanagida, Nature (London) **380**, 451 (1996).

[6] H. Noji, R. Yasuda, M. Yoshida, and J. K. Kinosita, Nature (London) **386**, 299 (1997).

[7] M. Morimatsu, H. Takagi, K. G. Ota, R. Iwamoto, T. Yanagida, and Y. Sako, Proc. Natl. Acad. Sci. U.S.A. **104**, 18 013 (2007).

[8] A. Baba and T. Komatsuzaki, Proc. Natl. Acad. Sci. U.S.A. **104**, 19 297 (2007).

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

[10] D. J. Wales, Curr. Opin. Struct. Biol. **20**, 1 (2009).

[11] Y. Levy, J. Jortner, and R. S. Berry, Phys. Chem. Chem. Phys. **4**, 5052 (2002).

[12] P. Kienker, Proc. R. Soc. B **236**, 269 (1989).

[13] H. Ito, S. Amari, and K. Kobayashi, IEEE Trans. Inf. Theory **38**, 324 (1992).

[14] D. Colquhoun and A. G. Hawkes, Proc. R. Soc. B **211**, 205 (1981).

[15] F. Qin, A. Auerbach, and F. Sachs, Proc. R. Soc. B **264**, 375 (1997).

[16] R. J. Bauer, B. F. Bowman, and J. L. Kenyon, Biophys. J. **52**, 961 (1987).

[17] W. J. Bruno, J. Yang, and J. E. Pearson, Proc. Natl. Acad. Sci. U.S.A. **102**, 6326 (2005).

[18] D. R. Fredkin and J. A. Rice, J. Appl. Probab. **23**, 208 (1986).

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[20] A. A. Filyukov and V. Y. Karpov, J. Eng. Phys. **13**, 624 (1967).

[21] A. A. Filyukov and V. Y. Karpov, J. Eng. Phys. **13**, 798 (1967).

[22] C. Monthus, J. Stat. Mech. 03 (2011) P03008.

[23] E. Smith, Rep. Prog. Phys. **74**, 046601 (2011).

[24] E. T. Jaynes, Annu. Rev. Phys. Chem. **31**, 579 (1980).

[25] G. Stock, K. Ghosh, and K. A. Dill, J. Chem. Phys. **128**, 194102 (2008).

[26] We note that the flat prior is implicitly assumed in Eqs. (1) and (3) to ensure that the state and transition probabilities of the HMM are uniform (i.e., with no preferable state and transition) when the constraints from the time series (the measured SSDs) are absent.

[27] C. B. Li, H. Yang, and T. Komatsuzaki, J. Phys. Chem. B **113**, 14 732 (2009).

[28] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinai, *Ergodic Theory* (Springer-Verlag, Berlin, 1982).

[29] P. Gaspard, *Chaos, Scattering and Statistical Mechanics* (Cambridge University Press, Cambridge, England, 1998).

[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.111.058301 for a detailed description of the HMM construction procedures.

[31] $f_a(t_a)$ and $f_b(t_b)$, are the marginals of $f_{ab}(t_a, t_b)$ and $f_{ba}(t_b, t_a)$.

[32] E. Yeramian and P. Claverie, Nature (London) **326**, 169 (1987).

[33] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 2007).

[34] J. Cao, J. Phys. Chem. B **115**, 5493 (2011).

[35] H. Yang, G. Luo, P. Karnchanaphanurach, T.-M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie, Science **302**, 262 (2003).

[36] The turnover trace in Fig. 2(d) contains only one symbol, i.e., the dark intensity level. The formulation presented here for multisymbol trajectories can be easily applied to this case by labeling the current cycle (or dark dwell) as one symbol (e.g., call it a) and the next cycle as another (e.g., b) and treating the state transitions associated with the two symbols identically (see also Supplemental Material [30]).

[37] O. Flomenbom and R. J. Silbey, Proc. Natl. Acad. Sci. U.S.A. **103**, 10 907 (2006).

[38] S. Pressé, J. Lee, and K. A. Dill, J. Phys. Chem. B **117**, 495 (2013).

[39] A. Caflisch, Curr. Opin. Struct. Biol. **16**, 71 (2006).