Direct Determination of Kinetic Rates from Single-Molecule Photon Arrival Trajectories Using Hidden Markov Models

Michael Andrec,* Ronald M. Levy,* and David S. Talaga*

Department of Chemistry and Chemical Biology and BioMaPS Institute, Rutgers, The State University of New Jersey, 610 Taylor Road, Piscataway, New Jersey 08854-8087

Received: May 30, 2003; In Final Form: July 17, 2003

The measurement of fluorescence from single protein molecules has become an important new tool in the study of dynamic processes, allowing for the direct visualization of the motions experienced by individual proteins and macromolecular complexes. The data from such single-molecule experiments are in the form of photon trajectories, consisting of arrival times and wavelength information on individual photons. The analysis of photon trajectories can be difficult, particularly if the motions are occurring at rates comparable to the photon arrival rate or in the presence of noise. In this paper, we introduce the use of hidden Markov models (HMMs) for the analysis of photon trajectory data that operate using the photon data directly, without the need for ensemble averaging of the data as implied by correlation function analysis. Using a simple kinetic model, we examine the relationship between the uncertainty in the estimates of the motional rate and the photon detection rate. Remarkably, we obtain relative uncertainties in the rate constants of as little as 3% even when the interconversion rate is equal to the photon detection rate, and the uncertainty increases to only 10% when the interconversion rate is 10 times the photon detection rate. This suggests that useful information can be obtained for much faster kinetic regimes than have typically been studied. We also examine the impact of background photons on the determination of the rate and demonstrate that the HMM-based approach is robust, displaying small uncertainties for background photon arrival rates approaching that of the signal. These results not only are relevant in establishing the theoretical limits on precision, but are also useful in the context of experimental design. Finally, to demonstrate how the methodology can be extended to more complex kinetic models and how it can allow one to make use of the full power of statistics for purposes of model evaluation and selection, we consider a four-state kinetic model for protein conformational transitions previously studied by Schenter et al. (*J. Phys. Chem. A* **1999**, *103*, 10477). We show how an HMM can be used as an alternative to higher-order correlation function analysis for the detection of "conformational memory" and apparent non-Markovian dynamics arising from such temporally inhomogeneous kinetic schemes.

Introduction

Single-molecule measurements are providing insight into many phenomena that were previously intractable because of the ensemble averaging present in bulk measurements.[1−8] In particular, the dynamics of conformationally heterogeneous systems are benefiting from single-molecule studies. Protein folding and conformational dynamics,[9−15] enzymology,[16−20] ribozyme function,[21] bacterial light harvesting,[12,22,23] and protein−nucleic acid interactions[24] are just a few examples of complex systems that have benefitted from the application of single-molecule techniques.

One goal of single-molecule measurements has been to extract the rate of a dynamic process from a single-molecule trajectory. Single-molecule experiments have been used to obtain kinetic rate information about a variety of biological processes, including protein conformational dynamics and folding, enzymatic turnovers, RNA and DNA conformational changes, and fluctuations and function of large biological assemblies. The rates that have been determined have ranged from 1000 to 0.01 Hz. Indirect evidence of the presence of faster processes is common from heterogeneity detected in distributions determined from binned data. However, experimental and data analysis limitations have prevented their quantification.

Comparisons of rates between single molecules can show evidence of heterogeneity or conformational memory.[12,13,19,25] In the context of a molecular system, conformational memory or intermittency[26] results from transitions between unobservable states that modulate the dynamics of the observable states and can result in apparent dynamics of the observable states being non-Markovian, even if the underlying dynamics involving the observable and unobservable states is Markovian. Rigler and co-workers have reported non-Markovian dynamics and molecule-to-molecule differences in activity in the rate of single enzymatic turnovers. They characterized the dynamics in terms of a non-Markovian function that is sensitive to memory in the trajectory.[19] Dovichi and co-workers reported that differences in single alkaline phosphatase catalytic activity result from differing degrees of glycosylation or protease degradation using the total intensity of a fluorescence product turned over during a set incubation time.[27] Such conclusions require that rates be extracted from observations of single molecules and that reliable uncertainty estimates can be made; otherwise, the heterogeneity between molecules cannot be determined with confidence.

* Corresponding authors. E-mail: andrec@lutece.rutgers.edu; ronlevy@lutece.rutgers.edu; talaga@rutchem.rutgers.edu.

10.1021/jp035514+ CCC: $25.00
Published on Web 09/03/2003

The capability of routinely making single-molecule measurements has driven the need for new methods of analyzing single-molecule data that take full advantage of the new and increased information they provide.[28] Essentially three approaches have been used to quantify the rates of single-molecule fluctuations: the fitting of dwell-time histograms, the analysis of the dependence of the shape of a distribution on binning time, and the calculation of correlation functions. Of these methods, only higher-order correlation functions seem to be likely to fully utilize the information present in single-molecule measurements.[18,19,25,29]

The most commonly employed of these three methods is the fitting of dwell-time histograms. When a single-molecule trajectory has sufficient contrast between states, thresholds can be applied to distinguish the states of the molecule. These thresholds are typically chosen manually and can introduce subjectivity into the analysis. Runs of each state are tallied to give histograms of the state dwell times, allowing for the determination of kinetic parameters by exponential fitting. Typically, this technique is limited to systems showing large modulations of the fluorescence signal. Binning of the data is also required, and this limits the temporal resolution of the measurement to be 1 or 2 orders of magnitude lower than the photon count rate to overcome the effects of shot noise. To mitigate the effects of shot noise, some investigators have applied filters to the data prior to applying a threshold.[11] This can substantially improve the time resolution of the experiment by mitigating some of the effects of shot noise, but there is still the difficulty associated with choosing a threshold.

Distribution narrowing has been used to estimate rates in cases where clear assignment of states is not possible. If the data are acquired at sufficiently high temporal resolution, they can be "rebinned" at a lower resolution, effectively averaging over some of the conformational fluctuations by causing exchange between different portions of a distribution. The bin-width dependence of a distribution can allow the time of interchange to be estimated by analogy with motional narrowing of spectral features in wavelength-resolved bulk spectroscopic measurements. This rebinning technique has been demonstrated for the conformational fluctuations in polypeptides and proteins[9,12,14] and appears to be useful for making estimates of interchange times when clear contrast between interchanging states does not exist and adequate trajectories are not available to determine correlation functions.

Correlation analysis is also commonly used and can provide a great deal of information regarding the time scales of fluctuations in the system. Correlation functions can be formally defined in terms of integrals over time with infinite limits. Conceptually, this corresponds to replacing the bulk ensemble average with a single-molecule time average, but such an approach can lead to difficulties for time scales that are not at least an order of magnitude faster than the average total observation time of a single molecule. A single molecule does not typically sample enough of its fluctuation spectrum during a single measurement to allow robust correlation analysis. In practice, a large number of trajectories must be averaged to obtain adequate mathematical accuracy,[11,19,30,31] particularly for the higher-order correlation functions that are sensitive to memory effects and temporal heterogeneity.[18,19,25,29] This prevents the examination of differences between single molecules. Finally, it can be difficult to determine the degree to which a model successfully describes the data using correlation functions.

One type of single-molecule experiment involves the observation of fluorescence fluctuations from an individual member of an equilibrium ensemble. Confocal microscopy coupled with high-sensitivity detection for time-correlated single-photon counting can monitor changes in fluorescence polarization, spectrum, lifetime, and intensity that arise from fluctuations in the system.[32] Single-molecule fluorescence measurements have some important fundamental limitations that restrict the rate, amount, and quality of information obtainable from the system. Because individual single fluorophores can emit only one photon at a time, they exhibit fluorescence anti-bunching at time scales very short compared to the fluorescence lifetime, thus limiting the maximum average observable count rate. Organic dyes are typically used as labels and always have a finite cross section for photobleaching. This limits the total number of photons that can be observed on average from a single molecule. Furthermore, in solution, there will be contributions from spontaneous Raman scattering of the solvent. Even though Raman scattering is weak, the high concentration of the solvent relative to a single molecule makes it a significant source of background in single-molecule fluorescence measurements. Background photons are uncorrelated with the state of the system and therefore degrade the average information content of the photon stream.

Converting the stream of detected photons, or photon arrival trajectory, into knowledge regarding the unobservable and dynamically changing state of the molecule is the goal of single-molecule data analysis and the topic of this paper. In this paper, we present a novel application of a statistical analysis method for extracting information about dynamic processes from single-molecule photon arrival trajectories. We specifically address the problem of extracting the rate of conversion between states and the number of states involved. We include treatment of the statistical uncertainties present in this type of single-molecule measurement and analysis to allow the determination of the significance of any differences observed between molecules. Our method allows us to demonstrate the fundamental limits of precision for determining this dynamic information by applying it to simulated data and to determine the degree to which experimental limitations due to background and detector crosstalk further limit the determination of dynamic information. Quantification of the fundamental precision limits of parameters derived from single-molecule trajectories has important ramifications for experimental design and interpretation.

We have in mind a single-molecule fluorescence measurement in which the molecular dynamics of interest will result in the signal switching from one detection channel to the other. We call this the "two-color problem". Many single-molecule phenomena can be interpreted within this context, including spectral diffusion, fluorescence anisotropy, and FRET colocalization. Single-molecule measurements are limited in precision because of the finite number of kinetic transitions in the observation period. We show how to calculate this "kinetic shot noise" limit. The arrival of photons is stochastic and occurs at a finite rate. We quantify the degree to which these characteristics limit the fastest time scales that can be accurately measured. The knowledge of such limits is critical in experimental design, as it allows for the estimation of the lowest possible intensity that will still permit measurement of the fastest time scale of interest, which is important in minimizing the effect of photobleaching. We show how background and crosstalk between detector channels degrades the accuracy of the rates calculated. We show that our data analysis methods give substantial improvements in the time scales that can be measured for single-molecule photon arrival trajectories.

The utility of maximum-likelihood methods for analysis of single-molecule experiments has been previously noted.[33−36] For

example, single-molecule energy transfer distribution measurements often report energy transfer yields that are negative or greater than 1. This unphysical result has been attributed to the broadening of the distribution due to finite statistical sampling (i.e., shot noise). This is a result of directly calculating molecular properties as one would from bulk measurements. A likelihood-based approach would not give such unphysical results, because the likelihood function includes information regarding the physical process generating the signal. As a result, the most likely parameters that give rise to the observed signal can be determined and will not include artifacts from shot-noise broadening.

The methods we describe operate directly on the photon arrival trajectory of a single molecule by evaluating a likelihood function without the need for averaging over many molecules such as is required for correlation functions. The likelihood function is defined by the solutions of the master equation for the kinetic process of interest and incorporates [by means of a hidden Markov model (HMM)] the corruption of the molecular state information due to backround photons and spectral crosstalk. The HMM formalism also allows us to directly model the effects of temporal heterogeneity, which can be considered to be the corruption of the molecular state information by the spectroscopic degeneracy of multiple molecular states. We demonstrate that this likelihood-based approach yields unbiased estimates of the molecular interconversion rate from the raw data stream with little or no user "tweaking" of the algorithm and that the uncertainty in the estimate of that rate remains low even when the interconversion rate reaches or exceeds the photon detection rate. We also show that the HMM approach is remarkably robust with respect to degradation of the signal by background and crosstalk photons. These results not only confirm the utility of the methodology, but also are useful in experimental design. We demonstrate how HMM-based methods, together with statistical model selection, can be used as an alternative to higher-order correlation function analysis for the detection of intermittency or temporal heterogeneity with a simple example involving a kinetic model previously studied by Schenter et al.[18] in the context of non-Markovian fluctuations of enzymatic reaction rates.

Theory and Computational Methods

The State Trajectory. For clarity of presentation in this paper, we primarily consider two-state kinetic models, i.e., a molecule that can exist in one of two discrete states, A and B, that interconvert with Poisson kinetics at rates k_1 and k_2

$$\mathrm{A} \underset{k_1}{\overset{k_2}{\rightleftharpoons}} \mathrm{B} \tag{1}$$

Suppose that we are given complete knowledge of the state of a molecule at all times t ($0 \geq t \geq T$), which we will call the "state trajectory" of the molecule. Although this state trajectory cannot generally be observed by single-molecule methods, it is of interest as a theoretical limit corresponding to "complete information" about the system. Because the time spent in a given state for a Poisson process is exponentially distributed,[37,38] given k_1 and k_2, we can easily calculate the likelihood for any state trajectory D

$$\begin{aligned} P(D|k_1,k_2) &= \left(\prod_{i=1}^{N_\mathrm{A}} k_1 \mathrm{e}^{-k_1 t_i^{(\mathrm{A})}}\right)\left(\prod_{i=1}^{N_\mathrm{B}} k_2 \mathrm{e}^{-k_2 t_i^{(\mathrm{B})}}\right) \\ &= \left(k_1^{N_\mathrm{A}} \mathrm{e}^{-k_1 T_\mathrm{A}}\right)\left(k_2^{N_\mathrm{B}} \mathrm{e}^{-k_2 T_\mathrm{B}}\right) \end{aligned} \tag{2}$$

where N_A and N_B are the numbers of residences in states A and B, respectively; $t_i^{(\mathrm{A})}$ and $t_i^{(\mathrm{B})}$ are the lengths of the ith residences in states A and B, respectively; and T_A and $T_\mathrm{B} = T - T_\mathrm{A}$ are the total amounts of time spent in states A and B, respectively. The maximum-likelihood estimates of the rates k_1 and k_2 are those that maximize $P(D|k_1,k_2)$ and are given by $N_\mathrm{A}/T_\mathrm{A}$ and $N_\mathrm{B}/T_\mathrm{B}$, respectively. Alternatively, one can adopt a Bayesian approach, which allows one to estimate not only the optimal values of k_1 and k_2, but also their uncertainties, in a conceptually straightforward manner.[39] In this formulation, the joint posterior probability of k_1 and k_2 given the data, $P(k_1,k_2|D)$, can be evaluated using Bayes' theorem

$$P(k_1,k_2|D) \propto P(D|k_1,k_2)\, P(k_1,k_2) \tag{3}$$

where $P(D|k_1,k_2)$ is the likelihood of the data as above and $P(k_1,k_2)$ is a prior probability over the rates. If we consider a uniform prior, i.e., one that is a nonzero constant for positive values of k_1 and k_2 and zero otherwise, then $P(k_1,k_2|D)$ is equal to the likelihood function to within a normalization factor (for positive k_1 and k_2). The shape of the $P(k_1,k_2|D)$ surface as a function of k_1 and k_2 constitutes a representation of our knowledge of k_1 and k_2 given the data. The normalization of eq 2 with respect to k_1 and k_2 is straightforward, and the posterior probability is given by a product of gamma distributions

$$P(k_1,k_2|D) = \left(\frac{T_\mathrm{A}^{N_\mathrm{A}+1}}{\Gamma(N_\mathrm{A}+1)} k_1^{N_\mathrm{A}} \mathrm{e}^{-T_\mathrm{A}k_1}\right)\left(\frac{T_\mathrm{B}^{N_\mathrm{B}+1}}{\Gamma(N_\mathrm{B}+1)} k_2^{N_\mathrm{B}} \mathrm{e}^{-T_\mathrm{B}k_2}\right) \tag{4}$$

The mean and variance of k_1 are given by $(N_\mathrm{A}+1)/T_\mathrm{A}$ and $(N_\mathrm{A}+1)/T_\mathrm{A}^2$, respectively, and similarly for k_2 (with A replaced by B).[40] Therefore, the estimates of k_1 and k_2 in the complete information limit are statistically independent and depend only on the total residence time and number of residences in each state.

The Noiseless Photon Trajectory and Markov Processes. As mentioned above, one cannot, in general, observe the state trajectory directly. At best, one can know the state of the molecule at a finite number of times corresponding to the arrivals of fluorescence photons at the detector (which, for example, can be determined by a Poisson process unrelated to the A $\leftrightarrow$ B interconversion). We will call such a finitely sampled state trajectory a "noiseless photon trajectory". If the underlying kinetics in eq 1 is Poisson, then the sequence of states in a noiseless photon trajectory is a Markov chain corresponding to the well-known "random telegraph process".[38] The transition probabilities of the random telegraph process are determined by the rates k_1 and k_2 and the time between photon arrivals Δt, and they are expressed by the conditional probabilities (solutions of the master equation) for being in state S_i at a time Δt after being in S_j, $P(S_i|S_j,k_1,k_2,\Delta t)$

$$\begin{aligned} P(\mathrm{A}|\mathrm{A},k_1,k_2,\Delta t) &= \frac{k_2}{k_1+k_2} + \frac{k_1}{k_1+k_2}\mathrm{e}^{-(k_1+k_2)\Delta t} \\ P(\mathrm{A}|\mathrm{B},k_1,k_2,\Delta t) &= \frac{k_2}{k_1+k_2}[1-\mathrm{e}^{-(k_1+k_2)\Delta t}] \\ P(\mathrm{B}|\mathrm{A},k_1,k_2,\Delta t) &= \frac{k_1}{k_1+k_2}[1-\mathrm{e}^{-(k_1+k_2)\Delta t}] \\ P(\mathrm{B}|\mathrm{B},k_1,k_2,\Delta t) &= \frac{k_1}{k_1+k_2} + \frac{k_2}{k_1+k_2}\mathrm{e}^{-(k_1+k_2)\Delta t} \end{aligned} \tag{5}$$

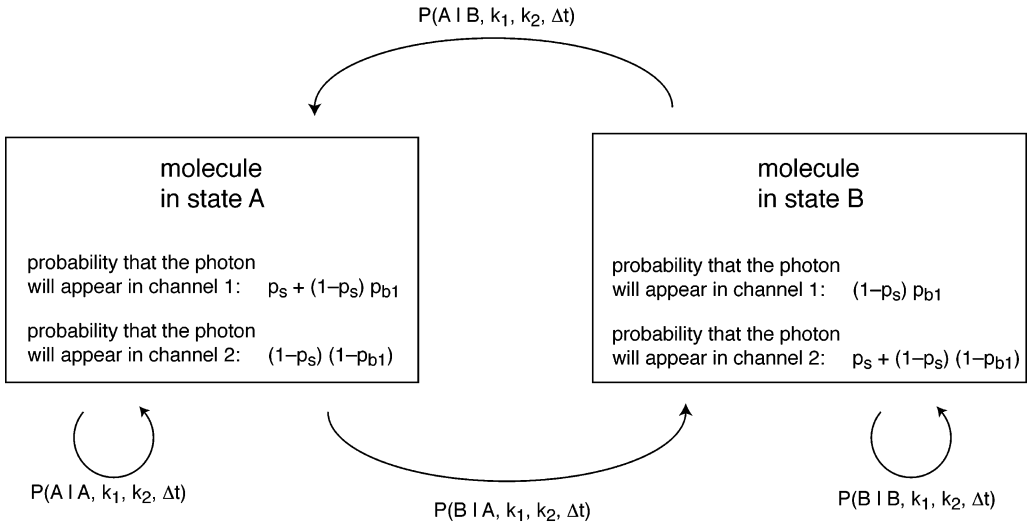


Figure 1. Schematic diagram of the hidden Markov model used to analyze single-molecule photon trajectory data using a two-state exchange model (eq 1) that allows for the presence of background photon noise. Each box represents the state of the molecule at the time of a photon detection event. The model can "transition" from box to box (possibly returning to the box it came from) depending on the transition probabilities $P(S_i|S_j,k_1,k_2,\Delta t)$ as given by eq 5, where Δt is the time elapsed since the previous photon detection event. Each transition corresponds to the detection of a photon. The photon can arrive in either channel 1 or channel 2, with relative probabilities given in each box (known as the "emission probabilities"). A formally identical HMM can be used to model spectral crosstalk and the combination of background scatter and crosstalk; however, the meanings of the emission probabilities would then be given by eqs 12 and 13, respectively.

Because the transition probabilities depend on the time between photon arrivals Δt and because those times are themselves a random variable, the transition probabilities are not constant (as they are in standard Markov chains). However, this is not a problem, because for any given data set, all of the Δt values are known, and therefore, all of the transition probabilities can be calculated as functions of k_1 and k_2 using eq 5.

The posterior distribution of k_1 and k_2 given a noiseless photon trajectory D is no longer given by gamma distributions, as we do not, in general, have any knowledge about the state of the molecule at times other than photon arrivals. However, we can make use of the Markovian nature of the process, which allows us to write the overall likelihood of the data as the product of conditional probabilities

$$P(D|k_1,k_2) = P(D_1|k_1,k_2)\prod_{i=1}^{N-1} P(D_{i+1}|D_i,k_1,k_2,\Delta t_i) \qquad (6)$$

where D is the data vector (D_1, ..., D_{N-1}, D_N) corresponding to the state of the molecule (A or B) at each of the photon arrival times and Δt_i is the time elapsed between the ith and $(i+1)$st photon. The first factor in eq 6 is simply the equilibrium probability of state i (A or B), and the following conditional probabilities are given by eq 5. Although this no longer gives a simple closed-form solution as in eq 2, the maximum-likelihood estimates of the rates k_1 and k_2 can be found by nonlinear optimization, while the posterior probability can be normalized using numerical integration, and visualized using a two-dimensional contour plot. However, we have chosen to show results below only for the one-dimensional case (i.e., assuming $k = k_1 = k_2$) to simplify the graphical presentation.

The Noisy Photon Trajectory and Hidden Markov Models. The preceding analysis assumed that, for each photon arrival, we have perfect knowledge of the state of the molecule at that time. As discussed in the Introduction, this is not normally the case for experimental single-molecule fluorescence data, which routinely contain noise resulting from Raman scattering background and spectral crosstalk. For example, consider the case of data that are ideal except for the presence of background scattering. Background photon data are completely uncorrelated with the state of the molecule and can be described completely by the probability of the arrival of a background photon in channel 1, which we denote as p_{b1}. The probability that a background photon will arrive in channel 2 is simply $1 - p_{b1}$. Given that the molecule is in state A, the probability, p_1, that a photon will arrive in channel 1 (as opposed to channel 2) is given by

$$P(p_1|\mathrm{A}) = p_s + (1 - p_s)p_{b1} \qquad (7)$$

where p_s is the probability that any given photon is a signal photon. The corresponding probability for channel 2 is

$$P(p_2|\mathrm{A}) = 1 - P(p_1|\mathrm{A}) = (1 - p_s)(1 - p_{b1}) \qquad (8)$$

while the analogous results for molecular state B are

$$P(p_1|\mathrm{B}) = (1 - p_s)p_{b1} \qquad (9)$$

and

$$P(p_2|\mathrm{B}) = p_s + (1 - p_s)(1 - p_{b1}) \qquad (10)$$

Therefore, the observed data (the arrival channel of each photon) can be thought of as a probabilistic function of an unobserved Markov process (the state of the molecule), which is the classic definition of a hidden Markov model (HMM) (Figure 1).[41]

HMMs have been extensively used in fields as diverse as speech recognition, bioinformatics, neuroscience, climatology, and finance, and the associated methodology has developed rapidly.[42−44] In HMM language, the states of the unobserved Markov chain are called the "hidden states" (in our case, A and B), and each hidden state has associated with it an "emission probability" for each observable (eqs 7−10 in our case). In contrast to standard HMM approaches, the transition probabilities between hidden states are not constant, but depend on the time elapsed since the previous photon arrival, and they are given by the random telegraph process master equation solutions

(eq 5). Given values for the emission and transition probabilities and full knowledge of the hidden states, one could easily calculate the likelihood of a noisy photon trajectory by simply modifying eq 6 to include emission probabilities. However, the hidden state sequence is unknown, making the evaluation of the likelihood more difficult. On the surface, calculation of the likelihood would involve optimizing or averaging over all possible hidden state trajectories, which is computationally impossible for reasonably sized data sets. However, the special structure of the HMM allows one to perform the optimization or averaging in a recursive manner, thereby avoiding the combinatorial explosion implied by an explicit enumeration of all possible hidden state sequences.[42−44] In our case, the recursion can be constructed by noting that the joint probability $f_t(\mathrm{A}) = P(\omega_1,\omega_2,...,\omega_t,h_t{=}\mathrm{A})$ that the arrival channel sequence is (ω_1, ω_2, ..., ω_t) and that the molecule is in hidden state h_t = A at time t can be written as

$$f_{t+1}(\mathrm{A}) = P(\omega_{1+1}|\mathrm{A})[P(\mathrm{A}|\mathrm{A},\Delta t)\, f_t(\mathrm{A}) + P(\mathrm{A}|\mathrm{B},\Delta t)\, f_t(\mathrm{B})] \tag{11}$$

where the emission probabilities $P(\omega_1|\mathrm{A})$ are given by eqs 7 and 8 and the transition probabilities are given by eq 5. An analogous recursion can be written for $f_{t+1}(\mathrm{B})$ in terms of $f_t(\mathrm{A})$ and $f_t(\mathrm{B})$. The recursion is initiated using $f_1(\mathrm{A}) = P(\omega_1,h_1{=}\mathrm{A}) = P(\omega_1|\mathrm{A})\, P(\mathrm{A})$, where $P(\mathrm{A})$ is the equilibrium probability of the molecule being in state A, and similarly for $f_1(\mathrm{B})$. The final overall averaged likelihood is simply $f_N(\mathrm{A}) + f_N(\mathrm{B})$, where N is the number of photons. Equation 11 is known as the "forward" recurrence relation. An equivalent "backward" recurrence can also be formulated by beginning at the final data point and proceeding in the opposite direction using the conditional probabilities $b_t(\mathrm{A}) = P(\omega_{t+1},\omega_{t+2},...,\omega_N|h_t{=}\mathrm{A})$ and $b_t(\mathrm{B}) = P(\omega_{t+1},\omega_{t+2},...,\omega_N|h_t{=}\mathrm{B})$.[42−44] Traditionally, an expectation-maximization procedure known as the Baum−Welch algorithm has been used to estimate unknown HMM parameters;[42,43] however, in principle, any nonlinear optimization algorithm can be used.[41] In our case, we have estimated the properties of the posterior probability of k_1, k_2, and the background scattering parameters by Monte Carlo sampling (see below).

The effect of spectral crosstalk can be modeled using exactly the same HMM as for background photons. However, the emission probabilities are now given by

$$\begin{aligned} P(p_1|\mathrm{A}) &= 1 - p_{\mathrm{xA2}} \\ P(p_2|\mathrm{A}) &= p_{\mathrm{xA2}} \\ P(p_1|\mathrm{B}) &= p_{\mathrm{xB1}} \\ P(p_2|\mathrm{B}) &= 1 - p_{\mathrm{xB1}} \end{aligned} \tag{12}$$

where p_{xA2} and p_{xB1} are the leakage probabilities of state A into channel 2 and state B into channel 1, respectively. The emission probabilities for modeling background photons and crosstalk simultaneously are simply the combination of eqs 7−10 and eq 12

$$\begin{aligned} P(p_1|\mathrm{A}) &= p_{\mathrm{s}}(1 - p_{\mathrm{xA2}}) + (1 - p_{\mathrm{s}})p_{\mathrm{b1}} \\ P(p_2|\mathrm{A}) &= p_{\mathrm{s}}p_{\mathrm{xA2}} + (1 - p_{\mathrm{s}})(1 - p_{\mathrm{b1}}) \\ P(p_1|\mathrm{B}) &= p_{\mathrm{s}}p_{\mathrm{xB1}} + (1 - p_{\mathrm{s}})p_{\mathrm{b1}} \\ P(p_2|\mathrm{B}) &= p_{\mathrm{s}}(1 - p_{\mathrm{xB1}}) + (1 - p_{\mathrm{s}})(1 - p_{\mathrm{b1}}) \end{aligned} \tag{13}$$

Although eq 13 shows that a straightforward two-state HMM can be used to model the simultaneous presence of background and crosstalk, it should be noted that, without prior information, one cannot use the HMM parameters to estimate the relative amounts of background and crosstalk, as both processes contribute to the model via the emission probabilities $P(p_1|\mathrm{A})$ and $P(p_1|\mathrm{B})$ in a way that prevents their simultaneous estimation. However, a lack of knowledge of the relative amounts of background and crosstalk does not prevent the use of the HMM for the determination of the rates of the kinetic process.

Computational Methods. The analyses described in this paper were performed using synthetic data. Data were generated by first constructing a state trajectory for the molecule based on the kinetic scheme of eq 1. This was done by choosing an initial state based on the equilibrium probabilities and generating waiting times distributed according to an exponential distribution with rate k_1 if the molecule is in state A and k_2 if the molecule is in state B.[38] A set of photon detection times were then generated with interphoton times exponentially distributed with a photon detection rate k_{p}. For each photon detection event, the state trajectory can be used to determine the channel in which that photon would, under ideal circumstances, be detected (i.e., channel 1 for state A and channel 2 for state B). The resulting list of photon arrival times and channels represents a realization of a noiseless photon trajectory consistent with the rates k_1, k_2, and k_{p}.

The effect of background scattering and spectral crosstalk were simulated by modifying the photon arrival channels of the above noiseless photon trajectories. To simulate background scattering, each photon in the noiseless photon trajectory was replaced by a background photon with probability $1 - p_{\mathrm{s}}$ and was assigned to channel 1 or 2 with probabilities p_{b1} and $1 - p_{\mathrm{b1}}$, respectively. Similarly, spectral crosstalk was simulated by changing each channel 1 photon in the noiseless trajectory to channel 2 with probability p_{xA2} and each channel 2 photon to channel 1 with probability p_{xB1}. This procedure was also performed on all signal photons in the background-corrupted trajectory to generate a trajectory containing both crosstalk and background.

All parameter estimation was done from a Bayesian perspective under a uniform prior in closed form for state trajectory data, by direct evaluation of the posterior probability for noiseless photon trajectory data or by means of Metropolis Monte Carlo sampling over the HMM parameters[45,46] for models containing noise. For all of the analyses presented here, the noise levels (i.e., the emission probabilities) were considered to be unknown and were treated as adjustable parameters in the model. The Monte Carlo sampling was based on the joint posterior probability of the HMM model parameters calculated directly using the forward recursion with rescaling to prevent numerical underflow.[41,44,47] The result of a Monte Carlo simulation is a set of points in the parameter space that are distributed according to their posterior probabilities, from which it is straightforward to estimate moments (e.g., means and standard deviations) and other summary statistics. The Metropolis proposals were generated from a multivariate normal density with means and covariance matrix chosen in an iterative manner as described in ref 48 to maintain a rejection rate of approximately 40%.

Results and Discussion

The Complete Information Limit. We began by generating 100-ms-long synthetic state trajectories as described above using exchange rates $k_1 = k_2 = k$ of 0.1, 0.5, 1, 5, 10, 50, 100, 500, 1000, 5000, and 10 000 ms^{-1}. Although such trajectories are

TABLE 1: Results for Likelihood-Based Analyses of Synthetic Single-Molecule Trajectory Data

	state trajectory[a]		noiseless photon trajectory[b]		background[c]		crosstalk[d]		background + crosstalk[e]	
true k (ms^{-1})	estimate (ms^{-1})[f]	rel. uncert. (%)[g]	estimate (ms^{-1})[f]	rel. uncert. (%)[g]	estimate (ms^{-1})[f]	rel. uncert. (%)[g]	estimate (ms^{-1})[f]	rel. uncert. (%)[g]	estimate (ms^{-1})[f]	rel. uncert. (%)[g]
0.1	0.140 ± 0.037	37	0.110 ± 0.033	33	0.111 ± 0.035	35	0.112 ± 0.035	35	0.112 ± 0.036	36
0.5	0.550 ± 0.074	15	0.526 ± 0.073	15	0.539 ± 0.077	15	0.555 ± 0.083	17	0.550 ± 0.086	17
1	1.04 ± 0.10	10	0.99 ± 0.10	10	0.98 ± 0.11	11	0.99 ± 0.11	11	1.02 ± 0.12	12
5	4.91 ± 0.22	4.4	5.06 ± 0.25	5.0	4.89 ± 0.28	5.6	4.97 ± 0.36	7.1	4.77 ± 0.36	7.2
10	9.73 ± 0.31	3.1	9.76 ± 0.38	3.8	9.53 ± 0.46	4.6	9.73 ± 0.61	6.1	9.25 ± 0.66	6.6
50	48.9 ± 0.7	1.4	49.4 ± 1.5	3.1	51.2 ± 2.4	4.9	49.9 ± 3.8	7.6	49.8 ± 4.5	8.9
100	99.2 ± 1.0	1.0	97.3 ± 3.3	3.4	95.8 ± 5.3	5.3	94.7 ± 8.6	8.6	95 ± 11	11
500	499.5 ± 2.2	0.4	502 ± 31	6.1	500 ± 46	9.3	483 ± 88	18	501 ± 88	22
1000	996.2 ± 3.2	0.3	1118 ± 97	9.8	1143 ± 174	17	1297 ± 272	27	1177 ± 370	37
5000	5014.7 ± 7.1	0.14	6715 ± 1625	25	5422 ± 1495	30	$5489 \pm —$[h]	—[h]	$5179 \pm —$[h]	—[h]
10 000	9980 ± 10	0.10	9120 ± 2182	37	$17\,703 \pm 10\,672$	107	$19\,253 \pm —$[h]	—[h]	$36\,122 \pm —$[h]	—[h]

[a] Complete knowledge of the molecular state at all times for an observation period of 100 ms. [b] State trajectory restricted to knowledge of the molecular state at photon arrivals ($k_p = 100$ ms^{-1}). [c] Noiseless photon trajectory with background photons ($p_s = 0.909$, $p_{b1} = 0.5$). [d] Noiseless photon trajectory with crosstalk photons ($p_{xA2} = p_{xB1} = 0.15$). [e] Noiseless photon trajectory with background and crosstalk photons (above parameters). [f] Mean ± standard deviation of marginal posterior probability density of k. [g] (Standard deviation of marginal posterior probability density of k)/(true k) [h] Standard deviation of marginal posterior probability density is undefined, and only the maximum-likelihood estimate of k can be obtained.

not directly experimentally observable, they represent a lower limit on the uncertainty in the estimates of the exchange rate and provide us with the fundamental limit for parameter estimation from a finite-length trajectory. We considered each synthetic state trajectory as if it were experimental data and asked how well we could infer the exchange rate k using eq 4 (under the assumption $k = k_1 = k_2$). As might be expected, the estimated values for k are extremely good; however, the relative uncertainties in those estimates vary over more than 3 orders of magnitude from 37% at $k = 0.1$ ms^{-1} to 0.01% at $k = 10\,000$ ms^{-1} (Table 1). This result follows from the fact that the relative uncertainty decreases as the square root of the number of A ↔ B interconversions observed. Because, for a given finite observation time, we observe more interconversions when k is larger, the relative uncertainty will decrease monotonically as k increases. As discussed above, this relative uncertainty represents a theoretical lower bound, because it assumes complete knowledge of the state of the molecule at all times during the observation period.

Noiseless Photon Trajectories. To investigate the loss of information due to finite sampling of the state trajectory, we generated noiseless photon trajectories based on the above state trajectories using a photon detection rate, k_p, of 100 ms^{-1}. For each trajectory, the logarithm of the likelihood under the assumption of $k_1 = k_2 = k$ was calculated by directly evaluating the sum of the logarithms of the conditional probabilities in eq 6. The mean and standard deviation of k under the posterior distribution were calculated using numerical quadrature and are shown in Table 1 and Figure 2. Because the results depend only on the relative magnitude of k compared to k_p and the total observation time T, the abscissa of Figure 2 is given in dimensionless units of k/k_p.

The dependence of relative uncertainty in the estimate on k is interesting, in that it approaches the theoretical complete information limit (where the uncertainty is defined completely by the relative size of k and the inverse of the observation time) for very small values of k/k_p, but it begins to increase substantially for $k/k_p > 0.1$. This is understandable, because when $k \ll k_p$, there are many photon arrivals for every residence time; therefore, very little loss of information occurs as a result of the finite sampling of the state trajectory. However, as k/k_p approaches and exceeds unity, we lose information about the precise length of each residence. In addition, it becomes

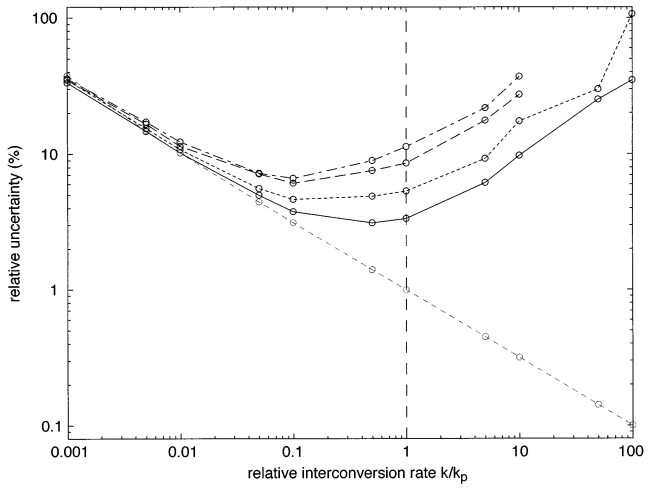


Figure 2. Dependence of the relative uncertainty in the estimation of k as a function of the ratio k/k_p, where k_p is the photon detection rate (data from Table 1). The dashed gray data are for the complete information limit derived from analysis of the state trajectory data (a value of $k_p = 100$ ms^{-1} was used for plotting purposes only) and are completely determined by the relative size of k and the inverse of the total observation time. The solid line corresponds to noiseless photon trajectories generated by sampling the state trajectories with a Poisson rate of $k_p = 100$ ms^{-1}. The remaining lines correspond to the addition of varying amounts of noise to the noiseless photon trajectories: background scatter with $p_s = 0.909$, $p_{b1} = 0.5$ (dotted), crosstalk with $p_{xA2} = p_{xB1} = 0.15$ (dashed), and both (dot−dashed).

increasingly likely that a particular residence will be completely unobservable because no photon arrives during that time. Thus, the relative uncertainty has contributions from two competing effects: the finite number of transitions when k is small and approaches T^{-1} (kinetic shot noise) and the finite sampling of the state trajectory when k is large and approaches k_p. This results in a V-shaped dependence of relative uncertainty on k with a minimum in the neighborhood of $k/k_p = 0.5$ (Figure 2, solid curve).

It is remarkable that one can obtain relative uncertainties of ~3% even when the interconversion rate $k/k_p = 1$ and that the uncertainty increases to only 10% when $k/k_p = 10$. This suggests that useful information can be obtained for much faster kinetic regimes than have typically been studied, and it is a significant benefit arising from measuring the photon arrival trajectory over

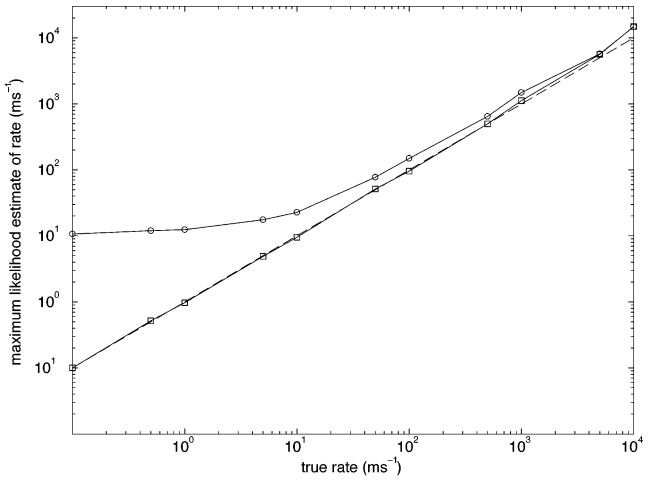


Figure 3. Maximum-likelihood estimate of the rate (i.e., the mode of the posterior under a uniform prior) obtained using a naive application of eq 7 to photon trajectory data containing background scatter with parameters $p_s = 0.909$ and $p_{b1} = 0.5$ (circles) and maximum-likelihood estimates obtained using a hidden Markov model that incorporates noise (squares) as a function of the true rate k. The line of unit slope (corresponding to perfectly unbiased estimation) is shown in dashes.

binned measurements. Because the photons stochastically sample the state trajectory, there are a substantial fraction of photons closer together temporally than the mean interphoton time. This effectively increases the bandwidth of the measurement by an order of magnitude or more over measurements binned at the photon count rate.

In addition to demonstrating the power of the direct likelihood-based analysis of photon trajectory data, Figure 2 can also be used in the context of experimental design. In general, to minimize the effect of photobleaching, one would like to use the lowest possible laser intensity that will still allow measurement of the fastest time scale of interest. If one has an approximate guess for the fastest interconversion process in the system of interest, then for a given relative uncertainty, one can use the curves of Figure 2 to determine the lowest photon detection rate that will allow the measurement of that rate to the desired precision. For example, if one required a precision of a few percent, then one would need to use sufficient laser power to obtain a photon arrival rate at least 5 times the expected interconversion rate (in the absence of noise). On the other hand, if one were willing to tolerate a relative uncertainty approaching 10%, then one could decrease the laser power so that the photon arrival rate was slower than the interconversion rate.

Noisy Photon Trajectories and Hidden Markov Models. The most serious source of nonideality in real photon trajectories is the presence of spectral crosstalk and background photons. This results in a random fraction of the observed photons being anticorrelated or uncorrelated with the molecular state. In other words, some fraction of the observed photons will be "lying", i.e., their arrival channel will be the opposite of the channel that would be expected on the basis of the actual state of the molecule. To investigate the effect of background photons on the naive use of eq 6, we perturbed the noiseless photon trajectories to simulate background photons with $p_s = 0.909$ (i.e., a signal-to-background ratio of ~10:1) and $p_{b1} = 0.5$ (a worst-case scenario of equal background in both channels). The results are shown in Figure 3 (circles). The effect of excluding noise from the model is quite catastrophic when k is small, resulting in misestimates of k by more than a factor of 100 when $k = 0.1$ ms^{-1}, but it becomes less severe as k increases. The very poor performance of eq 6 in the presence of background photons is due to the fact that the model does not include noise, but rather assumes that every photon is "telling the truth". For interconversion rates that are slow relative to the photon arrival rate, this leads to data where the molecule is clearly in a single state for a long period but with one photon occasionally "straying" into the other channel. Such data are extremely unlikely under a simple Poisson model, and use of such a model by the naive application of eq 6 results in grossly inflated estimates of the interconversion rate. At larger values of k, the background photons tend to "blend in" more with the signal, thereby causing a much lower systematic error.

One can incorporate the presence of background photons into the likelihood-based approach using an HMM as discussed above, and this results in estimates that are much closer to the true value than those obtained by the naive application of eq 6 (Figure 3, squares). However, the price paid for this improved accuracy is a decrease in precision, particularly when k is large (Figure 2, dotted line). This is not surprising, given that the HMM results are obtained by implicitly averaging over all 2^N possible hidden variable state sequences. However, when k is small, the relative uncertainty remains near the complete information limit. This is because very few of these 2^N sequences are likely under a Poisson model for exchange, allowing the HMM to determine which photons are lying with a high degree of confidence. Similar results are obtained with the introduction of crosstalk or the combination of background and crosstalk (Table 1 and Figure 2, dashed and dot−dashed lines).

In our analysis, we assumed that the noise photon rate and distribution is completely unknown. However, such knowledge could be obtained experimentally from the pattern of photon arrivals in the absence of sample. One might ask whether and how much such a priori knowledge of the noise properties could improve the estimate of the interconversion rate. This question is relevant from the perspective of experimental design: one would want to know how much effort to put into experimentally characterizing the noise characteristics of the hardware. In general, one will observe an improvement only if there is substantial correlation between the uncertainties in k and the HMM emission probabilities. We have observed that such correlation is practically zero when k is small or when the noise level is low (Figure 4A). Therefore, prior information about the amount of crosstalk and/or background scatter will not improve the estimate of k in this limit. Some weak correlation is observed as k and/or the noise level increases (Figure 4B), suggesting that very careful experimental measurement of the background and crosstalk parameters might give a slight improvement in the estimate of k. The marginal gain in the precision and accuracy appears to be quite small, however, and does not justify the expenditure of significant effort.

To investigate the ability to infer k as a function of the amount of background photons, we generated 13 data sets using $k = 50$ ms^{-1} and signal photon probabilities, p_s, ranging from 0.976 (signal-to-background ratio $S/B \equiv p_s/(1 - p_s) = 40$) to 0.231 ($S/B = 0.3$) using equal background photon probabilities in each channel ($p_{b1} = 0.5$) and repeated the analyses of Figure 2. Those results (along with the results for $p_s = 0.909$ described above) are shown as the solid curve in Figure 5. Remarkably, the relative uncertainty in k remains below 20% even for S/B levels approaching 1. Although the increase in relative uncertainty as a function of background level is approximately exponential, the rate of increase is nonetheless surprisingly slow, indicating that the likelihood-based HMM analysis strategy is quite robust for small or moderate values of k/k_p. We repeated the above analysis using an asymmetric background photon distribution

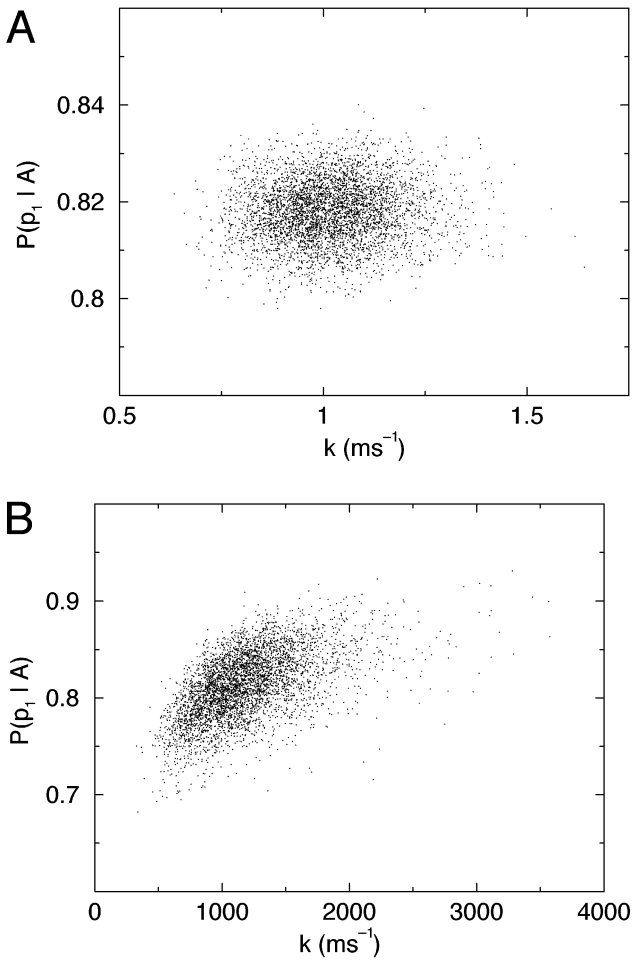


Figure 4. Scatter plot of Metropolis Monte Carlo output for the analysis of photon trajectories containing both background and crosstalk noise ($P(p_1|A) = 0.818\ 15$) generated using k = (A) 1 and (B) 1000 ms^{-1} using an HMM analysis. Each panel corresponds to a projection of the full set of Monte Carlo samples from the posterior probability distribution onto a plane corresponding to the interconversion rate k and the emission probability for state A. The fact that the cloud of points is elongated and tilted from the horizontal in panel B is a graphical indication that there is a weak correlation between the inferred rate and emission probability, whereas the lack of such behavior in panel A indicates the lack of correlation.

($p_{b1} = 0.8$), and the results are shown in the dashed curve of Figure 5. The relative uncertainties are very similar to the $p_{b1} = 0.5$ case, with only a small decrease in uncertainty at very high background levels. Overall, the robustness of the analysis is negligibly affected by the distribution of background photons between channels and depends only on the overall signal-to-background ratio. Because the models for background and crosstalk are mathematically identical, we expect similar results for the effect of asymmetry in crosstalk.

Extension to Models Exhibiting Temporal Heterogeneity. We have shown that likelihood-based methods that make use of hidden Markov models are a robust tool for estimating interconversion rates from single-molecule photon trajectories in the context of two-state jump kinetics. They provide not only reliable parameter estimates, but also statistically rigorous estimates of the uncertainty in those estimates. Another advantage of HMMs in the context of photon trajectory analysis is that they can be easily generalized to more complex non-two-state models, and provide a statistical context in which to evaluate the relative goodness-of-fit of different models. As an

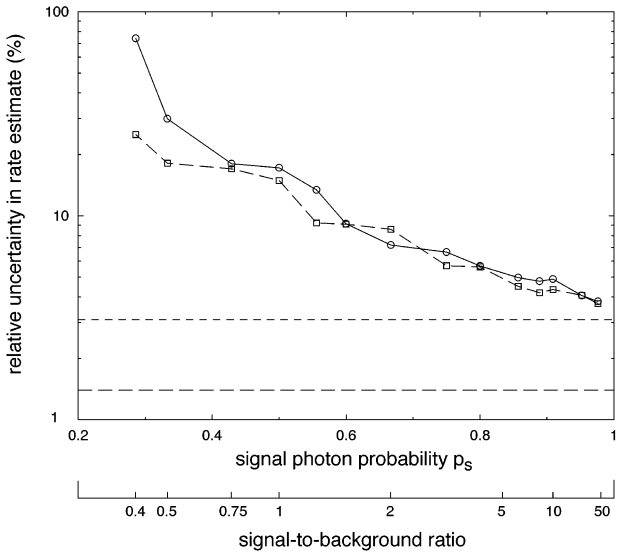
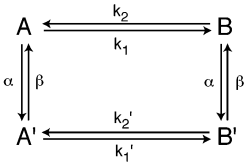


Figure 5. Dependence of the relative uncertainty in the estimation of k as a function of the background photon level for a 100-ms photon trajectory generated using $k = 50\ ms^{-1}$, $k_p = 100\ ms^{-1}$, $p_{b1} = 0.5$ (solid/circles), and $p_{b1} = 0.8$ (dashed/squares). The signal-to-background ratio is defined to be $p_s/(1 - p_s)$. The lower long-dashed horizontal line represents the theoretical lower bound on the relative uncertainty corresponding to the complete information limit, whereas the upper horizontal short-dashed line represents the $p_s = 0$ limit corresponding to sampling the state trajectory with a Poisson rate of $k_p = 100\ ms^{-1}$.

example, consider the following four-state model studied by Schenter and co-workers[18]



(14)

where the states A and A′ (and B and B′) are spectroscopically indistinguishable. This degeneracy results in apparent dynamics that is non-Markovian and is said to exhibit conformational memory or temporal heterogeneity. The memory effect is the result of not including enough states in the description of the dynamics, and expansion of the model from the two spectroscopically distinguishable states to four results in simple Markovian dynamics. Models of this type have been used, for example, to model enzymatic reaction rates that are fluctuating as a result of conformational dynamics in the enzyme.[18] Alternatively, one can also view this kinetic scheme as a barrier-crossing model with two deep wells (corresponding to A/A′ and B/B′) where the well depths and/or barrier heights are fluctuating in a Poissonian manner. In previous work, this model has been fit using an approach based on correlation functions.[18] However, because the ordinary two-point correlation function does not contain sufficient information to reliably distinguish between the four-state model (eq 14) and the simpler two-state model (eq 1), multipoint correlation functions, which are more difficult to interpret and suffer from higher levels of noise, are required.

It is straightforward to formulate an HMM based on the kinetic scheme of eq 14. Let us assume for the moment that there is no background scatter and that the spectral crosstalk is zero, i.e., that all photons from states A and A′ arrive in channel 1 and all photons from states B and B′ arrive in channel 2. If we consider a model with four hidden states (A, A′, B, and B′), then the observed data is a deterministic function of the hidden

states, i.e., we have an HMM where the emission "probabilities" are identically 0 or 1. This is a special case of eq 11 extended to four hidden states. All that is needed are the transition probabilities analogous to those in eq 5. These are given by the elements of the matrix exponential $\exp(\mathbf{K}t)$, where $\mathbf{K}$ is the kinetic matrix

$$\mathbf{K} = \begin{pmatrix} -(k_1+\alpha) & k_2 & \beta & 0 \\ k_1 & -(k_2+\alpha) & 0 & \beta \\ \alpha & 0 & -(k_1'+\beta) & k_2' \\ 0 & \alpha & k_1' & -(k_2'+\beta) \end{pmatrix} \quad (15)$$

Whereas $\exp(\mathbf{K}t)$ can, in principle, be calculated in closed form for a four-state model, the general solution is sufficiently awkward that numerical solution is preferable. However, the symmetric case where $k_1 = k_2 = k$, $k_1' = k_2' = k'$, and $\alpha = \beta$ gives a relatively simple solution, the details of which are given in the Appendix.

For purposes of demonstration, we generated a 100-ms-long synthetic two-channel noiseless photon trajectory corresponding to the kinetic scheme of eq 14 in the symmetric limit with rate constants $k = 20\ \mathrm{ms}^{-1}$, $k' = 80\ \mathrm{ms}^{-1}$, and $\alpha = 10\ \mathrm{ms}^{-1}$ and a photon detection rate of $100\ \mathrm{ms}^{-1}$ using an approach similar to that described in the Computational Methods section above.[38] We analyzed that trajectory using the symmetric four-state HMM with Metropolis Monte Carlo sampling in the same manner as for the two-state model, except that the emission probabilities were taken to be 0 or 1 as appropriate and the transition probabilities were those given in the Appendix. The resulting parameter estimates (posterior mean ± standard deviation) were $k = 16.6 \pm 2.2\ \mathrm{ms}^{-1}$, $k' = 93.2 \pm 6.9\ \mathrm{ms}^{-1}$, and $\alpha = 14.4 \pm 4.8\ \mathrm{ms}^{-1}$. Analysis of the same data using a two-state model based on eq 1 with $k_1 = k_2 = k$ (eq 6) gave a parameter estimate of $k = 44.4 \pm 2.0\ \mathrm{ms}^{-1}$, whereas the HMM model based on eq 1 gave an estimate of $k = 41.7 \pm 1.6\ \mathrm{ms}^{-1}$. To assess the relative statistical merits of the three models, we must correct for the fact that adding parameters to any model will always improve the "fit", but not necessarily the statistical significance. One simple way in which this can be done is via the Bayes information criterion (BIC), which is given by

$$\mathrm{BIC} = -2L + d \ln N \quad (16)$$

where L is the maximized logarithm of the likelihood, d is the number of dimensions in the parameter space, and N is the number of data points.[49] The BIC can be used both to select the best fitting model and to determine the relative probabilities of the models.[49] In our case, the BIC showed that the four-state HMM is very convincingly the best model of the three in a statistical sense,[50] thereby demonstrating the effectiveness of an HMM-based model selection procedure as an alternative to the multipoint correlation function for the detection of temporal hetergeneity. Although we have used a noiseless photon trajectory for this simple example, the effect of noise on a four-state model could be incorporated simply by allowing the emission probabilities to deviate from the deterministic values of 0 and 1.

Conclusions

We have shown how likelihood-based statistical methods can be applied to the analysis of photon trajectories from single-molecule fluorescence experiments. The likelihood-based methods presented here will enhance the ability to extract information about motions in proteins and macromolecular complexes from single-molecule fluorescence experimental data. In particular, they directly make use of the information from individual photons without the need for binning, averaging, or other methods that transform the data before analysis. Although methods based on binning or averaging can make the subsequent analysis more straightforward or familiar, they can obscure the dynamical information, particularly if the dynamics are occurring on time scales comparable to or faster than the binning or averaging. The gain in bandwidth obtained by the direct analysis of the photon trajectory is quite large, suggesting that useful information could be obtained for dynamics occurring at rates more than an order of magnitude faster than the photon detection rate. In addition, the HMM methodology for the analysis of noisy photon trajectories is remarkably robust with respect to the amount of noise, which suggests that it will be of significant practical utility in the analysis of real-world data.

Likelihood-based methods, particularly in their Bayesian form, lead to natural and intuitive estimates of the uncertainties in the estimated parameters arising from both experimental noise and finite sampling. Realistic error estimates are critical in experimental work, for example, in assessing the significance of differences in rate constants under different experimental conditions. Likelihood-based methods provide such error estimates. In addition, these methods can be made to operate in a fully automatic mode or with only minimal analyst intervention. This is critical given the quantity of data that must be analyzed in typical experimental situations.

For reasons of presentational clarity, the models considered in this paper have, for the most part, been limited to two-state exchange processes with equal forward and reverse rates. Likelihood-based methods are certainly not limited to such simple models. For example, unequal forward and reverse rates represent a trivial extension and merely require additional visualization methods. Furthermore, we have shown how more complex models that contain larger numbers of states and exhibit temporal heterogeneity can be addressed. In principle, models of arbitrary complexity, including those with large numbers of states[18,51] or time-dependent rate constants,[52] can be fit to photon trajectory data using this methodology, provided that the conditional probabilities analogous to eq 5 for the dynamic process can be solved analytically or numerically. In addition, the HMM used to analyze noisy photon trajectory data can also be used in unmodified form to analyze data from other types of single-molecule fluorescence experiments, such as those based on resonance energy transfer. Given a choice of possible models, it becomes critical to have reliable means of evaluating the goodness-of-fit of each model and the statistical significance of improvements in fit between different models. Likelihood-based methods allow for such model evaluation and testing in a statistically rigorous manner.

Acknowledgment. This work was supported in part by the National Institutes of Health (GM 30580 and GM 64375) and the Research Corporation through a Research Innovation Award.

Appendix

The conditional probabilities analogous to eq 5 for the four-state kinetic scheme of eq 14 in the symmetric limit $k_1 = k_2 = k$, $k_1' = k_2' = k'$, and $\alpha = \beta$ are the elements of the matrix exponential $\exp(\mathbf{K}t)$, where $\mathbf{K}$ is the kinetic matrix

$$\mathbf{K} = \begin{pmatrix} -(k+\alpha) & k & \alpha & 0 \\ k & -(k+\alpha) & 0 & \alpha \\ \alpha & 0 & -(k'+\alpha) & k' \\ 0 & \alpha & k' & -(k'+\alpha) \end{pmatrix} \quad \text{(A1)}$$

The matrix exponential can be calculated from the eigenvalues λ_i and eigenvectors U_i using

$$\exp(\mathbf{K}t) = \mathbf{U}\mathbf{\Lambda}(t)\mathbf{U}^{-1} \quad \text{(A2)}$$

where **U** is the matrix of eigenvectors of **K** and

$$\mathbf{\Lambda}(t) = \begin{pmatrix} \exp(\lambda_1 t) & 0 & 0 & 0 \\ 0 & \exp(\lambda_2 t) & 0 & 0 \\ 0 & 0 & \exp(\lambda_3 t) & 0 \\ 0 & 0 & 0 & \exp(\lambda_4 t) \end{pmatrix} \quad \text{(A3)}$$

as described in ref 53. The resulting conditional probabilities are

$$P(\mathrm{A}|\mathrm{A},k,k',\alpha,\Delta t) = P(\mathrm{B}|\mathrm{B},k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t} + e^{-\kappa_+\Delta t} + e^{-\kappa_-\Delta t}) + \frac{k-k'}{4\gamma}(e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

$$P(\mathrm{A}'|\mathrm{A}',k,k',\alpha,\Delta t) = P(\mathrm{B}'|\mathrm{B}',k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t} + e^{-\kappa_+\Delta t} + e^{-\kappa_-\Delta t}) - \frac{k-k'}{4\gamma}(e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

$$P(\mathrm{A}|\mathrm{B},k,k',\alpha,\Delta t) = P(\mathrm{B}|\mathrm{A},k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t} - e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t}) - \frac{k-k'}{4\gamma}(e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

$$P(\mathrm{A}'|\mathrm{B}',k,k',\alpha,\Delta t) = P(\mathrm{B}'|\mathrm{A}',k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t} - e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t}) + \frac{k-k'}{4\gamma}(e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

$$P(\mathrm{A}'|\mathrm{A},k,k',\alpha,\Delta t) = P(\mathrm{B}'|\mathrm{B},k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t}) - \frac{\alpha}{4\gamma}(e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

$$P(\mathrm{B}'|\mathrm{A},k,k',\alpha,\Delta t) = P(\mathrm{A}'|\mathrm{B},k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t}) + \frac{\alpha}{4\gamma}(e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

$$P(\mathrm{A}|\mathrm{B}',k,k',\alpha,\Delta t) = P(\mathrm{B}|\mathrm{A}',k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t}) - \frac{1}{4\alpha\gamma}[(k-k')^2 - \gamma^2] \times (e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

$$P(\mathrm{A}|\mathrm{A}',k,k',\alpha,\Delta t) = P(\mathrm{B}|\mathrm{B}',k,k',\alpha,\Delta t) = \frac{1}{4}(1 + e^{-2\alpha\Delta t}) + \frac{1}{4\alpha\gamma}[(k-k')^2 - \gamma^2] \times (e^{-\kappa_+\Delta t} - e^{-\kappa_-\Delta t})$$

where

$$\gamma = \sqrt{\alpha^2 + (k-k')^2}$$

and

$$\kappa_\pm = \alpha + k + k' \pm \gamma$$

References and Notes

(1) Moerner, W. E. *J. Phys. Chem. B* **2002**, *106*, 910−927.

(2) Rigler, R.; Edman, L.; Foldes-Papp, Z.; Wennmalm, S. *Springer Ser. Chem. Phys.* **2001**, *67*, 177−194.

(3) Xie, X. S.; Lu, H. P. *Springer Ser. Chem. Phys.* **2001**, *67*, 227−240.

(4) Basche, T.; Nie, S.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10527−10528.

(5) Weiss, S. *Nat. Struct. Biol.* **2000**, *7*, 724−729.

(6) Tamarat, P.; Maali, A.; Lounis, B.; Orrit, M. *J. Phys. Chem. A* **2000**, *104*, 1−16.

(7) Ambrose, W. P.; Goodwin, P. M.; Jett, J. H.; van Orden, A.; Werner, J. H.; Keller, R. A. *Chem. Rev.* **1999**, *99*, 2929−2956.

(8) Nie, S.; Zare, R. N. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 567−596.

(9) Jia, Y.; Talaga, D. S.; Lau, W. L.; Lu, H. S. M.; DeGrado, W. F.; Hochstrasser, R. M. *Chem. Phys.* **1999**, *247*, 69−83.

(10) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5179−5184.

(11) Talaga, D. S.; Lau, W. L.; Roder, H.; Tang, J.; Jia, Y.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13021−13026.

(12) Talaga, D. S.; Jia, Y.; Bopp, M. A.; Sytnik, A.; DeGrado, W. A.; Cogdell, R. J.; Hochstrasser, R. M. *Springer Ser. Chem. Phys.* **2001**, *67*, 313−325.

(13) Rhoades, E.; Gussakovsky, E.; Haran, G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3197 3202.

(14) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743.

(15) Mei, E.; Tang, J.; Vanderkooi, J. M.; Hochstrasser, J. M. *J. Am. Chem. Soc.* **2003**, *125*, 2730−2735.

(16) Xie, X. S.; Lu, H. P.; Xun, L. In *Book of Abstracts, 216th ACS National Meeting, Boston, August 23−27, 1998*; American Chemical Society: Washington, DC, 1998; p PHYS-394.

(17) Ha, T.; Ting, A. Y.; Liang, J.; Caldwell, W. B.; Deniz, A. A.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 893−898.

(18) Schenter, G. K.; Lu, H. P.; Xie, X. S. *J. Phys. Chem. A* **1999**, *103*, 10477−10488.

(19) Edman, L.; Rigler, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8266−8271.

(20) Dovichi, N. J.; Polakowski, R.; Skelly, A.; Craig, D. B.; Wong, J. *Springer Ser. Chem. Phys.* **2001**, *67*, 241−256.

(21) Zhuang, X.; Bartley, L. E.; Babcock, H. P.; Russell, R.; Ha, T.; Herschlag, D.; Chu, S. *Science* **2000**, *288*, 2048−2051.

(22) Xie, X. S.; Dunn, R. C. In *Proceedings of the Nineteenth DOE Solar Photochemistry Research Conference*; U.S. Department of Energy: Washington, DC, 1995; pp 60−62.

(23) Ying, L.; Xie, X. S. *J. Phys. Chem. B* **1998**, *102*, 10399−10409.

(24) Rothwell, P. J.; Berger, S.; Kensch, O.; Felekyan, S.; Antonik, M.; Wohrl, B. M.; Restle, T.; Goody, R. S.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1655−1660.

(25) Lu, H. P.; Xun, L.; Xie, X. S. *Science* **1998**, *282*, 1877−1882.

(26) Wang, J.; Wolynes, P. *Phys. Rev. Lett.* **1995**, *74*, 4317−4320.

(27) Polakowski, R.; Craig, D. B.; Skelley, A.; Dovichi, N. J. *J. Am. Chem. Soc.* **2000**, *122*, 4853−4855.

(28) Kelley, A. M.; Michalet, X.; Weiss, S. *Science* **2001**, *292*, 1671−1672.

(29) Onuchic, J. N.; Wang, J.; Wolynes, P. G. *Chem. Phys.* **1999**, *247*, 175−184.

(30) Chesko, J. D.; Lu, H. P.; Xie, X. S. In *Book of Abstracts, 213th ACS National Meeting, San Francisco, April 13−17, 1997*; American Chemical Society: Washington, DC, 1997; p PHYS-043.

(31) Yang, H.; Xie, X. S. *J. Chem. Phys.* **2002**, *117*, 10965−10979.

(32) Kuhnemuth, R.; Seidel, C. A. M. *Single Mol.* **2001**, *2*, 251−254.

(33) Tellinghuisen, J.; Goodwin, P. M.; Ambrose, W. P.; Martin, J. C.; Keller, R. A. *Anal. Chem.* **1994**, *66*, 64−72.

(34) Smith, D. A.; Steffen, W.; Simmons, R. M.; Sleep, J. *Biophys. J.* **2001**, *81*, 2795−2816.

(35) Maus, M.; Cotlet, M.; Hofkens, J.; Gensch, T.; De Schryver, F. C.; Schaffer, J.; Seidel, C. A. M. *Anal. Chem.* **2001**, *73*, 2078−2086.

(36) Enderlein, J.; Goodwin, P. M.; van Orden, A.; Ambrose, W. P.; Erdmann, R.; Keller, R. A. *Chem. Phys. Lett.* **1997**, *270*, 464−470.

(37) Ross, S. M. *Introduction to Probability Models*, 7th ed.; Harcourt Academic Press: San Diego, CA, 2000.

(38) Gillespie, D. T. *Markov Processes: An Introduction for Physical Scientists*; Academic Press: Boston, 1992.

(39) Sivia, D. S. *Data Analysis: A Bayesian Tutorial*; Oxford University Press: Oxford, U.K., 1996.

(40) Johnson, N. L.; Kotz, S. *Distributions in Statistics: Continuous Univariate Distributions-1*; Houghton Mifflin Co.: Boston, 1970.

(41) Levinson, S. E.; Rabiner, L. R.; Sondhi, M. M. *Bell Syst. Technol. J.* **1983**, *62*, 1035−1074.

(42) Rabiner, L. R.; Juang, B. H. *IEEE ASSP Mag.* **1986**, *3*, 4−16.

(43) Durbin, R.; Eddy, S. R.; Krogh, A.; Mitchison, G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1998.
(44) Scott, S. L. *J. Am. Stat. Assoc.* **2002**, *97*, 337−351.
(45) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1091.
(46) Hastings, W. K. *Biometrika* **1970**, *57*, 97−109.
(47) Rabiner, L. R. *Proc. IEEE* **1989**, *77*, 257−286.
(48) Andrec, M.; Prestegard, J. H. *J. Magn. Reson.* **1998**, *130*, 217−232.
(49) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer-Verlag: New York, 2001.
(50) The three models (two-state noiseless, two-state HMM, and four-state noiseless HMM) give maximized logarithmic likelihoods of −4924.8, −4919.9, and −4886.5, respectively, leading to BIC values using eq 16 of 9859, 9867, and 9801. These BIC values can be translated into relative odds for each of the models of $\sim 10^{-13}$, 10^{-15}, and 1, respectively.
(51) Kolomeisky, A. B.; Widom, B. *J. Stat. Phys.* **1998**, *93*, 633−645.
(52) Weiss, G. H.; Masoliver, J. *Physica A* **2001**, *296*, 75−82.
(53) Karlin, S.; Taylor, H. M. *A First Course in Stochastic Processes*; Academic Press: New York, 1975.