

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/49819203>

Maximum likelihood-based analysis of single-molecule photon arrival trajectories

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · FEBRUARY 2011

Impact Factor: 2.95 · DOI: 10.1063/1.3544494 · Source: PubMed

CITATIONS

4

READS

20

2 AUTHORS, INCLUDING:



Marta Waligorska

Adam Mickiewicz University

3 PUBLICATIONS 10 CITATIONS

SEE PROFILE

Maximum likelihood-based analysis of single-molecule photon arrival trajectories

Marta Hajdziona and Andrzej Molski^{a)}

Faculty of Chemistry, Adam Mickiewicz University, Grunwaldzka 6, 60-780 Poznań, Poland

(Received 21 September 2010; accepted 3 January 2011; published online 2 February 2011)

In this work we explore the statistical properties of the maximum likelihood-based analysis of one-color photon arrival trajectories. This approach does not involve binning and, therefore, all of the information contained in an observed photon trajectory is used. We study the accuracy and precision of parameter estimates and the efficiency of the Akaike information criterion and the Bayesian information criterion (BIC) in selecting the true kinetic model. We focus on the low excitation regime where photon trajectories can be modeled as realizations of Markov modulated Poisson processes. The number of observed photons is the key parameter in determining model selection and parameter estimation. For example, the BIC can select the true three-state model from competing two-, three-, and four-state kinetic models even for relatively short trajectories made up of 2×10^3 photons. When the intensity levels are well-separated and 10^4 photons are observed, the two-state model parameters can be estimated with about 10% precision and those for a three-state model with about 20% precision. © 2011 American Institute of Physics. [doi:10.1063/1.3544494]

I. INTRODUCTION

When an immobilized single molecule is illuminated with laser light, the fluctuations (blinking) of the fluorescence intensity can report on the internal dynamics of the molecule. Modeling and statistical analysis of single-molecule fluorescence blinking has been a subject of great interest in single-molecule fluorescence spectroscopy.^{1–3} In a common on–off analysis a photon arrival time trajectory is binned into a fluorescence intensity trajectory and an appropriate threshold is used to separate a high (on) fluorescence state and a low (off) fluorescence state. Histograms of the on- and off-dwell times are then used to extract the on- and off-state escape rates. However, the outcome of such an analysis depends on the choice of the bin width and threshold. An increased bin width gives a better signal-to-noise ratio, but lowers the time resolution since rapid transitions cannot be detected. This is a drawback of all methods that involve binning of arrival time trajectories.

A sequence of photons emitted by a single molecule carries all the information about the internal dynamics of the molecule. Equivalent information is contained in the sequence of interarrival times (Fig. 1). Andreć *et al.*⁴ demonstrated that likelihood-based methods can be applied to the analysis of photon trajectories from single-molecule fluorescence experiments. In Ref. 4, a Bayesian perspective was adopted and Markov chain Monte Carlo sampling was employed to determine the model parameters from two-color photon trajectories. An analogous approach was applied to photon delay trajectories by Kou *et al.*⁵ In Refs. 4 and 5, models with two and four distinct states were considered in connection with the data analysis presented in Ref. 6.

Here we focus on the maximum likelihood-based analysis of one-color photon arrival trajectories that can be

viewed as realizations of Markov modulated Poisson processes (MMPPs).^{7,8} A MMPP is a point process of arrivals associated with a continuous time Markov chain. Here we interpret arrivals as photon detection instances, and the underlying Markov process as a representation of the dynamics of a single molecule. A MMPP can be constructed by varying the arrival (i.e., detection) rate of a Poisson process according to an m -state continuous time Markov chain. When the Markov chain is in state i , arrivals occur according to a Poisson process with rate I_i . The MMPP is parameterized by the rate constants k_{ij} for the jumps from state i to state j and the Poisson arrival rates I_1, I_2, \dots, I_m .

Recently, Jäger *et al.*⁸ applied the concept of a MMPP to quantify fluorescence quenching of a tetramethylrhodamine probe reporting on the reversible binding of Cu^{2+} ions to 2,2-bipyridine-4,4-dicarboxylic acid. Unlike in Refs. 4 and 5, where a Bayesian perspective was used, Jäger *et al.* estimated the model parameters by maximizing the likelihood function. As stated in Ref. 8, the number of detected photons in recorded sequences was high ($1\text{--}3 \times 10^6$ photons). However, in a typical single-molecule study the trajectory length is limited by photobleaching. Thus, the question arises whether the maximum likelihood analysis can be useful for shorter trajectories.

A fundamental issue in kinetic modeling is the distinguishability problem, i.e., the question whether there are multiple statistically equivalent models that can explain the data with equal probability. According to Ryden,⁹ when all the intensities I_1, I_2, \dots, I_m are distinct then the only equivalent MMPPs of the same order m are those obtained by a trivial relabeling of the states. However, when some of the intensities are equal, there exist competing models that explain the data equally likely, so that the true model cannot be identified unless additional information is available.

The present paper deals with the problem of model selection and parameter estimation in single-molecule

^{a)}Electronic mail: amolski@amu.edu.pl.

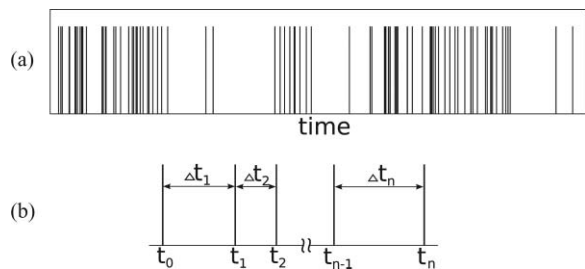


FIG. 1. (a) A photon arrival trajectory. (b) Representation in terms of inter-arrival times.

spectroscopy. First, we show that the maximum likelihood-based analysis of one-color photon trajectories can be applied not only to long trajectories as studied by Jäger *et al.*⁸ but also to short trajectories limited by photobleaching. Second, we quantify in a systematic way the bias and error of parameter estimates and show that this approach is superior to methods based on binning (e.g., on-off analysis and photon counting histogram). Third, our simulations justify the use of the Bayesian information criterion (BIC) for model selection. Fourth, we bring out and give an example of distinguishability problems that may arise when fluorescence trajectories are analyzed.

The outline of the paper is as follows. In Sec. II, we briefly recall the concept of a MMPP and present our procedure for data analysis. In Sec. III, we give illustrative examples of parameter estimation for two- and three-state models, and compare the present method with the standard on-off analysis and the estimation based on the photon counting histogram. In Sec. IV, we compare the BIC and Akaike information criterion (AIC) as tools to select the true model. In Sec. V, we address the distinguishability problem and present two three-state models that are statistically equivalent. A summary and comments are given in Sec. VI. Finally, the expectation-maximization (EM) algorithm used in this work is briefly reviewed in Appendix.

II. MODEL AND DATA ANALYSIS

A. Markov modulated Poisson processes (MMPPs)

Below we recall the concept of a MMPP and present a formula for calculating the likelihood of an arrival trajectory. A comprehensive review of MMPPs is given in Refs. 10–12.

We assume that the molecule has m internal states, and that its dynamics are determined by a generator matrix \mathbf{Q} with elements k_{ij} . If the molecule is in state i at time t , it can jump in an infinitesimal time interval $(t, t + dt)$ to a different state $j \neq i$ with conditional probability $k_{ij}dt$, or it can stay in state i with probability $1 - \sum_{j \neq i} k_{ij}dt$. The off-diagonal elements k_{ij} , $i \neq j$, of matrix \mathbf{Q} are positive and determine the jump rates between different states. The diagonal elements k_{ii} are negative, $k_{ii} = -\sum_{j \neq i} k_{ij}$, and determine the escape rate from states $i = 1, 2, \dots, m$.

The rate of arrivals is determined by an arrival rate matrix $\mathbf{\Lambda}$, where the diagonal elements are the Poisson arrival rates I_i for states $i = 1, \dots, m$, and the off-diagonal elements are all zero. When the molecule process stays in i in $(t, t + dt)$,

it generates an arrival with conditional probability $I_i dt$. Arrival instances t_0, t_1, \dots, t_n form a MMPP that is used here to model photon arrival time trajectories of single immobilized molecules. We focus on the statistics of the interarrival times $Y_k = \Delta t_k = t_k - t_{k-1}$.

Let $X(t)$ denote the state of the underlying rate processes at time t , and X_k , $k \geq 0$, denote the state right after a k th arrival, i.e., $X_k = X(t_k)$. Thus, $X_k = i$, if the arrival happened when the processes stayed in state i . Let Y_k , $k \geq 0$, denote the interval time between the $(k-1)$ th and the k th arrivals, $Y_k = \Delta t_k$, $Y_0 = 0$. The key observation for developing a formula for the likelihood of the observed sequence arrival times is that the bivariate process $\{(X_k, Y_k); k \geq 0\}$ is a Markov renewal process. Note that the states X_k at arrivals t_k are not observed. The likelihood of an arrival trajectory t_0, t_1, \dots, t_n is

$$L = \mathbf{P}^{\text{ini}} \prod_{k=1}^n \{\exp[(\mathbf{Q} - \mathbf{\Lambda})\Delta t_k] \mathbf{\Lambda}\} \mathbf{1}, \quad (1)$$

where $\mathbf{P}^{\text{ini}} = [P_1^{\text{ini}}, P_2^{\text{ini}}, \dots, P_n^{\text{ini}}]$ is a row vector of initial state probabilities of the underlying chain and $\mathbf{1}$ is a column vector of ones, i.e., $\mathbf{1} = [1, 1, \dots, 1]^T$. The meaning of the terms in Eq. (1) is the following. The state probabilities at t_0 are determined by \mathbf{P}^{ini} . Each matrix $\exp[(\mathbf{Q} - \mathbf{\Lambda})\Delta t_k]$ determines the conditional transition probabilities between states with no photon arrival in an interval (t_{k-1}, t_k) . The conditional probability densities of an arrival at t_k are given by $\mathbf{\Lambda}$. The final multiplication by vector $\mathbf{1}$ sums up the contribution from different states at t_n . A derivation of Eq. (1) can be found in Ref. 10. Analogous expression was obtained by Kou *et al.*⁵ for single-molecule photon delay trajectories, and by Gopich and Szabo¹³ for two-color photon sequences.

B. Data analysis

We simulated arrivals from MMPPs with different number of states m . The simulated arrival time trajectories were fitted with kinetic models by numerically maximizing the likelihood function Eq. (1).

In our parameter estimation study, arrival time trajectories were fitted with the true model, i.e., the one used to generate arrivals. Each parameter estimation experiment was repeated 10^3 times, and the estimated parameters \hat{k}_{ij} and \hat{I}_i were recorded. To facilitate comparison of different parameter sets, we scaled the estimated parameters by the true values used in the simulations, $\kappa_{ij} = \hat{k}_{ij}/k_{ij}$ and $\iota_i = \hat{I}_i/I_i$. The rates k_{ij} and intensities I_i have units of inverse time. Since the scaled estimates are unitless, we use the relative units for the rates and intensities, which makes the presentation slightly more general.

In our model selection study, each arrival time trajectory was fitted using not only the true model but also models with different number of states. The models selected as the right ones by the AIC and the BIC were recorded. Each model selection experiment was repeated to determine the percentage of successful model identifications.

TABLE I. The average (avg) and standard deviation (sdv) of 10^3 scaled estimates of the transition rates and fluorescence intensities for two-state models with $k_{12} = 1$, $I_1 = 10$, $I_2 = 40$, and different k_{21} . The trajectory length $n = 5 \times 10^4$ photons. The last column gives the number of fits that converged in less than 5×10^3 iterations.

k_{21}	avg (sdv)			avg (sdv)		No. of fits conv.
1.0	κ_{12}	1.001 (0.042)	ι_1	1.0000 (0.013)	1000	
	κ_{21}	1.001 (0.045)	ι_2	1.0000 (0.0060)		
10.0	κ_{12}	1.001 (0.058)	ι_1	1.0001 (0.0083)	1000	
	κ_{21}	1.001 (0.043)	ι_2	1.000 (0.019)		
50.0	κ_{12}	1.04 (0.32)	ι_1	0.999 (0.011)	971	
	κ_{21}	1.01 (0.16)	ι_2	1.01 (0.12)		
75.0	κ_{12}	0.77 (0.29)	ι_1	1.004 (0.010)	477	
	κ_{21}	0.96 (0.25)	ι_2	1.12 (0.21)		

Numerical maximization of the likelihood (1) was addressed in several studies.^{10,14,15} We carried out simulations and fitting using an expectation-maximization algorithm implemented in a C++ package IP2BMAP developed by Lindemann *et al.*¹⁴ For selected parameter sets, the simulation and fitting were repeated using an R package HIDDENMARKOV developed by Harte (<http://cran.at.r-project.org/web/packages/HiddenMarkov>) as well as a multidimensional optimization routine. The results were similar.

III. PARAMETER ESTIMATION

The simplest kinetic model is a two-state model:



where k_{12}, k_{21} are transition rates between states 1 and 2. The Poisson arrival intensities for states 1 and 2 are I_1 and I_2 . First we use the two-state model to show that even fast transitions can be quantified using the present approach and to demonstrate its superior time resolution over the standard on-off analysis and the analysis based the photon counting histogram.¹⁶

Table I compares the average and standard deviation of the scaled parameter estimates for a model with $k_{12} = 1$, $I_1 = 10$, $I_2 = 40$ when the rate k_{21} increases from 1 to 75. The trajectories are made up of $n = 5 \times 10^4$ photons. The transition rate $k_{21} = 1$ is slow compared with the intensity $I_2 = 40$;

however, it becomes fast when $k_{21} = 75$. Two points are worth noting. The first is that the estimates of the intensities are better than those of the rates. This holds true for all models studied in this work. As the transition rates between different states, rather than the fluorescence intensities, are interesting to the experimenter, we will focus the following presentation on the recovery of the transition rates. The second is that the parameter estimates get worse as the transition $2 \rightarrow 1$ gets faster. Nevertheless, the escape rates can be estimated even when the average dwell time in state 2, k_{21}^{-1} , is comparable with the average interphoton time I_2^{-1} .

Table II compares the maximum likelihood approach with the on-off analysis and the photon counting histogram (PCH) analysis. Clearly, the present approach produces better parameter estimates. This is related to the fact that the on-off analysis and the PCH analysis involve binning and, additionally, the on-off analysis involves thresholding.

Estimates of the rates improve when the fluorescence intensities I_1 and I_2 and the trajectory length n increase. Arrival trajectories for a two-state model with unit escape rates $k_{12} = k_{21} = 1$ were simulated for the trajectory length ranging from $n = 10^3$ to 5×10^4 . It follows from Table III that the rates can be recovered with a precision of 15% even for short trajectories made up of $n = 5 \times 10^3$ photons when the intensities are well-separated.

In the presence of Poissonian background noise, the procedure does not change and the intensities and the rates can be found by maximizing the same likelihood function. The recovered intensities I_i are now sums, $I_i = I_i^m + I^b$, of contributions from the molecule, I_i^m , and from background, I^b . The molecular intensities I_i^m can be obtained by shifting the extracted intensities I_i by I^b .

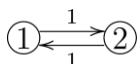
The presence of background lowers the quality of the recovered parameters as it lowers the relative difference between the intensity levels. The separation between two intensity values can be expressed as the ratio of the lower intensity, say I_1 , to the higher intensity, say I_2 . In the presence of background, the ratio $I_1/I_2 = (I_1^m + I^b)/(I_2^m + I^b)$ gets larger as the background intensity I^b increases. As the ratio I_1/I_2 approaches 1, the accuracy and precision of parameter estimates decreases.

How close the intensities can get before the present approach breaks down? To address this question we performed a series of simulations for a two-state model with $k_{12} = k_{21} = 1$, and $I_1 + I_2 = 50$, for different ratios I_1/I_2

TABLE II. Comparison of the maximum likelihood, on-off, and PCH estimates. The average and standard deviations of 10^3 scaled estimates of transition rates for different two-state models and the trajectory length $n = 5 \times 10^4$ photons. The sign “—” indicates that the procedure failed.

		$\textcircled{1} \xrightleftharpoons[0.1]{0.1} \textcircled{2}$	$\textcircled{1} \xrightleftharpoons[1]{1} \textcircled{2}$	$\textcircled{1} \xrightleftharpoons[1]{1} \textcircled{2}$
		$I_1 = 10, I_2 = 40$	$I_1 = 10, I_2 = 40$	$I_1 = 1, I_2 = 4$
PCH	κ_{12}	1.37 (0.67)	1.05 (0.10)	—
	κ_{21}	1.27 (0.53)	1.08 (0.14)	—
On-off	κ_{12}	1.05 (0.16)	0.563 (0.035)	1.134 (0.0247)
	κ_{21}	1.04 (0.17)	0.477 (0.030)	0.2500 (0.0047)
MMPP	κ_{12}	1.01 (0.11)	1.003 (0.043)	1.001 (0.047)
	κ_{21}	1.00 (0.10)	1.001 (0.043)	1.001 (0.050)

TABLE III. The average and standard deviation of 10^3 scaled estimates of transition rates for two-state models and different trajectory lengths n .

		$n = 10^3$	$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
$I_1 = 0.2, I_2 = 0.8$	κ_{12}	1.04 (0.68)	1.00 (0.33)	1.011 (0.22)	0.999 (0.067)
	κ_{21}	1.07 (0.83)	1.03 (0.35)	1.012 (0.22)	1.004 (0.081)
$I_1 = 1, I_2 = 4$	κ_{12}	1.05 (0.39)	1.01 (0.16)	1.004 (0.10)	1.001 (0.047)
	κ_{21}	1.04 (0.40)	1.00 (0.16)	1.009 (0.11)	1.001 (0.050)
$I_1 = 10, I_2 = 40$	κ_{12}	1.06 (0.33)	1.01 (0.13)	1.004 (0.096)	1.001 (0.042)
	κ_{21}	1.05 (0.33)	1.01 (0.14)	1.003 (0.097)	1.001 (0.045)

from 0.2 to 0.8 and trajectory lengths from $n = 5 \times 10^3$ to 5×10^4 photons. As expected, the bias and standard deviation of the recovered rate constants increase as the ratio I_1/I_2 increases. The best estimates are obtained for the longest trajectory $n = 5 \times 10^4$, where the relative bias remains less than 2% and the relative standard deviation less than 20% for $I_1/I_2 \leq 0.7$. For short trajectories, $n = 5 \times 10^3$, this accuracy and precision can be obtained only when $I_1/I_2 \leq 0.4$.

The maximum likelihood approach can be used for models with more than two states. For the sake of illustration, in Table IV we compare the parameter estimation for linear three- and four-state models with well-separated intensities, $I_1 = 10$, $I_2 = 40$, $I_3 = 80$, and $I_4 = 120$. For a given number of photons, the rate estimates are better for the three-state model since it has fewer parameters.

IV. MODEL SELECTION

In Sec. III we demonstrated that, given the true kinetic model, the maximum likelihood analysis of photon arrival trajectories is effective in estimating the model parameters. In this section we show that the maximum likelihood analysis can also be used for selecting the true model among competing models.

When different models are fitted to the same data, models with more fit parameters tend to produce a greater likelihood. One solution to this problem is to reward parsimony by penalizing the likelihood for the number of parameters in the fitted model. Akaike proposed a measure of the goodness of fit in-

cluding the effect of the number of parameters.¹⁷ The AIC is calculated as

$$\text{AIC} = -2 \ln L + 2p, \quad (3)$$

where L is the maximized likelihood for the estimated model and p is the number of free parameters to be estimated. Given an observed photon trajectory, a model is better by the AIC when its AIC in Eq. (3) is lower. A similar criterion called the BIC allows not only for the number of parameters in the fitted model but also for the number of data points n :¹⁸

$$\text{BIC} = -2 \ln L + p \ln n. \quad (4)$$

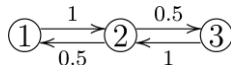
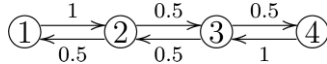
The BIC was used for model selection in Refs. 4 and 8.

To the best of our knowledge, the application of the AIC and BIC to MMPPs has not been theoretically justified. Here we use simulations to compare the performance of those two criteria and to show that the BIC works well even for short trajectories.

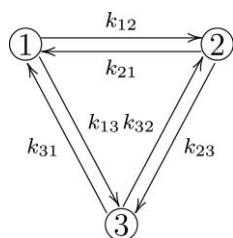
Table V compares the efficiency of the AIC and BIC for a three-state model with $I_1 = 10$, $I_2 = 40$, and $I_3 = 80$. The rate parameters were chosen so that the average dwell time in each state is 1. The BIC works better than the AIC for those particular model parameters. Interestingly, the BIC can select the true model with high efficiency (99%) even for short trajectories made up of 2×10^3 photons.

The model selection procedure is further illustrated in Fig. 2 that shows $\Delta\text{BIC} = \text{BIC} - \text{BIC}_{\min}$ together with its standard deviation as a function of the number of states m . BIC_{\min} is the smallest BIC value among the competing models. Thus, $\Delta\text{BIC} = 0$ for a model selected by the BIC. Two circular (C) models are considered,

TABLE IV. The average and standard deviations of 10^3 scaled estimates of transition rates for exemplary three- and four-state models with $I_1 = 10$, $I_2 = 40$, $I_3 = 80$, and $I_4 = 120$.

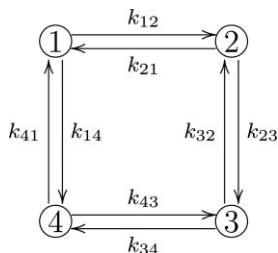
Model		5×10^3	10^4	5×10^4
	κ_{12}	1.05 (0.26)	1.02 (0.18)	1.005 (0.074)
	κ_{21}	1.02 (0.24)	1.01 (0.17)	1.003 (0.079)
	κ_{23}	1.03 (0.27)	1.02 (0.18)	1.007 (0.079)
	κ_{32}	1.03 (0.27)	1.02 (0.18)	1.002 (0.077)
	κ_{12}	1.14 (0.76)	1.04 (0.28)	1.01 (0.11)
	κ_{21}	1.06 (0.48)	1.03 (0.27)	1.00 (0.12)
	κ_{23}	1.09 (0.40)	1.03 (0.26)	1.00 (0.10)
	κ_{32}	1.03 (0.36)	1.02 (0.24)	1.00 (0.10)
	κ_{34}	1.11 (0.51)	1.05 (0.32)	1.00 (0.14)
	κ_{43}	1.11 (0.55)	1.07 (0.34)	1.01 (0.13)

3C



(5)

and 4C



(6)

along with linear models.

Figure 2 illustrates two points. First, the present approach can be applied to models with different topologies, for instance, circular models, where the condition of detailed balance imposes additional constraints on estimated rate constants (see Appendix for a brief comment). Second, ΔBIC is a random quantity whose fluctuations depend on the model and trajectory length. Since long trajectories were used, $n = 5 \times 10^4$ photons, the fluctuations of ΔBIC were small, and the true model was correctly selected in each case.

Once a model was selected, it is important to assess how confident we can be in that selection. A practical question is: How large an increase in BIC should be to make sure that this is not a random fluctuation? Clearly, the larger the BIC (or AIC) difference for a model, the less probable that model is. Apparently, as there no general results available to quantify this statement, one needs to resort to simulations in order to assess the performance of a model selection criterion.

V. EXAMPLE OF INDISTINGUISHABLE MODELS

In the previous sections we studied distinguishable models where each state has a different fluorescence intensity I_i . When two or more intensities I_i are equal, there exist models that are statistically equivalent to the true model. Those models produce the same likelihood of the data and cannot be distinguished from the true model unless additional information is available.

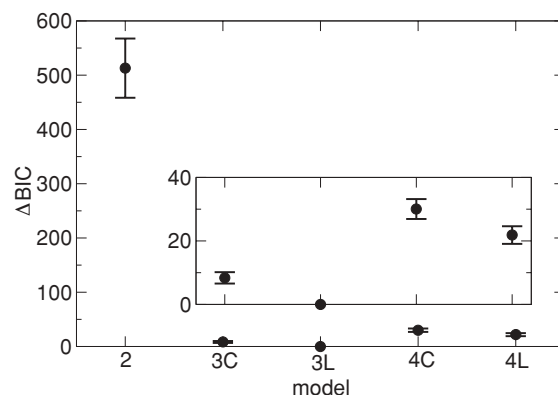


FIG. 2. Model selection using the BIC. For a three-state linear (3L) model, 10^2 trajectories with $n = 5 \times 10^4$ photon arrivals were generated and fitted using linear (L) and circular (C) kinetic models. The fitted models are located along the x axis with increasing number of states $m = 2, 3, 4$. The y axis gives $\Delta\text{BIC} = \text{BIC} - \text{BIC}_{\min}$ for each model. The error bars show the standard deviation of ΔBIC . The inset shows ΔBIC for $m = 3$ and 4 enlarged for clarity.

As an example we consider a MMPP where the generator matrix is

$$\mathbf{Q} = \begin{bmatrix} -25 & 5 & 20 \\ 50 & -50 & 0 \\ 1 & 0 & -1 \end{bmatrix} \text{ s}^{-1} \quad (7)$$

and the arrival rate matrix is

$$\mathbf{\Lambda} = \begin{bmatrix} 10^4 & 0 & 0 \\ 0 & 10^3 & 0 \\ 0 & 0 & 10^3 \end{bmatrix} \text{ s}^{-1}. \quad (8)$$

This three-state model corresponds roughly to model-3b from Ref. 8 for the case when the concentration of Cu^{2+} ions is $1 \mu\text{M}$. Note that we do not attempt to interpret the blinking studied in Ref. 8, but rather we use that experimental work as the motivation for the numerical values in Eqs. (7) and (8). Note also that in Ref. 8 the intensities were estimated independently, whereas here the maximum likelihood is used to estimate both the intensities and the rates.

The arrival rate matrix (8) has two equal intensities, which implies that there exist different statistically equivalent models. To find those models one can apply a similarity transformation to the true model. Two MMPP models $(\mathbf{Q}, \mathbf{\Lambda})$ and $(\mathbf{Q}', \mathbf{\Lambda}')$ are statistically equivalent when there exists a matrix \mathbf{S} where $\mathbf{S}\mathbf{1} = \mathbf{1}$, such that $\mathbf{S}\mathbf{Q}\mathbf{S}^{-1} = \mathbf{Q}'$ and $\mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1} = \mathbf{\Lambda}'$.

TABLE V. The efficiency of the AIC and BIC for a linear three-state model with transition rates and intensities as indicated. 10^3 photon trajectories of a given length n were generated and each photon trajectory was fitted with linear two-, three- and four-state models. The AIC and the BIC were applied to select the true model. The percentage of correct identifications is shown for a given number of photons n .

Simulated model	Size of fitted models	Selection criterion	n		
			2×10^3	5×10^3	10^4
$10 \xrightleftharpoons[0.5]{1} 40 \xrightleftharpoons[1]{0.5} 80$	2, 3, 4	AIC	94.3%	95.7%	94.8%
		BIC	99.0%	99.8%	100%

The model $(\mathbf{Q}', \mathbf{A})$ with the generator matrix

$$\mathbf{Q}' = \begin{bmatrix} -25 & 0 & 25 \\ 0 & -4.62963 & 4.62963 \\ 10.8 & 35.57037 & -46.37037 \end{bmatrix} \text{s}^{-1} \quad (9)$$

is statistically equivalent to the model defined in Eqs. (7) and (8). Equations (7) and (9) are examples of generator matrices for a model whose true kinetic parameters cannot be uniquely recovered from a single photon trajectory.

In model-3b studied in Ref. 8, one of the rates is concentration-dependent $k = k_{\text{assoc}}c$, where c is the ligand concentration, $c = [\text{Cu}^{2+}]$. When photon trajectories are measured at two or more concentrations c , k_{assoc} and other kinetic parameters can be recovered from a global (i.e., simultaneous) analysis of the photon trajectories. We do not pursue this issue further in the present work.

VI. SUMMARY AND COMMENTS

In this paper, we have studied the maximum likelihood-based approach to extract information from one-color photon arrival trajectories. This approach does not involve binning and, therefore, all of the information contained in an observed photon trajectory is used. We focused on the low excitation regime of immobilized molecules where photon trajectories can be modeled as Markov modulated Poisson processes. This is a good approximation when the times between detected photons are much longer than the excited-state relaxation time.¹³ When this is not the case, an appropriate likelihood function can be obtained using the concept of a Markov arrival process.¹⁴

The state-dependent fluorescence intensities and the rates of transitions between different states are obtained by numerical maximization of an appropriate likelihood function [Eq. (1)]. The true kinetic model can be selected by comparing the penalized maximized likelihood of competing kinetic models [Eqs. (3) and (4)]. We found that the number of available photons is the key parameter in determining the efficiency of model selection and the quality of parameter estimates. A longer trajectory allows for a more efficient model selection and better parameter estimation. For simple two- and three-state models, relatively short trajectories made up of 10^4 photons can be used for an efficient model selection and parameter estimation.

There are several implications of this study that are relevant for single-molecule spectroscopy. First, the maximum likelihood-based analysis of photon arrival trajectories can be applied not only to long trajectories as in Ref. 8 but also to short trajectories limited by photobleaching. Second, this approach has a better time resolution than the on-off analysis and the photon counting histogram analysis. Third, the maximum likelihood parameter estimation requires that the number of states and their connectivity are known. We found, however, that the Bayesian information criterion is effective in selecting the right model from a set of competing models. Thus, our simulation study justifies the use of this criterion in Ref. 8. Fourth, a simultaneous analysis of two or more trajectories may be needed to resolve the distinguishability issues

arising when the fluorescence intensities are not distinct. We hope to address this issue in a future study. Finally, our study further supports the application of likelihood-based methods in single-molecule fluorescence spectroscopy.

ACKNOWLEDGMENTS

This work has been supported by KBN (Grant Nos. NN 204 166736 and 204 017238).

APPENDIX: EXPECTATION-MAXIMIZATION ALGORITHM

In this appendix we briefly recall the concept of an EM algorithm and comment on its application to Markov modulated Poisson processes.

An expectation-maximization algorithm is a method for finding maximum likelihood estimates (MLEs) of parameters in statistical models, where the model depends on missing (unobserved) variables. General properties of the algorithm were demonstrated in Ref. 19 along with a large number of applications.

In single-molecule fluorescence spectroscopy, interphoton times, Y_k , can be measured but the state, $X(t)$, of a molecule is not observed. As pointed out by Ryden,¹⁰ it is useful to treat the entire trajectory $X(t)$ of the underlying Markov process as missing data. One can write the likelihood of the complete (observed + missing) data as

$$L^c(\Theta; Y, X) = p(Y, X|\Theta), \quad (\text{A1})$$

where $p(Y, X|\Theta)$ is the joined probability density of the complete data and Θ denotes the unknown parameters. The MLE of Θ is determined by maximizing the likelihood of the observed data:

$$L(\Theta; Y) = p(Y|\Theta) = \sum_x p(Y, X|\Theta). \quad (\text{A2})$$

In the present case, $\Theta = (\mathbf{Q}, \mathbf{A})$ and $L(\Theta; Y)$ is given by Eq. (1). The maximum of the likelihood function L can be found using a multidimensional nonlinear optimizer in connection with the marginal likelihood $L(\Theta; Y)$. An alternative is to use an EM algorithm in connection with the complete data likelihood $L^c(\Theta; Y, X)$.

EM is an iterative method which alternates between an expectation (E) step and a maximization (M) step. An E step computes the expectation of the complete data log-likelihood function with respect to the distribution of the missing data X given the current estimate Θ^0 of the parameters and the observed data Y :

$$LL^E(\Theta; \Theta^0, Y) = \sum_{X|\Theta^0} \log L^c(\Theta; Y, X). \quad (\text{A3})$$

A maximization step M computes parameters maximizing the expected log-likelihood found in the preceding E step:

$$\Theta^1 = \arg \max_{\Theta} LL^E(\Theta; \Theta^0, Y). \quad (\text{A4})$$

Explicit expressions for the complete data log-likelihood $\log L^c$ and the expectation LL^E of the complete data

log-likelihood function for a MMPP were given by Ryden in Ref. 10. An implementation of Ryden's algorithm can be found in an R package HIDDENMARKOV developed by Harte (<http://cran.at.r-project.org/web/packages/HiddenMarkov>). An EM algorithm applicable to broader class of point processes was developed by Lindemann *et al.* and implemented in a C++ package IP2BMAP.¹⁴

The EM algorithm can be directly applied to fit linear kinetic models; however, it needs to be modified to allow for detailed balance in circular models. Detailed balance holds at equilibrium and implies that all fluxes balance $P_i^{\text{eq}} k_{ij} = P_j^{\text{eq}} k_{ji}$, where \mathbf{P}^{eq} is the equilibrium distribution of states such that $\mathbf{P}^{\text{eq}} \mathbf{Q} = 0$. As pointed out in Ref. 20, the detailed balance condition can be written as

$$\text{diag}(\mathbf{P}^{\text{eq}}) \mathbf{Q} = [\text{diag}(\mathbf{P}^{\text{eq}}) \mathbf{Q}]^T, \quad (\text{A5})$$

where $\text{diag}(\mathbf{P}^{\text{eq}})$ is a diagonal matrix with elements P_i^{eq} along the diagonal and \mathbf{A}^T denotes the transpose of \mathbf{A} . An M step in a standard EM algorithm does not preserve detailed balance. To correct this, at each M step the following procedure was applied. First, estimates of \mathbf{Q} and \mathbf{P}^{eq} are obtained according to Ryden's prescription.¹⁰ Then a matrix

$$\mathbf{Q}_s = \{\text{diag}(\mathbf{P}^{\text{eq}}) \mathbf{Q} + [\text{diag}(\mathbf{P}^{\text{eq}}) \mathbf{Q}]^T\} / 2 \quad (\text{A6})$$

is calculated. The corrected rate matrix satisfying detailed balance is obtained as

$$\mathbf{Q} = [\text{diag}(\mathbf{P}^{\text{eq}})]^{-1} \mathbf{Q}_s, \quad (\text{A7})$$

where \mathbf{A}^{-1} denotes the inverse of \mathbf{A} . The corrected rate matrix is used in the next iteration step.

- ¹E. Barkai, Y. Jung, and R. Silbey, *Annu. Rev. Phys. Chem.* **55**, 457 (2004).
- ²M. Lippitz, F. Kulzer, and M. Orrit, *ChemPhysChem* **6**, 770 (2005).
- ³*Theory and Evaluations of Single-Molecule Signals*, edited by E. Barkai, F. Brown, M. Orrit, and H. Yang (World Scientific Publishing Co. Inc., River Edge, NJ, 2008).
- ⁴M. Andrec, R. M. Levy, and D. S. Talaga, *J. Phys. Chem. A* **107**, 7454 (2003).
- ⁵S. C. Kou, X. S. Xie, and J. S. Liu, *J. R. Stat. Soc., Ser. C, Appl. Stat.* **54**, 469 (2005).
- ⁶G. K. Schenter, H. P. Lu, and X. S. Xie, *J. Phys. Chem. A* **103**, 10477 (1999).
- ⁷T. Burzykowski, J. Szubiakowski, and T. Ryden, *Chem. Phys.* **288**, 291 (2003).
- ⁸M. Jäger, A. Kiel, D.-P. Herten, and F. Hamprecht, *ChemPhysChem* **10**, 2486 (2009).
- ⁹T. Ryden, *J. Appl. Probab.* **33**, 640 (1996).
- ¹⁰T. Ryden, *Comput. Stat. Data Anal.* **21**, 431 (1996).
- ¹¹W. Fischer and K. Meier-Hellstern, *Perform. Eval.* **18**, 149 (1992).
- ¹²T. Ryden, *Commun. Statist. Stoch. Models* **10**, 795 (1994).
- ¹³I. V. Gopich and A. Szabo, *J. Phys. Chem. B* **113**, 10965 (2009).
- ¹⁴A. Klemm, C. Lindemann, and M. Lohmann, *Perform. Eval.* **54**, 149 (2003).
- ¹⁵W. J. J. Roberts and Y. Ephraim, *IEEE Signal Process. Lett.* **13**, 373 (2006).
- ¹⁶M. Hajdziona and A. Molski, *Chem. Phys. Lett.* **470**, 363 (2009).
- ¹⁷H. Akaike, *IEEE Trans. Autom. Control* **19**, 716 (1974).
- ¹⁸G. Schwartz, *Ann. Stat.* **6**, 461 (1978).
- ¹⁹A. P. Dempster, N. M. Laird, and D. B. Rubin, *J. Royal Stat. Soc. Ser. B* **39**, 1 (1977) <http://www.jstor.org/stable/2984875>.
- ²⁰J. Yang, W. J. Bruno, W. S. Hlavacek, and J. E. Pearson, *Biophys. J.* **91**, 1136 (2006).