Markov processes follow from the principle of maximum caliber

Hao Ge,[1,2,3,a)] Steve Pressé,[4] Kingshuk Ghosh,[5] and Ken A. Dill[6]

[1]*Beijing International Center for Mathematical Research and Biodynamic Optical Imaging Center, Peking University, Beijing 100871, People's Republic of China*
[2]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA*
[3]*School of Mathematical Sciences and Centre for Computational Systems Biology, Fudan University, Shanghai 200433, People's Republic of China*
[4]*Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, USA*
[5]*Department of Physics and Astronomy, University of Denver, Denver, Colorado 80208, USA*
[6]*Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, USA*

Markov models are widely used to describe stochastic dynamics. Here, we show that Markov models follow directly from the dynamical principle of maximum caliber (Max Cal). Max Cal is a method of deriving dynamical models based on maximizing the path entropy subject to dynamical constraints. We give three different cases. First, we show that if constraints (or data) are given in the form of singlet statistics (average occupation probabilities), then maximizing the caliber predicts a time-independent process that is modeled by identical, independently distributed random variables. Second, we show that if constraints are given in the form of sequential pairwise statistics, then maximizing the caliber dictates that the kinetic process will be Markovian with a uniform initial distribution. Third, if the initial distribution is known and is not uniform we show that the only process that maximizes the path entropy is still the Markov process. We give an example of how Max Cal can be used to discriminate between different dynamical models given data. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.3681941]

I. INTRODUCTION

Dynamical processes are commonly modeled as Markov processes.[1,2] When making a dynamical model, what is the underlying justification for asserting a Markov model? We show here that Markov modeling can be justified on the grounds of the principle of maximum caliber (Max Cal), the dynamical analog of maximum entropy.[3] Max Cal provides a general foundation for stochastic dynamics in the same way that the principle of maximizing the entropy provides a foundation for equilibrium statistical physics.[4–8]

In Max Cal, one begins with knowledge of all the possible microscopic trajectories that a system could take in time. A small number of dynamical quantities are measured (far fewer than the total number of possible microscopic trajectories). These quantities can include known average macroscopic fluxes. The route entropy is then maximized subject to the dynamical constraints on these average quantities (called maximizing the caliber) to build a weighted ensemble of these microscopic trajectories consistent with the constrained averages. The ensemble is then used to infer other dynamical averages and fluctuations.[9–16] The first application of this basic idea is probably due to Filyukov and Karpov who, in 1967,[17] assumed dynamical microscopic trajectories that could be treated as Markov chains. They then maximized the caliber to parameterize the rates of their model which was assumed to be Markovian from the onset.

Here, we pose Filyukov and Karpov's problem in reverse and ask: given dynamical constraints, what models will

maximize the caliber? The answer to this question is one systematic recipe for building dynamical models. A common approach to dynamical modeling is to assert a particular kinetic model having a given set of rates as parameters. Then, the parameters of the model are fit to the experimental data. However, this type of approach has limitations: the models are often not unique, they are sometimes overparameterized, and they are not guaranteed to provide a principled description of the data beyond the known average quantities that are used as input. We are interested in answering the following question: Without invoking a hypothetical model in advance, how much can the data alone tell us about how to model the dynamics of the system underlying it?

II. THE MAXIMUM-CALIBER APPROACH

Suppose we have a stochastic process with a discrete state space $\Gamma = \{1, 2, \cdots, N\}$.

Consider its trajectories of length T and denote the probability for the trajectory $\{i_0, i_1, \cdots, i_T\}$ as $p_{i_0 i_1 \cdots i_T}$ where i_n is the state visited at time n. The path entropy, $S(T)$, is

$$S(T) = - \sum_{i_0, i_1, \cdots, i_T} p_{i_0 i_1 \cdots i_T} \log p_{i_0 i_1 \cdots i_T}. \tag{1}$$

Each variable i_n in the sum runs from 1 to N. Equation (1) is the measure which, when maximized subject to constraints, yields the least biased probability for the set of microscopic trajectories, $\{p_{i_0 i_1 \cdots i_T}\}$, consistent with observed constraints.[18,19] For a trajectory of length T = 1, $p_{1,2}$ is the probability of being in state 1 at time 0 and state 2 at time 1.

The sum Eq. (1) is over all possible combinations of states at time 0 and 1.

Furthermore, if $p_{i_0 i_1 \cdots i_T}$ is a path probability, then it must satisfy certain logical consistency requirements, as established by Kolmogorov.[20]

Specifically, for any T equal to 0 or to any positive integer, the probability distribution for a microscopic trajectory must satisfy

$$\sum_{i_{T+1}} p_{i_0 \cdots i_T i_{T+1}} = p_{i_0 \cdots i_T}, \tag{2}$$

i.e., the distribution of the process during the time interval $\{0, 1, 2, \cdots, T\}$ is the distribution of the process during the time interval $\{0, 1, 2, \cdots, T+1\}$ marginalized over $T+1$.

A. Constraint on the mean of the singlet distribution

First, we suppose we only know the mean number of time intervals during which the system dwells in each state m. Let, A_m be those quantities in a trajectory of length T and

$$A_m(T) = \sum_{i_0, i_1, \cdots, i_T} p_{i_0 i_1 \cdots i_T} \sum_{k=0}^{T} \delta_{i_k, m}, \tag{3}$$

where $\delta_{i,j} = 1$ when $i = j$ and 0 otherwise. Thus, in Eq. (3), $\sum_{k=0}^{T} \delta_{i_k, m}$ counts the number of times state m is visited. By averaging this quantity over all microscopic trajectories, we find the mean number of times this state is visited from time 0 to T. We, therefore, have that $\sum_m A_m = T + 1$. When data are given in terms of quantities like $A_m(T)$, we refer to these as "singlet statistics."

Now, we maximize the path entropy in Eq. (1) subject to constraints on our singlet statistic, imposed using Lagrange multipliers. That is, we maximize the Caliber, $S(T) - \sum_m \lambda_m A_m(T)$ with respect to each $p_{i_0 i_1 \cdots i_T}$ where λ_m is the Lagrange multiplier for state m. We therefore solve

$$\frac{\partial [S(T) - \sum_m \lambda_m A_m(T)]}{\partial p_{i_0 i_1 \cdots i_T}}$$

$$= 1 + \log p_{i_0 i_1 \cdots i_T} - \sum_m \lambda_m \sum_{k=0}^{T} \delta_{i_k, m} = 0,$$

which yields

$$p_{i_0 i_1 \cdots i_T} \propto e^{-\sum_m \lambda_m \sum_{k=0}^{T} \delta_{i_k, m}} = \prod_{k=0}^{T} e^{-\lambda_{i_k}}. \tag{4}$$

Note that these λ's are all implicitly dependent on the length of the microscopic trajectory in time, T.

Once normalized, the probability becomes

$$p_{i_0 i_1 \cdots i_T} = \frac{\prod_{k=0}^{T} e^{-\lambda_{i_k}}}{\sum_{i_0 i_1 \cdots i_T} \prod_{k=0}^{T} e^{-\lambda_{i_k}}}$$

$$= \frac{\prod_{k=0}^{T} e^{-\lambda_{i_k}}}{(\sum_{i=1}^{N} e^{-\lambda_i})^{T+1}} = \prod_{k=0}^{T} p_{i_k}(T), \tag{5}$$

where $p_{i_k}(T) = e^{-\lambda_{i_k}} / \sum_{i=1}^{N} e^{-\lambda_i}$ is the probability of being in state i_k at time k for a trajectory of length T. The trajec-

tory $p_{i_0 i_1 \cdots i_T}$ thus breaks up into the product of independent probabilities for each time interval.

Therefore, we have

$$A_m = \sum_{i_0, i_1, \cdots, i_T} p_{i_0 i_1 \cdots i_T} \sum_{k=0}^{T} \delta_{i_k, m} = (T + 1) p_m(T), \tag{6}$$

where $p_m(T) = \sum_i p_i \delta_{i, m}$ is interpreted as the probability of being found in state m over the course of time T.

As a final note on this example, we consider the result of imposing the consistency condition, Eq. (2), to constrain the distribution obtained above. From this we obtain for the time interval $[0, T+1]$,

$$\sum_{i_{T+1}} p_{i_0 \cdots i_T i_{T+1}} = \prod_{k=0}^{T} p_{i_k}(T+1) \left[\sum_{i_{T+1}} p_{i_{T+1}}(T+1) \right]$$

$$= \prod_{k=0}^{T} p_{i_k}(T+1). \tag{7}$$

Since $\sum_{i_{T+1}} p_{i_0 \cdots i_T i_{T+1}} = p_{i_0 i_1 \cdots i_T} = \prod_{k=0}^{T} p_{i_k}(T)$ due to the results from Max Cal for time interval $[0, T]$, it then follows that

$$\prod_{k=0}^{T} p_{i_k}(T+1) = \prod_{k=0}^{T} p_{i_k}(T), \tag{8}$$

which gives

$$p_{i_k}(T+1) = p_{i_k}(T) \tag{9}$$

for any i_k.

The proof from Eqs. (8) to (9) follows from summing both sides of Eq. (8) over all indices except i_k. Furthermore, since the state i_k is arbitrary, we recover Eq. (9).

So, under singlet constraints, maximizing the caliber leads to a model which is an identical, independent distributed process. Equation (5) is the statement of independence and Eq. (9) is the statement of the identical distributions.

B. Constraint on pairwise statistics

Now we consider instead the situation in which the constraint is on the pairwise statistics for each step $m \to n$ over the time period $[0, T]$, i.e.,

$$A_{mn} = \sum_{i_0, \cdots, i_T} p_{i_0 \cdots i_T} \sum_{k=0}^{T-1} \delta_{i_k, m} \delta_{i_{k+1}, n}. \tag{10}$$

Here, $\sum_{k=0}^{T-1} \delta_{i_k, m} \delta_{i_{k+1}, n}$ is just the number of occurrence of the transition $m \to n$ and $\sum_{m, n} A_{mn} = T$.

Now we maximize the path entropy in Eq. (1) subject to constraints on the pairwise statistics given by Eq. (10). We use the λ_{mn} quantities as Lagrange multipliers to constrain A_{mn}. Following a derivation similar to that of Eq. (5), we now get

$$p_{i_0 \cdots i_T} = \prod_{k=0}^{T-1} p_{i_k i_{k+1}} \propto e^{-\sum_{m, n} \lambda_{mn} \sum_{k=0}^{T-1} \delta_{i_k, m} \delta_{i_{k+1}, n}}, \tag{11}$$

where $p_{i_k i_{k+1}} \propto e^{-\lambda_{i_k i_{k+1}}}$. Following similar reasoning to Eqs. (7) and (8), we find

$$p_{i_k i_{k+1}}(T+1) = p_{i_k i_{k+1}}(T).\qquad(12)$$

Equations (11) and (12) are the main results of this subsection. In particular, Eq. (11) shows that when transitions are used as constraints the procedure of Max Cal yields a Markov process as the model where the probability of a microtrajectory is given by the product of the independent transition probabilities. Furthermore, Eq. (12) says that these transitions are independent of the total trajectory length.

C. What if the initial distribution is not uniformly distributed?

In Secs. I–II B, we did not use any information about the initial condition of the states; hence, we implicitly assumed uniform initial conditions. In many cases, however, the initial conditions are known and not uniform. How can we include this information within Max Cal? As an example, if a system is subjected to a temperature spike at time $t = 0$, prior to relaxation, its initial conditions will not be uniform. By this we mean that proteins are initially sampled from a particular (presumably unequilibrated) conformational state. T-jump experiments are useful, for instance, in studying the folding kinetics of proteins from a denatured state.[21]

Max Cal still applies in cases where the initial condition is not uniform. However, we must now condition the path entropy on the known initial state, $p_{i_0}(0) \equiv p_{i_0}(t = 0)$. This is an additional piece of information that can be regarded as a constraint on path probability. In general, we will denote the probability of occupying state i_n at time n as $p_{i_n}(n)$.

In other words, the following expression:

$$p_{i_0 i_1 \cdots i_T} = p_{i_0}(0) p_{i_1 \cdots i_T | i_0}\qquad(13)$$

must be substituted into the definition of the path entropy, Eq. (1), that is then to be maximized with respect to a further constraint available on $p_{i_0}(0)$. In the above $p_{i_1 \cdots i_T | i_0}$ denotes the probability of the path from time 1 to T conditioned on being in state i_0 at time 0.

Having pre-specified the initial conditions and now constraining pairwise statistics, we recover

$$p_{i_0 \cdots i_T} = p_{i_0}(0) \prod_{k=0}^{T-1} p_{i_k i_{k+1}}.\qquad(14)$$

In this case, the state occupancies, the p_i's, are not necessarily independent of time. A particular state occupancy p_{i_k} can depend on how much time has elapsed since $t = 0$. To see this explicitly, we note that $\sum_{i_0} p_{i_0}(0) p_{i_0 i_1}$ is now equal to $p_{i_1}(1)$ which might be different from the steady-state value (which is denoted as π_{i_1}). The steady-state value would instead be obtained from $\sum_{i_0} \pi_{i_0} p_{i_0 i_1}$. Upon specifying initial conditions, we have in general

$$\sum_{i_n} p_{i_n}(n) p_{i_n i_{n+1}} = p_{i_n}(n+1).\qquad(15)$$

When initial conditions are not specified, i.e., no constraints are imposed on p_{i_0}, then plugging Eq. (14) into the expression for the path entropy, Eq. (1), we have

$$S(T) = -\sum_{i_0,i_1,\cdots,i_T} p_{i_0}(0) \prod_{k=0}^{T-1} p_{i_k i_{k+1}} \log\left(p_{i_0}(0) \prod_{k=0}^{T-1} p_{i_k i_{k+1}}\right).$$

The $p_{i_0}(0)$ which maximizes this expression (i.e., taking the derivative of the above with respect to p_{i_0} and setting that derivative equal to zero) is the initial uniform distribution. By uniform we mean that the initial distribution is one with equal *a priori* weight assigned to each state. The same result is also obtained in statistical mechanics by maximizing the entropy over states without constraints. The initial distribution is, therefore, not the stationary distribution.

In theory, when knowledge of a stationary distribution is known *a priori*, i.e., when π_{i_0} is known, this value of p_{i_0} can be used as a prior. That is,

$$S(T) = -\sum_{i_0,i_1,\cdots,i_T} p_{i_0}(0) \prod_{k=0}^{T-1} p_{i_k i_{k+1}}$$
$$\times \log\left(p_{i_0}(0) \prod_{k=0}^{T-1} p_{i_k i_{k+1}}/\pi_{i_0}\right).\qquad(16)$$

In this case, value of $p_{i_0}(0)$ which maximizes the path entropy is found to be π_{i_0} upon normalization.

In practical terms, for a long trajectory, where each state is visited multiple times, the steady state is reliably obtained as follows. We define the *dynamical partition function*

$$Q_d(T) = \sum_{i_0,\cdots,i_T} e^{-\sum_{m,n} \lambda_{mn} \sum_{k=0}^{T-1} \delta_{i_k,m} \delta_{i_{k+1},n}}.\qquad(17)$$

Taking the derivative of this dynamical partition function gives the average number of time intervals of dwell, $A_{mn} = -\partial \log Q_d(T)/\partial \log \lambda_{mn}$ (resembling the way that taking derivatives of equilibrium partition functions yield equilibrium averages and higher cumulants). When T is large enough, we have $\sum_m A_{mn}(T)/T \to \pi_n$, which is the stationary distribution of state n.

Thus, even when the initial conditions are assumed to be uniform, rather than to select initial conditions from the stationary distribution, the system still tends towards the correct stationary state.

A note on data analysis is warranted. We have just seen that constraints determine the functional form of the probability for observing a particular trajectory in maximum caliber – in other words they justify the Markov chain. Max Cal thus yields the mathematical form for the likelihood function used to describe the probability of a particular trajectory. Parameterizing a trajectory – i.e., finding numerical values for all transition probabilities – is equivalent to maximizing the caliber.

III. MODEL-BUILDING IN MAX CAL IS DRIVEN BY THE DATA

Max Cal gives a systematic procedure for building kinetic models. To build a kinetic model, pair statistics are first used as constraints. But how do we know this is the correct model? If measured correlation functions show this constraint

to be insufficient, triplet statistics are otherwise used as constraints and so forth. Below we show a simple problem where pairwise statistics gives the correct distribution, but in another slightly complex example it makes the wrong prediction. The example further shows how we can improve the prediction by considering higher order statistics.

Suppose the underlying physics is just two-state, $A \underset{k_B}{\overset{k_A}{\rightleftharpoons}} B$ and the profiles of A and B are measured over time intervals of length δt. Hence, one could apply Max Cal by first considering a microtrajectory starting in state A, i.e., using $p_A(0) = 1$ and $p_B(0) = 0$. For example, suppose we measure $p_{AA} = 1 - k_A \delta t$ where p_{AA} is the probability that the system is still in state A at time δt, given that it was in state A at time $t = 0$. Since $p_{AB} + p_{AA} = 1$ then $p_{AB} = k_A \delta t$. The probability of staying some total time $t \equiv N \delta t$ in state A is $p_A(t) \equiv p_{AA}^N = (1 - k_A \delta t)^N = \exp(-k_A t)$ for long enough t. This would be your expectation based only on pairwise constraints. The same situation would occur for p_{BB}, p_{BA}, and $p_B(t)$. We know that for such a system the steady-state distribution is $p_A^{ss} = k_B/(k_A + k_B)$ and $p_B^{ss} = k_A/(k_A + k_B)$. These are obtained from $p_A^{ss} = p_{BA}/(p_{AB} + p_{BA})$ and $p_B^{ss} = p_{AB}/(p_{AB} + p_{BA})$. To summarize, Max Cal infers the correct steady-state distribution after repeated measurements, i.e., accurate measurements of p_{AB} and p_{BA}, though, in the absence of initial conditions, it infers uniform initial conditions. However, if the trajectory observed is sufficiently long, Max Cal does predict the correct steady-state occupation of each state provided all states are clearly resolved no matter how many states we have.

Now we consider the case where not all states are resolved. Consider the following reaction scheme as an example: $A \overset{k_A}{\rightarrow} B \overset{k_C}{\rightarrow} C$. Suppose that species A and B cannot be resolved experimentally. When not all states are resolved, transitions from the aggregated state are not Markov (technically not first-order Markov). The aggregate of A and B is labeled $\bar{A}$. In this case, the probability of dwelling in state $\bar{A}$ as a function of t, $p_{\bar{A}}(t)$, is

$$p_{\bar{A}}(t) = p_A(t) + p_B(t)$$
$$= (k_C - k_A)^{-1}(k_C e^{-k_A t} - k_A e^{-k_C t}). \quad (18)$$

Theoretically, there are two ways of backing out a model. One is to fit or to infer the number of exponential contributions to the decay of Eq. (18) and thereby infer the intermediate state B.[22] The other is to build the Markovian model from Max Cal through entropy maximization with finite-step constraints. In practice Max Cal is more general (it does not rely on exponentials) and is easier to apply.

Pairwise constraints, such as above, would predict exponential distribution for $p_{\bar{A}}(t)$ instead of Eq. (18). Clearly, this would be inconsistent with a measurement consistent with Eq. (18) instead. From pairwise constraints, we could predict quantities like the probability of observing $p_{i_0 i_1 i_2} = p_{\bar{A}\bar{A}\bar{A}}$ within δt. However, our prediction would be incorrect.

We, therefore, turn to data beyond two point constraints to see if we can improve our prediction. We use the set of three point constraints such as $p_{\bar{A}\bar{A}\bar{A}}$ based on the underlying model

$A \overset{k_A}{\rightarrow} B \overset{k_C}{\rightarrow} C$. The constraint $p_{\bar{A}\bar{A}\bar{A}}$ is computed as follows:

$$p_{\bar{A}\bar{A}\bar{A}} = p_{AAA} + p_{AAB} + p_{ABB}, \quad (19)$$

where $p_{AAA} = p_{AA}p_{AA} = (1 - k_A(\delta t/2))^2$, $p_{AAB} = p_{AA}p_{AB} = (1 - k_A(\delta t/2))k_A(\delta t/2)$, and $p_{ABB} = p_{AB}p_{BB} = k_A(\delta t/2)(1 - k_C(\delta t/2))$. Therefore,

$$p_{\bar{A}\bar{A}\bar{A}} = 1 - k_A k_C \left(\frac{\delta t}{2}\right)^2. \quad (20)$$

In a similar fashion, we find $p_{\bar{A}\bar{A}C} = k_A k_C (\delta t/2)^2$ and $p_{\bar{A}CC} = 0$ where $p_{\bar{A}\bar{A}C} = p_{ABC} = k_A k_C ((\delta t/2))^2$.

Now, expanding $p_{\bar{A}}(t)$, Eq. (18), to second order in δt, we have $p_{\bar{A}}(t) \sim 1 - k_A k_C \delta t^2/2$. Comparing this with $p_{\bar{A}\bar{A}\bar{A}}$, it is clear they both capture the correct k_A and k_C dependence to second order in δt though the pre-factors do not agree. Predictions of other higher order trajectories would also be in better agreement by having third-order constraints.

Just as we had before, this exercise can be carried out to even higher order and the difference in the pre-factor would be eliminated when we use more and more steps as constraints. For instance, using fourth-order time constraints we find

$$p_{\bar{A}\bar{A}\bar{A}\bar{A}} = 1 - 3k_A k_C \left[\left(\frac{\delta t}{3}\right)^2 - (k_A + k_C)\left(\frac{\delta t}{3}\right)^3 /3 \right]. \quad (21)$$

This should be compared with the expansion of Eq. (18) to the third order in δt that yields

$$p_{\bar{A}}(t) = 1 - \frac{k_A k_C}{2}[(\delta t)^2 - (k_A + k_C)(\delta t)^3/3]. \quad (22)$$

This time the prefactor of the second-order term in Eq. (21) changes to $1/3$ from $1/4$ in Eq. (20), and in the limit of large number of time steps, it would finally become the correct $1/2$. Thus, we find that with higher order constraints we capture more of the behavior of the full profile $p_{\bar{A}}(t)$ when the underlying process cannot be accurately determined by a two-point constraint. The example above outlines how to improve our choice of constraints depending on the data.

IV. CONCLUSIONS AND DISCUSSION

The dynamics of bulk chemical reaction kinetics are governed by the law of mass action. For chemical reactions in small systems like in cellular biochemical environments, the dynamics are stochastic, and many models assume Markovian dynamics. On the other hand, with growing microscopic biochemical data on dynamical systems, a statistical approach to chemical kinetics which is founded on the inferential principles of maximum entropy is desirable. For example, it is well known that exponential distribution for a positive random variable, and Gaussian distribution, both essential to chemical kinetics, can be justified using Jaynes "principle of maximum entropy."[7,8] It is in this spirit that the approach based on Max Cal for statistical dynamics was proposed.[3,9–12] Maximum caliber can be used both as a first principle from which to derive stochastic dynamical models and as a method of data analysis. We showed here that

Markovian dynamics can be justified from Max Cal; the Markov property is a natural consequence of maximizing a dynamical entropy over trajectories for time-independent processes. Our treatment can be generalized to handle dynamical systems having finite memory (time delays), corresponding to cases of inference used in time-series analysis.[23]

ACKNOWLEDGMENTS

[1] N. G. van Kampen, *Stochastic Processes in Chemistry and Physics* (North-Holland, Amsterdam, 1981).
[2] K. L. Chung, *Lectures from Markov Processes to Brownian Motion* (Springer-Verlag, New York, 1982).
[3] E. T. Jaynes, "Macroscopic prediction," in *Complex Systems – Operational Approaches in Neurobiology, Physics, and Computers*, edited by H. Haken (Springer-Verlag, Berlin, 1985).
[4] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, London, 2003).
[5] P. J. Steinbach, K. Chu, H. Frauenfelder, J. B. Johnson, D. C. Lamb, G. U. Nienhaus, T. B. Sauke, and R. D. Young, Biophys. J. **61**, 235 (1992).
[6] S. F. Gull and G. J. Daniell, Nature (London) **272**, 686 (1978).
[7] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[8] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[9] E. Seitaridou, M. M. Inamdar, R. Phillips, K. Ghosh, and K. A. Dill, J. Phys. Chem. B **111**, 2288 (2007).
[10] S. Pressé, K. Ghosh, R. Phillips, and K. A. Dill, Phys. Rev. E **82**, 031905 (2010).
[11] S. Pressé, K. Ghosh, and K. A. Dill, J. Phys. Chem. B **115**, 6202 (2011).
[12] M. Otten and G. Stock, J. Chem. Phys. **133**, 034119 (2010).
[13] G. Stock, K. Ghosh, and K. A. Dill, J. Chem. Phys. **128**, 194102 (2008).
[14] E. Smith, Rep. Prog. Phys. **74**, 046601 (2011).
[15] C. Monthus, J. Stat. Mech. **2011**, P03008.
[16] K. Ghosh, J. Chem. Phys. **134**, 195101 (2011).
[17] A. A. Filyukov and V. Ya. Karpov, Inzh.-Fiz. Zh. **13**, 798 (1967).
[18] A. K. Livesey and J. Skilling, Acta Cryst. **A41**, 113 (1985).
[19] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **IT-26**, 26 (1980).
[20] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin, 1933) [*Foundations of the Theory of Probability* (Chelsea, New York, 1950) (in English)].
[21] M. Gruebele, Annu. Rev. Phys. Chem. **50**, 485 (1999).
[22] H. P. Lu, L. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[23] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. (Springer-Verlag, New York, 2009).