Single molecule kinetics. II. Numerical Bayesian approach

James B. Witkoskie and Jianshu Cao
Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 11 February 2004; accepted 2 July 2004)

As discussed in the companion paper [J. B. Witkoskie and J. S. Cao, J. Chem. Phys. **121**, 6361 (2004), preceding paper], quantitative extraction of information from single molecule experiments by several proposed indicators is difficult since the experiments only observe certain characteristics of the system, even though the indicators can contain all available information. This paper shows how one can circumvent the shortcomings of these indicators by combining information extracted from indicators with a numerical Bayesian statistical approach. The Bayesian approach determines the relative probability of various models reproducing the entire sequence of the single molecules trajectory, instead of binning and averaging over the data, which removes much of this information.
 [DOI: 10.1063/1.1785784]

I. INTRODUCTION TO APPLICATION OF BAYESIAN STATISTICS

As discussed in a companion paper,[1] a formidable theoretical obstacle facing single molecule analysis is the extraction of pertinent information about stochastic chemical processes from noisy data. Several authors have proposed simple indicators to attempt to reveal these processes.[1] Generally, these indicators give qualitative indications of deviations from simple Poisson or renewal processes (i.e., simple single waiting time processes), but quantitative extraction of the characteristics of these processes is difficult if not impossible. The major obstacles to using these indicators are the data binning and data averaging that they require, which remove much of the available information. The indicators may also be sensitive to instrument resolution and background counts, whose effects may be difficult to incorporate in the analysis.

Despite these difficulties, these indicators can give valuable qualitative insight into the behavior of the system and can be used to reduce the number of possible models that describe the system. After developing this insight and reducing the number of models, a more robust numerical routine that does not require the reduction of information through data binning or data averaging is required to move beyond a qualitative description of the system. One candidate numerical routine that can perform a complete sequence analysis is Markov Chain Monte Carlo with Bayesian statistics.[2] Markov Chain Monte Carlo with Bayesian statistics starts from the use of Bayes formula, which compares the relative probability that two models, M and M', reproduce a data set,

$$\frac{P(M|D)}{P(M'|D)}=\frac{P(D|M)P(M)}{P(D|M')P(M')}. \tag{1}$$

In this expression $P(M|D)$ is the posterior probability of the model given the data, $P(D|M)$ is the probability of the data given the model, and $P(M)$ is the *a priori* probability of the model, which is often assumed to be uniform or log uniform or a conjugate prior.[3,4] In other words, from our previous experience we start with a given distribution of possible models, $P(M)$. After examining the data through the Bayesian formula, we modify our guess about the possible model parameters, which results in the new distribution $P(M|D)$. This quantity tells us about our certainty in predicting the values of parameters.

As a demonstration of the philosophical approach to single molecule problems, we will apply the Bayesian method to the simple four state model that we examine in the companion paper.[1] The kinetic scheme is outlined in Fig. 1. The model has four states and the interconversion between the states is governed by Poisson kinetics. Two states are "bright" with labels "$b1$" and "$b2$," and two states are "dark" with labels "$d1$" and "$d2$." Although we know that the molecule is either bright or dark at any instant in time, we do not know the exact state of the molecule, i.e., is it in state $b1$ or $b2$. This scenario is one of the simplest hidden Markov processes, which we discuss to some extent in the companion paper.[1] The equation for the probability density in this simple four state model is given by the simple kinetic equation,

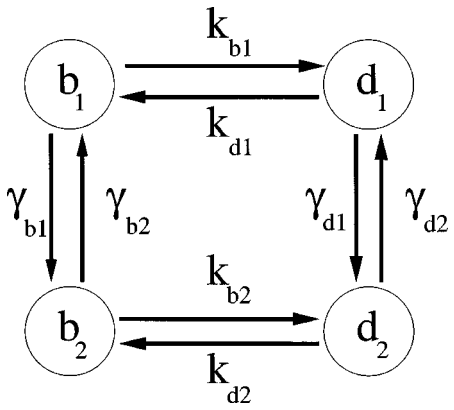

FIG. 1. A diagrammatic depiction of the kinetic scheme analyzed in this paper and the companion paper (Ref. 1).

$$\begin{bmatrix} \dot{\rho}_{b1} \\ \dot{\rho}_{b2} \\ \dot{\rho}_{d1} \\ \dot{\rho}_{d2} \end{bmatrix} = \begin{bmatrix} -(\gamma_{b1}+k_{b1}) & \gamma_{b2} & k_{d1} & 0 \\ \gamma_{b1} & -(\gamma_{b2}+k_{b2}) & 0 & k_{d2} \\ k_{b1} & 0 & -(\gamma_{d1}+k_{d1}) & \gamma_{d2} \\ 0 & k_{b2} & \gamma_{d1} & -(\gamma_{d2}+k_{d2}) \end{bmatrix} \begin{bmatrix} \rho_{b1} \\ \rho_{b2} \\ \rho_{d1} \\ \rho_{d2} \end{bmatrix}. \tag{2}$$

For numerical calculations we use the constants, $k_{b1}=0.75\,\mathrm{s}^{-1}$, $k_{d1}=0.50\,\mathrm{s}^{-1}$, $K_{b1/b2}=0.44$, $k_{b2}=k_{b1}K_{b1/b2}=0.33\,\mathrm{s}^{-1}$, $k_{d2}=k_{d1}K_{b1/b2}=0.22\,\mathrm{s}^{-1}$, and $\gamma_{b1}=\gamma_{b2}=\gamma_{d1}=\gamma_{d1}=\gamma=0.1\,\mathrm{s}^{-1}$. The four parameters we examine are k_{b1}, k_{d1}, $K_{b2/b1}$, and γ. The numerical values of these kinetic constants are chosen so that the waiting time distribution for both the bright and dark states are obviously not monoexponential, which we determined from the indicators presented in the companion paper,[1] but the rate of modulation between the two states does not cause apparent time separation in any of the indicators. We also perform the analysis with all of the parameters free, but this information is difficult to present in a visual form so we concentrate on the four parameter fit.

From the model system with the specified parameters, we generate a sequence of bright and dark states for 25 molecules with a duration of 300 s, which allows about 150 observed turnover events per molecule and 7500 pieces of data to determine the four parameters. The simulation details used to generate this data are presented in part 1a of the Appendix. These data sets are much smaller than the data sets collected in the experiments by Lu, Xun, and Xie.[5] By applying the ''event'' correlation indicator discussed in the companion paper[1] to this data, one is able to deduce that the bright and dark decays can both be fit with biexponentials which suggests that they both contain two states. The event density also indicates a memory effect, which is confirmed by the characteristic function. This information allows one to reasonably suggest the four state model as a candidate to describe the system.

Given the kinetic scheme, we can simply calculate $P(D|M)$ through iterative matrix multiplication. Given the initial state of the system, bright or dark, and the times of the transitions $\{t_i\}$ the probability of the data given the model is

$$P(D=\{t_i\}|M)=1^T\left[\prod_i \mathbf{K}e^{(\mathbf{K}_d+\Gamma)t_i}dt_i\right]\delta^{(\pm)}\rho_{\mathrm{eq}}, \tag{3}$$

where the matrix definitions follow those used in the companion paper and previous work.[1,6–8] $\delta^{(\pm)}$ is determined by the initial condition. Since we can calculate the relative probabilities, we can perform Monte Carlo on the probabilities to determine models that are consistent with the data. The exact method of calculating the probabilities and performing the Monte Carlo simulation are outlined in part 2 of the Appendix.

The approach can also be extended to experiments with more complex data, such as photon counting statistics discussed briefly in the companion paper,[1,9] and can incorporate data deficiencies or statistical fluctuations from sources other than the system such as instrument resolution and response.[9,10] For most of the analysis we assume the switching between states is sufficiently fast so that these times are sharp variables relative to the duration of the bright and dark states, but generalizations can easily be made in this framework since we do not require the system to have Poisson kinetics.[11] From our proposed model, we determine the probability of the initial condition and the probability of the transitions at the recorded times. The probability of the sequence is the result of iterative matrix multiplication.

For many optimization applications, the Monte Carlo approach avoids difficulties associated with gradient based likelihood maximization. Gradient based maximum likelihood approaches successfully determined point estimates of the most likely set of parameters for single membrane ion channel experiments, but the calculation only determines the best fitting parameters and the curvature of the likelihood function, $P(D|M)$ at this point.[12] The Monte Carlo approach can show more detail in the probability distribution, such as multiple minima with similar probabilities.[11] The shot noise in the data creates large uncertainties and make these maximum likelihood estimates inaccurate so a global estimation of the parameter distribution becomes important. Often, if the data are not sufficient or other difficulties arise that prevent the system from finding the most likely parameters, the probability density signals these difficulties.

II. APPLICATION OF BAYESIAN APPROACH

We applied the Bayesian algorithm outlined in part 2 of the Appendix to the data set generated by the algorithm discussed in part 1a of the Appendix. Since the kinetic rates are positive quantities and we do not initially know the magnitude of these rates a natural initial (*a priori*) distribution for the kinetic rates is log-uniform, $P(M)\propto 1/M$.[3] As a result the Monte Carlo jump sizes are proportional to the magnitude of the kinetic rates. This log-uniform distribution is not normalizable (improper), but the likelihood $P(D|M)$ will give the necessary truncation to prevent any problems associated with this normalization. Other *a priori* distributions are possible based on the analysis with other indicators. From the multiexponential fit determined by the event correlations, one may want to restrict the eigenvalues of the matrix to a small interval around the fitted parameters.

Figure 2 shows two-dimensional projections onto the principal axes of the posterior probability density for the parameters determined from a data set, and Table I contains the mean and covariance of these parameters. The predicted mean values for these constants are very accurate and the variances are fairly small, which shows that the data are sufficient to determine these parameters. We note that the maximum likelihood value is not necessarily in the center of the distribution, which shows that the distribution is not

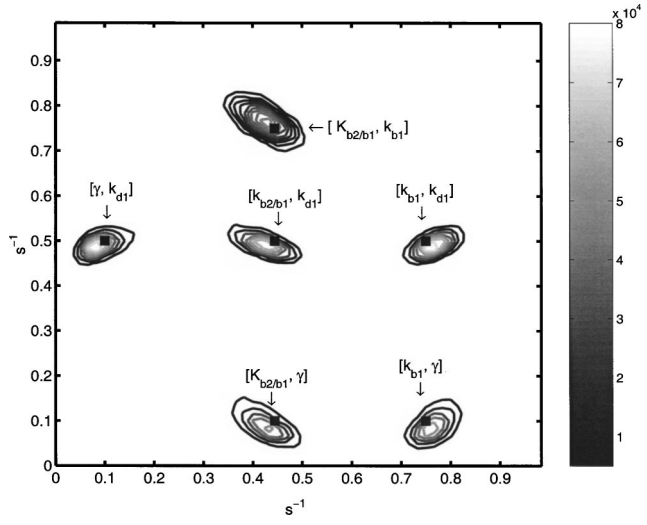

FIG. 2. Contour plots of two-dimensional projections onto the principal hyperplanes of the probability density generated by the Monte Carlo simulation that determines the four parameters for the model in Eq. (2). The gray scale is measured in number of points for $\approx 5\times 10^6$ Monte Carlo Samples. The black squares mark the positions of the parameters that generated the data. Each peak is pointed to by a two component label for that peak, such as $[\gamma, k_{d1}]$. The first entry corresponds to the horizontal axis and the second entry corresponds to the vertical axis. $K_{b2/b1}$ has the largest uncertainty since it must adjust itself to fluctuations in k_{b1} and k_{d1}.

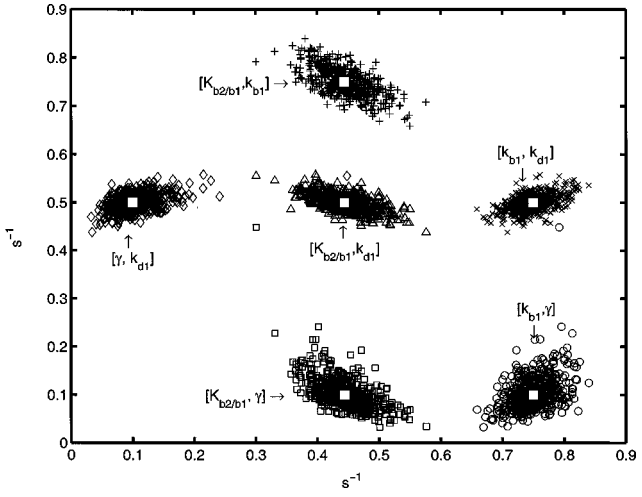

FIG. 3. Two-dimensional projections onto the principal hyperplane of the vectors of the best fitting parameters encountered during a Monte Carlo simulation for 500 different data sets. In comparison, Fig. 2 is the total probability determined from a single data set, but this plot shows the peak position for several data sets. The optimal parameters are distributed around the white squares that label the parameters that generated the data sets. Different symbols are used for each projection to show the single outlier. Similar to Fig. 2, each cluster is pointed to by a label such as $[\gamma, k_{d1}]$. The first entry corresponds to the horizontal axis and the second entry corresponds to the vertical axis.

Gaussian. Because the eigenvalues and eigenvectors depend on the parameters through an inherently nonlinear functional form, the asymmetry is not surprising. The largest uncertainty is in the constant $K_{b2/b1}$. This constant enters into the determination of the eigenvalues through multiplication with k_{b1} and k_{d1} so the additional uncertainty comes from compensation for fluctuation in these other quantities.

The predicted maximum likelihood estimate is slightly offset from the real parameters. Many sets of data examined in simulations converge to a roughly Gaussian distribution with a mean that is slightly offset from the actual parameters although some simulations fail to converge and a few simulations converged to a set of parameters that are far from the actual set of parameters used to generate the data. Even for data sets with as few as five molecules, many simulations predict a maximum likelihood estimates around the actual parameters, but more importantly, convergence to parameters far removed from the actual parameters are rare.

TABLE I. Mean and covariance for the Monte Carlo simulation presented in Fig. 2. Stability analysis determines normal modes with standard deviations of 5.1, 2.1, 1.8, 1.3×10^{-2}.

			Actual values	
	k_{b1}	k_{d1}	γ	$K_{b2/b1}$
	0.75	0.50	0.10	0.44
			Mean values	
	k_{b1}	k_{d1}	γ	$K_{b2/b1}$
	0.77	0.50	0.10	0.43
			Covariance $\times 10^3$	
	k_{b1}	k_{d1}	γ	$K_{b2/b1}$
k_{b1}	0.86	0.33	0.42	−0.73
k_{d1}		0.39	0.27	−0.50
γ			0.89	−0.74
$K_{b2/b1}$				1.4

The most likely point estimates encountered during a Monte Carlo simulation for 500 different sets of data with 25 molecules and 300 s trajectories are plotted in Fig. 3. Although the Bayesian philosophy concerns determining the entire probability distribution, the point estimates give good insight into the reproducibility of the simulation. From this plot it is apparent that the predicted maximum likelihood parameters are distributed around the actual parameters, except for one outlying data point. The distribution of the maximum likelihood estimates resemble the distributions of the probability distribution of the parameters for a single set of data. The stochastic nature of the underlying dynamics causes these offsets. This noise makes the use of simple maximum likelihood point estimates of the parameters statistically uncertain unless other analysis is performed.

To help determine the offset caused by the noise in real single molecule experiments, we can break a large data set into several subsets and perform optimization of the parameters with these subsets and compare optimal parameters, this is known as cross validation.[13] For this application, breaking data sets up by taking segments of single molecule sequences or by performing the analysis on different single molecule sequences has about the same effect on the cross validation. In fact, future analysis concerns a data set that contains a single long sequence.[9] Mixing these subsets will improve our predictions of offsets and allow us to understand the sensitivity of parameters to the intrinsic noise in these systems. Breaking the data up into smaller subsets will also allow the simulation to search larger regions of parameters space since the sensitivity of the Bayesian score scales linearly with the length of the sequence and number of mol-

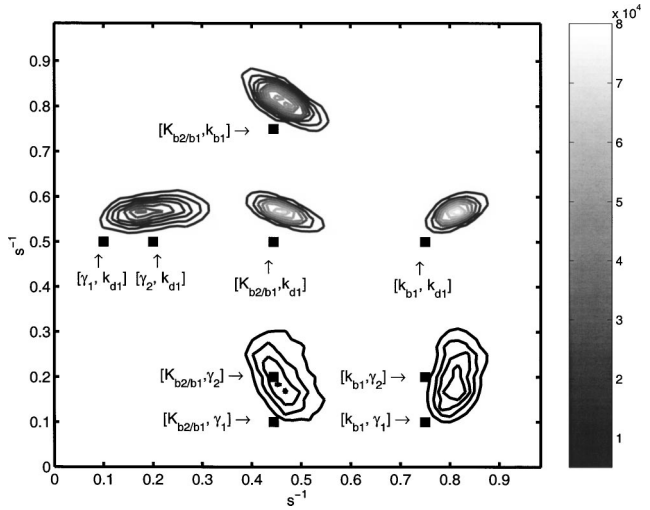

FIG. 4. Contour plots of two-dimensional projections of the probability density generated by the Monte Carlo simulation that determined the four parameter fit to the five parameter model discussed in Sec. II. The black squares mark the positions of the parameters that generated the data. The distribution of γ is much wider than the distribution presented in Fig. 2, which reflects the fact that this parameter is not defined for the model that generated the data. Similar to Fig. 2, each cluster is pointed to by a label such as $[\gamma, k_{d1}]$. The first entry corresponds to the horizontal axis and the second entry corresponds to the vertical axis.

ecules. In many way, the length of the sequences corresponds to a fictitious inverse Boltzmann temperature.

A. Variations of the simulation

The Monte Carlo simulation generally failed to converge to any value when the data sets are not consistent with any parametrization of the model. We tested this property by trying to fit the simple four parameter model to a sequence created from a stretched exponential waiting time distribution. Both the bright and dark waiting time distributions are given by $9/8e^{-[(9)/4t]^{1/2}}$, which has a characteristic time of 8/3, which is comparable to $k_{b1}^{-1}+k_{b2}^{-1}/2$ in Eq. (2). When the model tried to fit this data, it would either set many of its parameters to zero (large negative values for the log of the parameters) or wander through the parameter space without converging.

For situations where the difference between the proposed model and the actual kinetics is not as great as the stretched exponential, the manifestation of errors are more subtle. We demonstrate the subtlety by attempting to fit data generated from a five parameter model with the original four parameter model. The five parameter model is also a four state model with the same k_{b1}, k_{d1}, and $K_{b2/b1}$, but two different values for the γ's. For this demonstration, we set $\gamma_{b1}=\gamma_{d1}=0.1\,\mathrm{s}^{-1}$ and $\gamma_{b2}=\gamma_{d2}=0.2\,\mathrm{s}^{-1}$. Figure 4 and Table II summarize the results of one of the best fits of this optimization. Although all of the parameters are shifted relative to the true parameters, the greatest uncertainty appears in the single γ since this quantity is not well defined in the model that generated the data.

As with all parametrizations, the data requirements scale with the number of parameters. For more complex models, more data may be needed. Since the amount of data serves

TABLE II. Mean and covariance for the Monte Carlo simulation presented in Fig. 4. Note that the data are generated from a model with two γ's, but the fit is performed with a single γ. Stability analysis determines normal modes with standard deviations of 6.7, 4.2, 1.8, 1.6×10^{-2}. Because the model with a single γ did not generate the data, there is a fairly large variance for γ relative to the other parameters.

			Actual values	
	k_{b1}	k_{d1}	γ	$K_{b2/b1}$
	0.75	0.50	0.10/0.20	0.44
			Mean values	
	k_{b1}	k_{d1}	γ	$K_{b2/b1}$
	0.82	0.57	0.22	0.47
			Covariance $\times10^3$	
	k_{b1}	k_{d1}	γ	$K_{b2/b1}$
k_{b1}	0.88	0.33	0.52	−0.82
k_{d1}		0.42	0.35	−0.56
γ			4.0	−0.9
$K_{b2/b1}$				1.5

the role of temperature in the sampling of a partition function, large amounts of data correspond to a very low temperature and can result in a failure to explore the relevant parameters space due to trapping in local minima. This trapping phenomenon is analogous to diffusion on a rough potential energy surface.[14] We found that the use of 250 or more molecules with a trajectory length of 300 s leads to trapping in local minimum. This trapping shows that the surface is generally not monotonically decreasing to the global minimum. One should use standard approaches such as annealing to help the system find the global minimum.[15] One method of incorporating annealing is the addition and mixing of data during the simulation, i.e., exchanges data used in the optimization with unused data during the simulation. The mixing is important to avoid certain pieces of data from dominating the optimization and preventing convergence to parameters that are consistent with all of the data.

The effects of excessive data, such as trapping in local minima, are not a consideration even if we perform the optimization of all eight parameters with no constraints. Using 50 molecules, instead of 25, we generated data with the four parameter model, but let all eight parameters vary independently and then imposed $k_{b1}>k_{b2}$ at the end of each Monte Carlo iteration. Most of the simulations are able to locate a global minimum that is near the actual parameters. The results for one of these simulations is presented in Table III. The largest standard deviation in a mode is ~8.5%, which is a typical value for most simulations. For 25 molecules, the largest standard deviation of one of the normal modes was around ~15%–20%, which shows that the certainty in the parameters improve roughly as the expected $1/\sqrt{n}$ scaling as more data are added.

B. Four state model and the fluctuating bottleneck

The four state model is often compared with the fluctuating bottleneck model (i.e., the diffusion controlled reaction) in the literature.[6–8,16–18] The fluctuating bottleneck

TABLE III. Mean and covariance for the Monte Carlo simulation with eight parameters. Stability analysis determines the normal modes with standard deviations of 8.5, 5.8, 4.8, 3.4, 2.1, 1.7, 1.1, 0.9×10^{-2}.

				Actual values				
	k_{b1}	k_{b2}	k_{d1}	k_{d2}	γ_{b1}	γ_{b2}	γ_{d1}	γ_{d2}
	0.75	0.33	0.50	0.22	0.10	0.10	0.10	0.10
				Mean values				
	k_{b1}	k_{b2}	k_{d1}	k_{d2}	γ_{b1}	γ_{b2}	γ_{d1}	γ_{d2}
	0.75	0.32	0.50	0.19	0.12	0.09	0.11	0.14
				Covariance $\times10^3$				
	k_{b1}	k_{b2}	k_{d1}	k_{d2}	γ_{b1}	γ_{b2}	γ_{d1}	γ_{d2}
k_{b1}	3.0	1.0	1.1	0.3	1.8	0.25	0.88	−1.4
k_{b2}		1.5	0.85	0.75	0.78	−0.80	0.37	0.83
k_{d1}			1.0	0.46	0.85	−0.22	6.7	−0.46
k_{d2}				0.87	0.61	−0.19	0.36	−0.92
γ_{b1}					2.7	1.2	0.11	−0.97
γ_{b2}						2.1	0.27	−0.01
γ_{d1}							1.6	0.55
γ_{d2}								2.1

model describes a one-dimensional diffusion process in a harmonic well with a reaction rate that depends harmonically on the coordinate,

$$\partial_t P_{\pm}(t) = D\nabla^2 P_{\pm}(t) + \nabla[kxP_{\pm}(t)] - \kappa_{\pm}x^2 P_{\pm}(t) + \kappa_{\mp}x^2 P_{\mp}. \quad (4)$$

In the companion paper,[1] it was demonstrated that the difference between the characteristic function of the fluctuating bottleneck model with $D=k=\kappa_+=1$ and $\kappa_{-1}=K_{\mathrm{eq}}\kappa_+=2$, and a four state model with $\gamma_{b1}=\gamma_{d1}=0.289\,706$, $\gamma_{b2}=\gamma_{d2}=1.710\,29$, $k_{b1}=1/2k_{d1}=0.417\,953$, $k_{b2}=1/2k_{d2}=4.436\,15$ is small with the maximum deviation in the characteristic function of ≈5%, which is smaller than the approximate noise levels for the reasonable amounts of data.

We consider the time traces of 25 molecules for 100 time units generated from both the fluctuating bottleneck and the four state model. For the given model parameters, each molecule performs ≈150 turn-overs. From this data, we attempt to find a four state model that optimizes the fit to both sets of data. Applications of Bayesian statistics to a continuous diffusion model will be discussed in applications to single photon experiments.[9] Table IV compares the means and covariances for a typical run (typical average value and variance). The first important observation is that the optimal fitting parameters to the fluctuating bottleneck model are different from the parameters suggested by Brown's procedure (even if the amount of data is increased).[16] Maximizing statistical overlap of the sequences between two models is actually a nontrivial problem, and Brown's parametrization only matches correlation functions, which does not necessarily maximize the overlap of probability.

From the covariance matrices it is apparent that k_{b2} and K_{eq} have comparable variances for both data sets, but the variances of k_{b1}, γ_{b1}, and γ_{b2} are over twice as large for the fits to the fluctuating bottleneck model. The larger uncertainty in the parameters can be used as a flag to suggest the exploration of other models, which can then be compared through the use of the Bayesian score. Similar to the choice of optimal parameters, one can use Bayesian statistics to choose from models with different physical features or complexity. Comparison of seemingly disjoint models has a rich history with several aspects including determination of when to increase the number of parameters, so this step is omitted here for brevity, but several references address model comparison through Bayesian methods.[19] Many of these comparisons use the Bayesian score that we calculate to perform our Monte Carlo simulations.

III. CONCLUSION

Single molecule experiments offer an opportunity to gain significant insight into the physics of glasses, biomolecules, and other complex systems. The insight is limited by both the amount of collected data and the analysis performed on the data. Previously proposed indicators and other quantities used in the analysis of single molecule experiments contain useful information that give qualitative insight into the physics of the system, as demonstrated in previous references.[1,6–8] The useful information includes various time

TABLE IV. Mean and covariance for the Monte Carlo simulations attempting to fit the four state model to data generated by a four state model and by the fluctuating bottleneck model. Stability analysis determines the normal modes with standard deviations of 40, 21, 14, 8.6, 4.1×10^{-2} for fitting the four state model to itself and standard deviations of 44, 23, 20, 8.1, 5.8×10^{-2} for fitting the four state model to the fluctuating bottleneck.

			Actual values for four state		
	k_{b1}	k_{b2}	K_{eq}	γ_{b1}	γ_{b2}
	0.42	4.4	2.0	0.29	1.7
			Mean values (four state/bottleneck)		
	k_{b1}	k_{b2}	K_{eq}	γ_{b1}	γ_{b2}
	0.44/0.41	4.6/4.1	2.0/2.0	0.26/0.43	1.7/2.2
			Covariance $\times10^3$ (four state/bottleneck)		
	k_{b1}	k_{b2}	K_{eq}	γ_{b1}	γ_{b2}
k_{b1}	0.39/0.82	1.9/4.9	−0.67/−0.57	−0.20/−0.95	−0.18/−0.21
k_{b2}		43./51.	−8.4/−7.2	−1.0/−4.0	8.4/17
K_{eq}			6.4/6.6	$1.0\cdot10^{-2}/2.0\cdot10^{-2}$	−0.81/−0.45
γ_{b1}				0.55/2.3	1.8/4.5
γ_{b2}					20./40.

constants for the relaxation of the system and connectivity between these relaxation times (i.e., memory effects), but the extraction of quantitative information from these indicators is difficult because the indicators require data binning and data averaging which remove large amounts of useful information.

To move beyond indicator analysis requires a numerical method that does not require ill-conditioned data inversion or averages out information contained in the data. Bayesian analysis with Monte Carlo optimization is one strong candidate. Implementing Bayesian analysis still requires the use of the previous indicators to determine constraints on possible models, such as the number of states or restrictions on eigenvalues, but the Bayesian approach gives quantitative estimates of the parameters and uncertainties in these parameters. By incorporating the constraints discovered through the indicators into the *a priori* distribution of possible models, one can use Monte Carlo with the Bayesian score as the Boltzmann energy to optimize the parameters. Applications of this approach to a simple four state model demonstrate its capabilities to reproduce the correct parametrization from a limited data sets and give uncertainties in these parameters. The Bayesian approach also has the ability to distinguish different models as shown by the comparison of the fluctuating bottleneck and four state models.

Many other scenarios exist and should be explored, such as the role of photon statistics and continuous distributions of states. The Bayesian approach performed well in our tests of some of these other scenarios, but additional analysis requires motivation from applications to real systems so we will not go into detail about other simulations with computer generated data. Preliminary results from the analysis of single photon emission events from fluorescence quenching experiment are very promising.[9,20] Of all the existing approaches, the Bayesian approach is the most reliable and robust method of numerically analyzing single molecule data, and we encourage experimentalists to explore the application of this approach to their single molecule data. The Bayesian approach will become a standard method of single molecule analysis in the future.

ACKNOWLEDGMENTS

We thank Professor X. S. Xie for sharing his manuscript, which also demonstrates the validity of the Monte Carlo approach. This research is supported by the AT&T Research Fund Award, the NSF Career Award (Grant No. Che-0093210), and the Camille and Henry Dreyfus Teacher-Scholar award.

APPENDIX: SIMULATION DETAILS

1. Generation of data

a. Four state model

The four state numerical example that we examine in this paper is depicted in Fig. 1 and discussed in the companion paper.[1] The kinetics for the system are given by the kinetic matrix equation, Eq. (2). From the kinetic rates the steady state solution, ρ_{eq} can be determined. For each of the 25 molecular trajectories the initial state is randomly sampled from this steady state. After choosing the initial state, the time of a transition from this state to either the other connected states is drawn from an exponential distribution with characteristic time $(\gamma_s+k_s)^{-1}$, where $s=b1$, $b2$, $d1$, $d2$ denotes the current state. Once this time is chosen, the new state is chosen. The probability of making an unseen ''bright-bright'' or ''dark-dark'' transition is given by $\gamma_s/(\gamma_s+k_s)$, and one minus this quantity is the probability of making a visible transition. If a visible transition is observed, the transition time is recorded as part of the single molecule trajectory. After the new state is chosen, the simulation is continued until 300 s has elapsed, but it is possible to incorporate photo-bleaching events. Although we assume that the transitions are sharp in this paper, we can easily simulate systems with broader transition regions by adding the uncertainty in the transition time.

b. Fluctuating bottleneck model

A similar simulation method is used to generate the data for the fluctuating bottleneck model. The equilibrium distribution is given by

$$(\rho_{\mathrm{eq}})_{\pm}=\frac{k_{\mp}}{k_{\pm}+k_{\mp}}\frac{1}{\sqrt{2\pi k/D}}e^{-(k/2D)x^2}, \tag{A1}$$

This distribution can be easily sampled. Given the current position x, we choose a small time step $\delta t\approx D\times 10^{-6}$. We calculate the probability that the system reacts $1-e^{-k_{\pm}x^2\delta t}$. If the system reacts we record the time of the event. Then we propagate the system under normal Brownian motion in a harmonic oscillator to choose a new x value, so that our temporal resolution in $\approx 10^{-6}D$, which is much smaller that the kinetic rates $\sim 4.5D$.

2. The Monte Carlo algorithm

Given the data generated in part 1a of the Appendix, we attempt to determine the relative likelihood of various possible parameters through a standard Metropolis Monte Carlo algorithm. More complex algorithms may be necessary, depending on the amount and complexity of the data, as well as the model being considered. The probability of a specific sequence is easily written in matrix notation as

$$P(D|M)=1^T\left[\prod_i \mathbf{K}e^{(\mathbf{K}_d+\Gamma)t_i}dt_i\right]\delta_{\pm}\rho_{eq}, \tag{A2}$$

where the definitions of the matrices follows the previous paper and the $\delta_{\pm}$ selects the initial condition.[1] dt_i are the small widths of the time bins, which will generally be determined by instrument concerns as well as the photon emission rates from the bright and dark states as discussed in the companion paper.[1] If dt_i is large relative to the kinetic constants, the appropriate binning needs to be considered. Note that the elements of $e^{(\mathbf{K}_d+\Gamma)t_i}$ must be evaluated through standard eigenmode analysis. Since we care about relative probabilities of different kinetic matrices, we can neglect the bin sizes dt_i. For a long sequence the matrix products quickly go to zero. To prevent difficulties associated with zeroing out of the matrix products after each multiplication of $\mathbf{K}e^{(\mathbf{K}_d+\Gamma)t_i}$

we renormalize the resulting vector, $\rho_i = \eta_i \mathbf{K} e^{(\mathbf{K}_d+\Gamma)t_i}\rho_{i-1}$. The constant η_i is chosen so that $\Sigma_j(\rho_i)j=1$. Here i denotes the number of matrix multiplications. The sum of the logs of all of the renormalization constants is the Bayesian score, $B=-\Sigma \ln(\eta_i)$. Except for a constant correction of $\Sigma \ln(dt_i)$, the Bayesian score is the log of $P(D|M)$ which is necessary to apply Bayes formula. The Bayesian score plays the same role as the Boltzmann factor, $\beta E(\{t_i\})$ in statistical mechanics.

The Bayesian score allows us to perform Monte Carlo importance sampling to sample the posterior distribution of parameters that represent the data. We start the simulation at a random parameterization. At iteration j we have a set of parameters with Bayesian score B_j. We sample a new set of parameters $j+1$ by multiplying the current parameters by e^r, where r is a uniform random number symmetric around the origin with step size Δ, i.e., $-\Delta<r<\Delta$. This step corresponds to a Markov Chain Monte Carlo importance sampling with an improper log-uniform *a priori* distribution, $P(M)dM \propto 1/M dM$. This distribution is improper since it is not normalizable, but as can be seen from the numerical examples in this paper, the prior does not play a large role in the final distribution for these models. We choose this importance sampling since the kinetic constants must be positive and we cannot *a priori* set their magnitude so a log-uniform distribution (Jeffrey's prior) is the most appropriate choice.[3] If we are imposing an ordering to preserve uniqueness, i.e., $k_{b1}>k_{b2}$, then after we choose the new values, we relabel the states to maintain this ordering. Once the new parameters $j+1$ are chosen we calculate the Bayesian score for these parameters. Following standard importance sampling in Monte Carlo if $B_{j+1}>B_j$ we accept the new set of parameters, else we conditionally accept the new parameters with probability $e^{B_{j+1}-B_j}$.

For a short sequence and a small number of states, we can readily evaluate the Bayesian score for every possible choice of parameters. For larger sets of states, more complicated models or data a more intelligent methods of sampling is required, which we will discuss in a paper where we apply Bayesian methods to data from single molecule experiments.[9]

[1] J. B. Witkoskie and J. S. Cao, J. Chem. Phys. **121**, 6361 (2004), preceding paper.
[2] C. P. Robert, T. Ryden, and D. M. Titterington, J. Stat. Soc. B Method **62**, 57 (2000).
[3] S. Assoudou and B. Essebbar, Commun. Stat: Theory Meth. **32**, 2163 (2003).
[4] P. H. Garthwaite, Commun. Stat.-Simul. Comput. **23**, 871 (1994).
[5] H. P. Lu, L. Y. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[6] J. S. Cao, Chem. Phys. Lett. **327**, 38 (2000).
[7] S. L. Yang and J. S. Cao, J. Phys. Chem. B **105**, 6536 (2001).
[8] S. L. Yang and J. S. Cao, J. Phys. Chem. B **117**, 10996 (2002).
[9] J. B. Witkoskie and J. S. Cao (unpublished).
[10] S. C. Kou, X. S. Xie, and J. S. Lui (unpublished).
[11] M. E. A. Hodgson and P. J. Green, Proc. R. Soc. London, Ser. A **455**, 3425 (1999).
[12] F. Qin, A. Auerbach, and F. Sachs, Proc. R. Soc. London, Ser. B **264**, 375 (1997).
[13] T. Higuchi, Ann. Inst. Stat. Math. **43**, 469 (1991).
[14] P. G. Debenedetti and F. H. Stillinger, Nature (London) **410**, 259 (2001).
[15] M. K. Sen and P. L. Stoffa, Geophys. J. Int. **108**, 281 (1992).
[16] F. L. H. Brown, Phys. Rev. Lett. **90**, 028302 (2003).
[17] I. V. Gopich and A. Szabo, J. Chem. Phys. **118**, 454 (2003).
[18] J. Wang and P. Wolynes, Phys. Rev. Lett. **74**, 4317 (1995).
[19] G. D'Agostini, Rep. Prog. Phys. **66**, 1383 (2003).
[20] H. Yang, G. Luo, P. Karnchanaphanurach, T. M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie, Science **302**, 262 (2003).