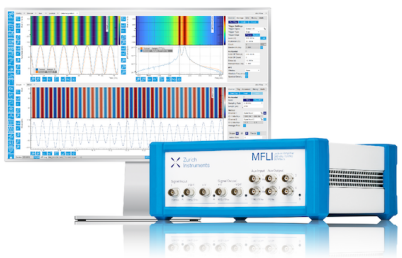
RESEARCH ARTICLE | JULY 21 2010

Maximum caliber inference of nonequilibrium processes

Moritz Otten; Gerhard Stock

J. Chem. Phys. 133, 034119 (2010)
https://doi.org/10.1063/1.3455333



Maximum caliber inference of nonequilibrium processes

Moritz Otten and Gerhard Stock[a)]
Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, 79104 Freiburg, Germany

(Received 26 March 2010; accepted 28 May 2010; published online 21 July 2010)

Thirty years ago, Jaynes suggested a general theoretical approach to nonequilibrium statistical mechanics, called maximum caliber (MaxCal) [Annu. Rev. Phys. Chem. **31**, 579 (1980)]. MaxCal is a variational principle for dynamics in the same spirit that maximum entropy is a variational principle for equilibrium statistical mechanics. Motivated by the success of maximum entropy inference methods for equilibrium problems, in this work the MaxCal formulation is applied to the inference of nonequilibrium processes. That is, given some time-dependent observables of a dynamical process, one constructs a model that reproduces these input data and moreover, predicts the underlying dynamics of the system. For example, the observables could be some time-resolved measurements of the folding of a protein, which are described by a few-state model of the free energy landscape of the system. MaxCal then calculates the probabilities of an ensemble of trajectories such that on average the data are reproduced. From this probability distribution, any dynamical quantity of the system can be calculated, including population probabilities, fluxes, or waiting time distributions. After briefly reviewing the formalism, the practical numerical implementation of MaxCal in the case of an inference problem is discussed. Adopting various few-state models of increasing complexity, it is demonstrated that the MaxCal principle indeed works as a practical method of inference: The scheme is fairly robust and yields correct results as long as the input data are sufficient. As the method is unbiased and general, it can deal with any kind of time dependency such as oscillatory transients and multitime decays. © *2010 American Institute of Physics*. [doi:10.1063/1.3455333]

I. INTRODUCTION

The dynamics and function of complex molecular systems can be studied in great detail via time-resolved multidimensional spectroscopy.[1] As an illustrative example, let us consider a (fictive) laser experiment on the photoinduced folding reaction of a protein. We assume that the experiment provides two observables R_α and R_β shown in Fig. 1(A): R_α reports on the rise of a quantity (e.g., the amount of helicity in the protein structure) that is obviously not present in the initial state but characterizes the final state. R_β monitors the decay of a quantity (e.g., the radius of gyration) that is initially at maximum and decreases as the system relaxes to equilibrium. To interpret these findings, one needs to identify the underlying dynamical process that gives rise to these results. In particular, one would like to construct the multidimensional free energy landscape of the system, which exhibits the correct number, energy, and location of the system's metastable states and barriers.[2–6]

As a simple model that generates the time evolution of the observables R_α and R_β, Fig. 1(B) displays a system with four metastable conformational states. It contains the initially prepared unfolded state **I**, the final folded state **F**, and two intermediate states, a more populated state **1** and a less populated state **2**, which act as an alternative pathway. The states are defined in the projection plane of the two observable quantities, referred to as the (R_α,R_β)-map. For example, $(R_\alpha,R_\beta)=(0,1)$ for initial state **I**. Although the model is correct (in the sense that it reproduces the available information R_α and R_β) and maybe plausible (in the sense that, e.g., the chosen number of states is anticipated from experiment), it is usually not unique. This is because the available information is not sufficient to identify a unique process as the source of the observable data. While numerous approaches for so-called inverse problems have been developed,[7] most of these inference methods deal with time-independent problems. The inference of time-dependent nonequilibrium processes, on the other hand, is usually based on a memoryless jump process as generated by a discrete Markov state model.[8–12] In many molecular systems, however, the underlying assumption of a time scale separation between fast and slow interstate transitions may break down.[13]

In this work, we wish to develop a general inference method for nonequilibrium processes. The approach is based on a time-dependent extension of the maximum entropy formulation introduced by Jaynes.[14–16] Before discussing the time-dependent case, it is instructive to first sketch the basic idea of the well-established maximum entropy inference of an equilibrium problem. To this end, we consider a system with a given set of microstates $j=1,2,3,\ldots$. While the equilibrium probabilities p_j of those microstates are unknown, we know one or several macroscopically observable quantities $\langle A_m\rangle$, which are given by an average over all microstates,

$$\langle A_m\rangle = \sum_j p_j A_{mj}, \tag{1}$$

where A_{mj} denotes the value of variables A_m in state j, thus characterizing this state. By defining the entropy of the sys-

a)Electronic mail: stock@physik.uni-freiburg.de.

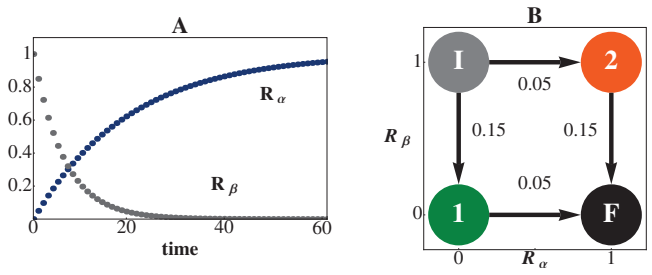


FIG. 1. (a) Illustration of a photoinduced protein folding reaction as monitored by the time-dependent observables R_α and R_β, which may reflect, e.g., the helicity of the system and its radius of gyration, respectively. (b) This time evolution can be generated by a four-state system, where the states are clearly defined by the two observables R_α and R_β. All reaction rates are given in number of transitions per time unit.

tem as $S=-\Sigma_j p_j \ln p_j$, the equilibrium probabilities p_j can be determined by maximizing this entropy function under the constraints given by Eq. (1). Considering a thermodynamic system with total energy $\langle E\rangle$, for example, this procedure directly leads to Boltzmann's distribution law, $p_j = e^{-\beta E_j}/\Sigma_i e^{-\beta E_i}$, with β being the inverse temperature. This simple example shows that statistical mechanics can be posed as an inference problem: Given the macroscopic measurements $\langle A_m\rangle$, what is the underlying distribution of probabilities p_j? By construction, the maximum entropy method gives the distribution with the highest multiplicity that satisfies the given constraints: an equal distribution if no information is available, an exponential distribution if only the mean of a quantity is given (as in Boltzmann's distribution law), a Gaussian distribution if the mean and the variance of a quantity are given, and so on. The maximum entropy inference method has been applied in numerous fields.[17] A recent application in the field of chemical physics, for example, was to construct the conformational distribution of peptides from nuclear magnetic resonance data.[18]

To account for nonequilibrium processes, one needs to generalize the maximum entropy formulation to time-dependent problems.[19–21] In this case we have some time-dependent quantities A_m with averages

$$\langle A_m(t)\rangle = \sum_j p_j A_{mj}(t). \tag{2}$$

Instead of the equilibrium probability of a microstate in Eq. (1), the p_j now denotes the probability of a *microtrajectory*, e.g., the path of a specific single-particle. These trajectories represent individual realizations of the considered dynamical process and $A_{mj}(t)$ describes the time evolution of observable A_m along trajectory j. As a consequence, the resulting entropy $\mathcal{C}=-\Sigma_j p_j \ln p_j$ will be a *functional* or path integral[22] of the $\{p_j\}$. In direct analogy to the equilibrium case, we maximize this functional under constraint (2) to find the weights of the individual dynamical paths, p_j. Jaynes called this functional "Caliber" since it refers to the cross sectional area of a tube, which partly determines the flow in a dynamic process.[20]

The maximum caliber (MaxCal) formulation was originally suggested by Jaynes as a variational principle to predict dynamical properties of systems.[20] The approach has been subject to some formal study,[23,24] and was recently used to derive transport laws such as Fick's law of particle transport[25,26] as well as various nonequilibrium statistical formulations of two-state dynamics.[27] Furthermore, the MaxCal principle has been recently applied to describe single-particle experiments.[28] In general, MaxCal should be useful for systems, such as in biology and nanotechnology, where the numbers of particles are small and where there is some interest in knowing the distribution of trajectories.[29,30]

The aim of this work is to apply MaxCal to the inference of nonequilibrium processes such as presented in Fig. 1. With this end in mind, we first briefly review the MaxCal formalism and describe its practical numerical implementation in the case of an inference problem. The performance of the approach is demonstrated by adopting various few-state models of increasing complexity: several Markov state models, a hidden-state model that gives rise to biexponential kinetics, and a model that exhibits time-dependent transition rates. The potential of the method as well as its virtues and shortcomings are discussed in some detail.

II. THEORY AND IMPLEMENTATION

The MaxCal formulation can be summarized as follows (see Ref. 27 for a more detailed description). Let us assume that we know the observables $\langle A_m\rangle$ $(m=1,\ldots,M)$, given by Eq. (2) as a function of the path weights p_j. The "caliber" of the underlying process can then be written as

$$\mathcal{C}(\{p_j\}) = -\sum_j p_j \ln p_j + \mu \sum_j p_j + \sum_m \lambda_m \sum_j p_j A_{mj}. \tag{3}$$

The first term represents the Shannon entropy of the process, which becomes maximal for the path distribution $\{p_j\}$ with the highest multiplicity. The Lagrange multipliers μ and λ_n enforce that the distribution is normalized (i.e., that $\Sigma_j p_j=1$) and that average (2) is satisfied. The sum goes over all possible microtrajectories j, whose total number, in general, grows exponentially with the duration of the process. To find the weights of these individual dynamical paths, p_j, we maximize caliber (3) by setting $\delta\mathcal{C}/\delta p_j=0$. This gives for the path weights

$$p_j = Q_{\mathrm{d}}^{-1} \exp\left\{-\sum_m \lambda_m A_{mj}\right\}, \tag{4}$$

where $Q_{\mathrm{d}}=\Sigma_j \exp\{-\Sigma_m \lambda_m A_{mj}\}$ denotes the dynamical partition function, which can be shown to contain all information on the dynamical process.[27] Hence we need to solve the set of M nonlinear equations

$$\langle A_m\rangle = \sum_j p_j(\{\lambda_m\}) A_{mj} \tag{5}$$

for the Lagrange multipliers $\{\lambda_m\}$, which can be accomplished by using standard procedures. So far, however, the quantities λ_m and A_{mj} are both constant, i.e., the path weights p_j as well as the observables $\langle A_m\rangle$ actually do not depend on time. Hence caliber function (3) is only suitable for time-independent observables of the dynamical process, such as the average number of transitions.[27]

To allow for explicitly time-dependent observables, we assume that the observables are given as time series

$\{\langle A_m(0)\rangle, \langle A_m(1)\rangle, \ldots, \langle A_m(T-1)\rangle\}$ with T points. As a consequence, Eq. (2) yields $M \times T$ dynamical constraints. The caliber of the underlying process can then be written as[27]

$$\mathcal{C}(\{p_j\}) = -\sum_j p_j \ln p_j + \mu \sum_j p_j + \sum_j p_j \sum_{m,t} \lambda_{mt} A_{mj}(t), \quad (6)$$

where $t \leq T-1$. Maximizing this function, we obtain the path weights

$$p_j = Q_d^{-1} \exp\left\{-\sum_{m,t} \lambda_{mt} A_{mj}(t)\right\}. \quad (7)$$

Finally, the Lagrange multipliers $\{\lambda_{mt}\}$ are obtained by solving the $M \times T$ equations

$$\langle A_m(t)\rangle = \sum_j p_j(\lambda_{mt}) A_{mj}(t). \quad (8)$$

To evaluate Eq. (8) in practice, we need to define the underlying microtrajectories $\{j\}$, that is, we have to introduce a state or configuration space, in which the trajectories live. Given M as continuous observables $R_m(t) \equiv \langle A_m(t)\rangle$, an obvious choice is the M-dimensional configuration space spanned by these coordinates. In practice, however, one often finds the situation that the system does not adopt all possible coordinate values, but rather can be described in terms of a few discrete states

$$k = 1, \ldots, N, \quad (9)$$

with well-defined values of the coordinates R_m [see, e.g., Fig. 1(B)]. The suitable number of states N is *a priori* not known. In practice, it may be anticipated from experiment or be determined by an iterative computational approach. As a further simplification, one may restrict the possible transitions from and to each state, e.g., by allowing only transitions to neighboring states. Given these definitions, a microtrajectory j can be defined as the discrete-valued time series $k_j(t)$. Denoting the time-dependent population probability of the kth state by $P_k(t)$, the observables can be calculated by

$$R_m(t) = \sum_{k=1}^{N} P_k(t) R_{mk}, \quad (10)$$

where R_{mk} represents the value of observable R_m in state k.

As an example, let us apply the MaxCal formulation to the inference of the folding problem introduced in Fig. 1. In this case, we have $M=2$ observables, $R_\alpha(t)$ and $R_\beta(t)$, which are given as time series $\{R_m(0), \ldots, R_m(T-1)\}$ with $T=61$ points (as at $t=60$ the system has largely relaxed to equilibrium). As explained above, we choose $N=4$ conformational states which are connected as indicated in Fig. 1(B). Assuming that the system is initially in state **I**, we can calculate all possible trajectories $k_j(t)$ for $0 \leq t \leq 60$. Moreover, we calculate the values of the observables $R_{mk_j(t)}$ along these trajectories from the (R_α, R_β)-map in Fig. 1(B). Inserting these values, we solve the $M \times T = 2 \times 61$ nonlinear equation (8) for the 2×61 Lagrange multipliers $\{\lambda_{mt}\}$ using standard methods. The solution of these equation is unique since the caliber $\mathcal{C}$ represents a convex function of the Lagrange multipliers.[31] Hence it is evident if the algorithm cannot find a set of Lagrange multipliers that satisfy constraint (8), for example, if the chosen model is not suited or general enough. Contrariwise, satisfying all constraints means that the chosen model is complex enough to describe the dynamics. This might be the basis for a technique that iteratively raises the complexity of the system in order to find the simplest model that works to produce the data. As shown below, the scheme is fairly robust and yields correct results as long as the input data are sufficient.

III. COMPUTATIONAL RESULTS

To demonstrate the performance and the potential of the MaxCal inference method, we adopt various models of increasing complexity: various Markov models as the one shown in Fig. 1, a hidden-state model that gives rise to biexponential kinetics (Fig. 6), and a model that exhibits time-dependent transition rates (Fig. 7). As explained below, all these processes can be realized in terms of simple rate equations of some state population probabilities $P_k(t)$. As a reference, we therefore first solve the rate equation $\dot{\mathbf{P}} = -\mathcal{T}\mathbf{P}$ in order to obtain the exact state populations $\mathbf{P}(t) = (P_1(t), \ldots, P_N(t))^\dagger$ via

$$\mathbf{P}(t) = \prod_{\tau=1}^{t} \mathcal{T}(\tau)\mathbf{P}(0), \quad (11)$$

where, in general, the transition matrix $\mathcal{T}$ may depend on time. In all cases, we assume that we start in state **I**, i.e., $\mathbf{P}(0) = (1,0,0\ldots,0)^\dagger$. From these populations, the time-dependent observables $R_m(t)$ under consideration are readily calculated via Eq. (10). Usually, we will consider two observables, $R_\alpha(t)$ and $R_\beta(t)$, which serve as examples of typical macroscopic averages given from an experiment. These observables then serve as input data for the MaxCal inference procedure described above. By calculating all possible trajectories $k_j(t)$ and their path weights p_j, the goal is to reproduce the state populations $P_k(t)$ through

$$P_k(t) = \sum_j p_j \delta_{k,k_j(t)}. \quad (12)$$

In the considered case of four states and two constraints the problem is underdetermined. Using three constraints plus the normalization condition $\Sigma_k P_k(t) = 1$, the populations $P_k(t)$ can be simply recovered by directly solving Eq. (10) for all t, i.e., the problem is determined.

A. Discrete Markov process

As a proof-of-principle example, we consider the four-state Markov model shown in Fig. 1(B). The transition rates given in the figure results in the (time-independent) transition matrix

$$\mathcal{T} = \begin{pmatrix} 0.8 & 0 & 0 & 0 \\ 0.15 & 0.95 & 0 & 0 \\ 0.05 & 0 & 0.85 & 0 \\ 0 & 0.05 & 0.15 & 1 \end{pmatrix}. \quad (13)$$

Evaluation of Eq. (11) gives the time evolution of the state populations $P_k(t)$ $(k=1,\ldots,4)$ shown in Fig. 2(A). As is ex-

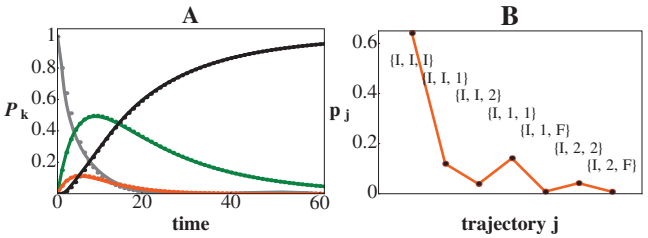


FIG. 2. (a) Time evolution of the population probabilities $P_k(t)$ of states **I** (gray), **1** (green), **2** (red), and **F** (black), as obtained for the four-state Markov system shown in Fig. 1. (b) Probability distribution of all possible trajectories after the first two time steps (in arbitrary order). In both cases, the reference calculations obtained from Eq. (11) (dots) are compared to the MaxCal inference results (full lines).

pected from the simple structure of the transition matrix, we find that the initially prepared state **I** decays exponentially with the overall rate 0.2 (transition rates are given in number of transitions per time unit). The population flows into state **1** with the rate 0.15 and into state **2** with the rate 0.05, which subsequently decay into the final state **F** with rates 0.05 and 0.15, respectively. Hence, state **1** is more metastable than state **2** and exhibits higher population at intermediate times. After $T=60$ time steps, the process has approximately reached its equilibrium state $\mathbf{P}(\infty)=(0,0,0,1)^{\dagger}$. By using Eq. (10), we obtain the time-dependent observables $R_\alpha(t)$ and $R_\beta(t)$.

We now view the example as an inference problem and suppose that we only know the two time-dependent observables $R_\alpha(t)$ and $R_\beta(t)$ as well as the location of the states with respect to R_α and R_β as shown in Fig. 1(B). Using these observables, the MaxCal inference procedure allows us to compute all possible trajectories $k_j(t)$ and their path weights p_j. Subsequently, the trajectory distribution $\{p_j\}$ can be used to calculate any possible observable of the system. As an illustration, Fig. 2(B) shows the path weights after the first two time steps. Assuming that at time $t=0$ all trajectories start in state **I**, at $t=1$ we have one trajectory still sitting in state **1**, one that moved to state **2**, and one that moved to state **3**. In the next and all succeeding steps, each trajectory spawns (at most) three new trajectories, which either stay in its state or change to the two connected states. Proceeding this way, we find at time $T=2$ the trajectory distribution $\{p_j\}$ shown in Fig. 2(B). Due to the relatively high metastability (0.8) of the initial state **I**, the trajectory j with highest probability p_j is the one that stays in state **I**. Denoting this trajectory by $(\mathbf{I},\mathbf{I},\mathbf{I})$, the next most likely case is trajectory $(\mathbf{I},\mathbf{1},\mathbf{1})$, which changes to state 1 in the first step and stays there. Also at longer times, trajectories of the kind $(\mathbf{I},\mathbf{1},\mathbf{1},\ldots\mathbf{1})$ occur with high probability because of the high metastability of state **I**.

Also shown in Fig. 2(B) are the exact path weights which were obtained from the transition matrix. We see that for this simple first example the MaxCal inference method is in perfect agreement with the reference data. (The accuracy is similar for longer times; $T=2$ was only chosen to keep the figure simple.) The agreement of the trajectory distribution $\{p_j\}$ means that the dynamics is fully reproduced in all observable quantities since the resulting *dynamical partition function* $Q_{\mathrm{d}}=\Sigma_j p_j$ has the same meaning as the common partition function in equilibrium statistical mechanics.[27] In Fig. 2 we also compare the inferred time evolution of the state populations $P_k(t)$, which indeed matches the reference results.

B. Generalizations

To explain the main idea of the MaxCal inference method, we have adopted a quite simple example which assumed (i) that the number of states of the system is known, (ii) that only forward trajectories are considered, and (iii) that we have enough observables (with hopefully complementary information) as input data. Furthermore the choice of parameters in this example might have been beneficial for the excellent performance of the method. To discuss the virtues and shortcomings of the method, in what follows we want to relax these assumptions and discuss more general situations.

1. Choice of model

In designing the model, we have assumed a four-state system showing only forward transitions. In a given application, however, it might be difficult to appraise if these presumptions are true. To validate the assumption of four states within the MaxCal inference approach, one can simply try to predict the observables by using a model with less or more states. In our example, the existence of two time scales for the observables R_α and R_β clearly implies a minimum number of three states for the Markov model. Hence, we tried to find the (R_α,R_β) coordinates of a single intermediate state which reproduces the correct input data $R_\alpha(t)$ and $R_\beta(t)$. (The coordinates of the initial and final states are fixed by the well-defined initial and equilibrium conditions, respectively.) Numerical studies showed that a three-state model can reproduce the four-state results only in very special cases (e.g., when the two intermediate states have very similar coordinates). On the other hand, one can certainly find models with more states that reproduce the given observables. Hence, the usual procedure is to search for the simplest model (with the least number of states) that does the job.

2. Effective versus diffusive trajectories

As a second major assumption, we have restricted the discussion to unidirectional trajectories, that is, we have neglected possible backflow between the states. To study the performance of the MaxCal inference in the more general case, we adopt a model which includes backflow to the transitions of our original model, see Fig. 3(A). Choosing the backflow rates smaller than the forward rates, one obtains quite similar results for the observables [Fig. 3(B)] and the state populations [Fig. 3(C)] with above. Interestingly, the MaxCal inference reproduces the dynamics just as well as for the unidirectional model. It should be stressed, though, that the numerical effort to propagate unidirectional trajectories only scales as $\propto T^2$ (with T being the number of time steps), while the effort for a general N-state system scales as $\propto N^T$. In practice, it is therefore advantageous to construct an "effective" unidirectional model, whose forward rates are somewhat smaller in order to account for backflow effects.

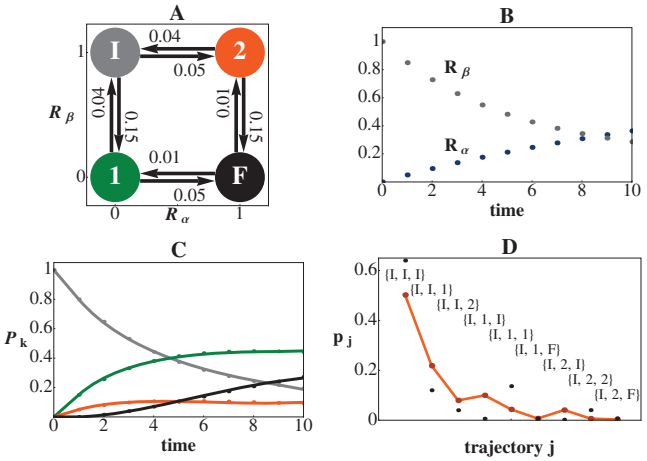


FIG. 3. (a) A four-state model that includes backflow between the transitions yields (b) a time evolution of the observables $R_m(t)$ and (c) the state populations $P_k(t)$ that is similar to the results for the unidirectional model in Figs. 1 and 2. (d) The trajectory probability distribution after the first two time steps, however, is different for the two models. In all cases, reference calculations (dots) are compared to MaxCal inference results (full lines).

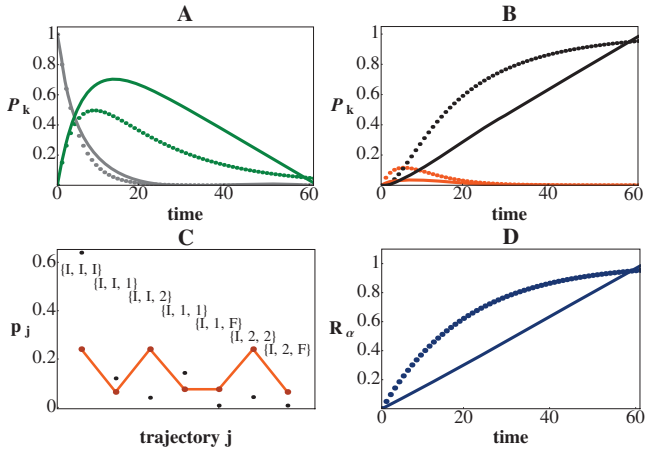


FIG. 4. [(a) and (b)] MaxCal prediction of the state probabilities $P_k(t)$ of the four-state model shown in Fig. 1, assuming that only a single observable, R_β, is known. Although the overall trends of $P_k(t)$ of the reference calculation (dots) are recovered by the MaxCal results (full lines), (c) the inferred trajectory distributions after the first two time steps differs significantly. (d) As a consequence, MaxCal is able to predict only roughly the other reaction coordinate $R_\alpha(t)$.

It is interesting to note that although the state populations are perfectly reproduced, the MaxCal inference yields a trajectory probability distribution which is quite different from the transition matrix based result, see Fig. 3(D). That is, all trajectories that remain in a state [e.g., $(\mathbf{I},\mathbf{I},\mathbf{I})$ and $(\mathbf{I},\mathbf{1},\mathbf{1})$] are underestimated, while all hopping trajectories [e.g., $(\mathbf{I},\mathbf{1},\mathbf{I})$] are overestimated. This reflects the fact that the
MaxCal principle aims to keep the trajectory probability distribution as smooth as possible, while the model under consideration is characterized by relatively high metastabilities. To quantify this effect, one may define the variance σ_j of a trajectory as its number of transition divided by its duration T. Calculating the macroscopic average via $\sigma=\Sigma p_j\sigma_j$, we obtain $\sigma=0.59$ from the transition matrix calculation and $\sigma=0.96$ from the MaxCal inference.

The example shows that one needs to be cautious with the interpretation of dynamical processes in terms of mean values only. Since there is no evidence of high metastability from the input data, an unbiased method such as MaxCal correctly predicts random walklike rather than metastable trajectories, i.e., trajectories that move less. The necessary information to discriminate the two scenarios of "effective versus diffusive" trajectories can be obtained from single-molecule experiments. In fact, by monitoring the time evolution of a single colloidal particle in a double-well potential, Wu *et al.* succeeded to experimentally obtain the first few moments of the underlying trajectory distribution, which were found to be in excellent agreement with accompanying MaxCal calculations.[28] We note in passing that one can implicitly enforce the metastability of the system by restricting the model to unidirectional trajectories. This is why the trajectory distributions of MaxCal inference and transition matrix coincide well in our first example.

3. Inference from insufficient information

Whether or not the result of an inference matches the original data depends on whether or not the data characterize the system sufficiently and in a way that the optimization method "thinks" that the original source is the most reasonable one, according to its rationale. In the case of the four-state system considered above, we have found that the two constraints, R_α and R_β, are sufficient to achieve a correct modeling of the state populations. To study the performance of the MaxCal inference method in the case of insufficient input data, in the following we only use one observable, R_β, see Fig. 4. Projecting the (R_α,R_β)-map in Fig. 1 on the remaining reaction coordinate R_β, we find that states **I** and **2** as well as states **1** and **F** overlap. As a consequence, the inference scheme assigns the same probabilities to a trajectory that switches within the pairs of indistinguishable states or remains in one of these states. This can be seen from the inferred trajectory distribution after the first two time steps [Fig. 4(C)], which assigns trajectories $(\mathbf{I},\mathbf{I},\mathbf{I})$, $(\mathbf{I},\mathbf{I},\mathbf{2})$, $(\mathbf{I},\mathbf{1},\mathbf{1})$, and $(\mathbf{I},\mathbf{2},\mathbf{F})$ equal probabilities.

As a considerable failure of the inference, moreover, one would expect that at long times the system ends up to equal parts in the indistinguishable states **1** and **F**. Nonetheless, the MaxCal prediction of the state probabilities [Figs. 4(A) and 4(B)] is still in qualitative agreement with the reference data and also forecasts the time evolution of the other reaction coordinate $R_\alpha(t)$ [Fig. 4(D)] at least roughly. In particular, the system decays completely in the final state **F** at long times. The latter finding is a consequence of our simplifying assumption on unidirectional trajectories, which affects that for $t\to T$ all trajectories populate state **F** except for three which remain in the other three states. Due to the normalization of the trajectory probabilities, the relative population of the intermediate states goes to zero. Including backflow to the model, however, MaxCal predicts at long times a 50/50 population of the indistinguishable states **1** and **F** as expected.

4. Information content of input data

To study how well the macroscopic averages describe the underlying process, we have tested numerous sets of

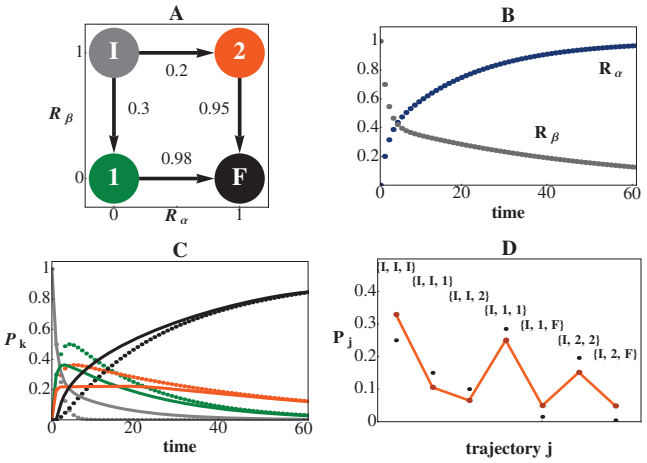


FIG. 5. (a) A model with large rates for the transitions from the initial state **I** to the intermediate states **1** and **2** and small rates for the transitions from the intermediates to the final state **F** gives rise to biexponential behavior of the observables $R_m(t)$ shown in (b). Due to the high metastability of the intermediate states, however, there is no clear indication of the initial decay of state **I** or the increase of state **F**. (d) Hence the trajectory distribution overestimates the populations of the initial and final states and underestimates the populations of the intermediate states. Moreover, the MaxCal state populations (c) predict a biexponential (and thus slower) decay of the initial state, which in turn also causes the populations of the intermediates to rise too slow.

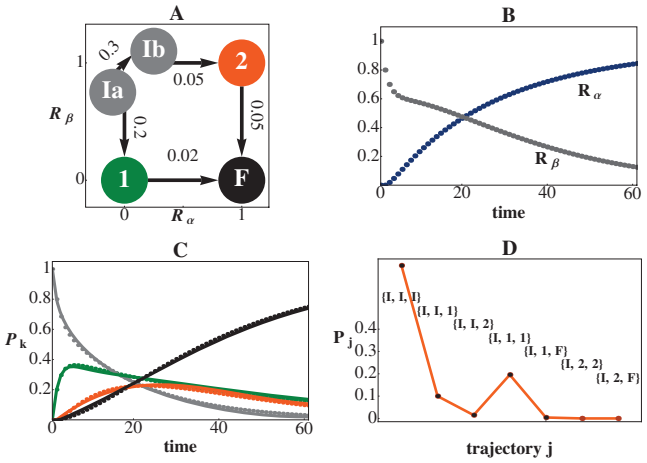


FIG. 6. (a) Hidden state model with state **I** consisting of two indistinguishable substates **Ia** and **Ib** which evolve differently in time. This gives rise to non-Markovian population dynamics of states **I** and **1** [(panel (c)], which in turn causes biexponential behavior of the observables $R_m(t)$ shown in panel (b). The trajectory distribution in panel (d) reveals that the MaxCal results (full lines) perfectly match the reference data (dots).

model parameters. In general it was found that the chosen situation of two given averages, R_α and R_β, is sufficient information to obtain at least qualitative description of the original four-state dynamics. As a challenge in that respect, in the following we consider a system whose macroscopic averages show biexponential kinetics, which may suggest that the state populations decay or rise biexponentially too. As shown in Fig. 5, though, biexponential behavior of the observables can also be obtained by assuming large rates for the transitions from the initial state to the intermediate states and small rates for the transitions from the intermediates to the final state. For short times, $R_\alpha(t)$ changes along the **I** $\rightarrow$ **1** transition while $R_\beta(t)$ changes along the **I** $\rightarrow$ **2** transition.

The data-driven MaxCal results, on the other hand, predict a biexponential (and thus slower) decay of the initial state, which in turn causes the populations of the intermediates to rise too slow. This behavior is again explained by the trajectory distribution shown in Fig. 5(D). Due to the high metastability of the intermediate states, there is no clear indication that state **I** has decayed after a short period. Since the MaxCal principle aims to keep the trajectory probability distribution as flat as possible, it therefore overestimates the population of the initial state and underestimates the population of the intermediate states. Similarly, because of the high metastability of the intermediate states, the population flow into state **F** is overestimated. Again, we find that caution is required when we attempt to interpret dynamical information that merely reflects the average behavior of the system.

C. Inference of non-Markovian dynamics

Although the MaxCal method has proven to work well for the variety of examples discussed above, the inference of discrete Markov processes can be achieved with less effort by maximum likelihood methods, which explicitly exploit the Markovian nature of the dynamics.[9,10] To demonstrate the potential of the MaxCal approach method, we therefore present two applications that defy a straightforward Markov state modeling.

1. Hidden state model

To provide a challenge for the MaxCal inference approach, we have considered in Fig. 5 a biexponential behavior of the observables $R_m(t)$, which was nonetheless caused by a standard Markov state model. Let us now consider the—maybe more common—case that the population dynamics of the system itself is non-Markovian and therefore causes biexponential kinetics of the observables. Biexponential population dynamics may occur, for example, if a state of the model (say, state **I**) actually consists of two substates (**Ia** and **Ib**), which are indistinguishable by the available observables (i.e., one of them is "hidden") but evolve differently in time. This gives rise to non-Markovian dynamics of the resulting four-state system in the sense that this dynamics cannot be constructed by a 4×4 transition matrix as there are two rates that describe the decay of state **I**. (The dynamics of the full five-state system is still Markovian, of course.) Figure 6 shows such a scenario where we assume that we start at $t=0$ in state **Ia**, which then decays rapidly into states **Ib** and **1**. After a few time steps, the population of state Ia has decayed, which is reflected in an abrupt stop of the population increase of state **1**. At the same time, state **Ib** decays slowly into state **2**, which acts as a kinetic trap. Due to the high metastability of the intermediate states, we observe a slow increase of the population of the final state **F**. By construction, we find for this example a somewhat delayed, slow rise of $R_\alpha(t)$, while the non-Markovian population dynamics of state **I** causes the biexponential kinetics of $R_\beta(t)$, see Fig. 6(B).

From the predicted state populations and the trajectory distribution displayed in Figs. 6(C) and 6(D), we find that the

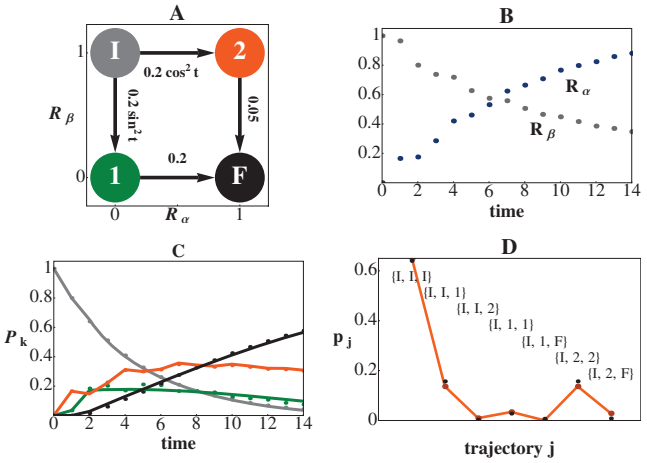


FIG. 7. (a) Model with time-dependent transition rates, in which the outgoing flux from the initial state **I** alternates in time between intermediates **1** and **2**. Shown are the resulting observables (b) $R_m(t)$, (c) the state populations $P_k(t)$, and (d) the trajectory distribution. The MaxCal results (full lines) are seen to be in excellent agreement with the reference data (dots).

MaxCal inference of the non-Markovian dynamics perfectly matches the reference results. This is because the two transitions, $\mathbf{I}\rightarrow\mathbf{1}$ and $\mathbf{I}\rightarrow\mathbf{2}$, are well described by the observables $R_\beta(t)$ and $R_\alpha(t)$, respectively. In other words, if the input data exhibit biexponential kinetics, MaxCal also assumes biexponential state dynamics. The example shows that MaxCal does not need to know the "full" state space (in which all transitions can be described by rate equations) because the approach does not assume Markovian dynamics at any point.

2. Time-dependent transition rates

Arguably the most general dynamical behavior of a few-state system can be obtained by allowing the transition rates to become time-dependent, e.g., driven by an external source. Figure 7 shows a simple example for this phenomenon where the outgoing flux from initial state **I** alternates in time between intermediates **1** and **2**. As a consequence, we find oscillations of the state populations of the intermediates as well as of both observables $R_\alpha(t)$ and $R_\beta(t)$. Since the dynamics is well accounted for by the observables, the MaxCal inference nicely describes the oscillating populations of the intermediates. Again, the bias-free ansatz of the MaxCal approach facilitates to infer even very general dynamical behavior.

IV. CONCLUDING REMARKS

Jaynes[32] once noted that his important research papers such as the maximum entropy formulation of statistical mechanics[14] or the Jaynes–Cummings model in quantum optics[33] usually went through a 20-year incubation period before they became widely recognized. The MaxCal principle was proposed by Jaynes even 30 years ago and only recently the elegance and power of this appealing formulation has been elucidated. In this work we have applied the MaxCal formulation to the inference of few-state nonequilibrium processes. Maybe the most important result of our study is that this fairly abstract and general variational principle, which at a first sight seems to be mostly of formal interest, indeed works in practice to solve inference problems! As the Caliber function is given as an integral over all possible paths, the numerical effort for a general N-state system scales exponentially as $\propto N^T$, with T being the number of time steps. Dealing with metastable systems, it is therefore advantageous to construct an effective unidirectional model, which only scales as $\propto T^2$. Alternatively, one could adopt a stochastic sampling method such as a Gillespie-type algorithm,[34] in order to obtain an ensemble of trajectories which subsequently is used to maximize the caliber functional.

The performance of the approach has been demonstrated by adopting various few-state models of increasing complexity including several Markov models, a hidden-state model, and an externally driven model. In all cases, it has been found that the MaxCal inference scheme is fairly robust and yields correct results as long as the input data are sufficient. Since the MaxCal principle aims to keep the trajectory probability distribution as flat as possible, the method is "democratic" in that it attempts to treat every possible transition with equal probability, as long as this fits the available information. While the method relies on the content of the given information, it is completely unbiased with respect to the type of dynamics such as Markovianity. Therefore it can, in principle, reproduce any kind of time dependencies such as oscillatory transients and multitime decays. Finally, our study has demonstrated that one needs to be cautious with the interpretation of dynamical information that merely reflects some average behavior of the system. Characteristic behavior such as multitime decays can be exhibited by very different systems, thus hampering a simple identification of the underlying process.

ACKNOWLEDGMENTS

We thank Ken Dill and Kingshuk Ghosh for numerous inspiring discussions and Ken Dill, moreover, for his kind hospitality during our visit in summer 2007. This work has been supported by the Frankfurt Center for Scientific Computing, the Fonds der Chemischen Industrie, and the Deutsche Forschungsgemeinschaft.

[1] For a recent review, see S. Mukamel, Y. Tanimura, and P. Hamm, Acc. Chem. Res. **42**, 1207 (2009).
[2] J. N. Onuchic, Z. L. Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[3] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[4] M. Gruebele, Curr. Opin. Struct. Biol. **12**, 161 (2002).
[5] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
[6] G. Gfeller, P. De Los Rios, A. Caflisch, and F. Rao, Proc. Natl. Acad. Sci. U.S.A. **104**, 1817 (2007).
[7] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Elsevier, Amsterdam, 1994).
[8] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, J. Chem. Phys. **126**, 155101 (2007).
[9] F. Noé, J. Chem. Phys. **128**, 244103 (2008).
[10] S. Bacallado, J. Chodera, and V. Pande, J. Chem. Phys. **131**, 045106 (2009).
[11] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. Weikl, Proc. Natl. Acad. Sci. U.S.A. **106**, 19011 (2009).
[12] M. Sarich, F. Noé, and C. Schütte, "On the approximation of Markov state models," Multiscale Model. Simul. (in press).

[13] A. Altis, M. Otten, P. H. Nguyen, R. Hegger, and G. Stock, J. Chem. Phys. **128**, 245102 (2008).
[14] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[15] See: http://bayes.wustl.edu/ for numerous helpful references on the Maximum Entropy formalism.
[16] K. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology* (Garland Science, New York, 2003).
[17] B. Buck, in *Maximum Entropy in Action*, edited by B. Buck and V. A. Macaulay (Oxford University Press, Oxford, 1991), p. 248.
[18] M. Habeck, W. Rieping, and M. Nilges, J. Magn. Reson. **177**, 160 (2005).
[19] E. T. Jaynes, in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT Press, Cambridge, 1978), p. 15.
[20] E. T. Jaynes, Annu. Rev. Phys. Chem. **31**, 579 (1980).
[21] E. T. Jaynes, in *Complex Systems—Operational Approaches*, edited by H. Haken (Springer, Berlin, 1985), pp. 254–269.
[22] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[23] H. Haken, Z. Phys. B: Condens. Matter **63**, 505 (1986).
[24] J. P. Dougherty, Philos. Trans. R. Soc. London, Ser. A **346**, 259 (1994).
[25] K. Ghosh, K. Dill, M. M. Inamdar, E. Seitaridou, and R. Phillips, Am. J. Phys. **74**, 123 (2006).
[26] E. Seitaridou, M. Inamdar, R. Phillips, K. Ghosh, and K. Dill, J. Phys. Chem. **111**, 2288 (2007).
[27] G. Stock, K. Ghosh, and K. A. Dill, J. Chem. Phys. **128**, 194102 (2008).
[28] D. Wu, K. Ghosh, M. Inamdar, H. J. Lee, S. Fraser, K. Dill, and R. Phillips, Phys. Rev. Lett. **103**, 050603 (2009).
[29] R. Blossey, *Computational Biology: A Statistical Mechanics Perspective* (Chapman and Hall, New York, 2006).
[30] D. A. Beard and H. Qian, *Chemical Biophysics: Quantitative Analysis of Cellular Systems* (Cambridge University Press, Cambridge, 2008).
[31] L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).
[32] E. T. Jaynes, *Physics and Probability* (Cambridge University Press, Cambridge, 1993), p. 261.
[33] E. Jaynes and F. Cummings, Proc. IEEE **51**, 89 (1963).
[34] D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007).