Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements

Lucas P. Watkins and Haw Yang*

Department of Chemistry, University of California at Berkeley, D46 Hildebrand Hall, Berkeley, California 94720

Received: July 22, 2004; In Final Form: October 21, 2004

We present a method for the analysis of optical single molecule emission data that exhibit discrete intensity jumps. This new method uses a generalized likelihood ratio test that determines the location of an intensity change point based on individual photon arrival times. This test is applied recursively to an entire single molecule intensity trajectory, thus finding each change points. Expectation−maximization clustering and the Bayesian information criterion is then used for accurate determination of the true number of states accessible to the system. This procedure allows rigorous and quantitative determination of intensity change points without the artificial time resolution limitations that arise from binning and thresholding.

1. Introduction

Complex systems generally exhibit a broad distribution in both their dynamic and static properties. As such, their properties are significantly obscured by the ensemble average. Time-resolved optical single-molecule spectroscopy is particularly well suited to the study of these systems.[1] However, the Poisson detection noise inherent in such measurements severely complicates analysis and interpretation.[2] In this context, the determination of intensity change points is worthy of particular study, since available methods often involve subjective parameters and are not statistically robust.

Intensity change points, discrete jumps in detected intensity followed by some period of constant intensity, occur in many systems and often arise from processes that occur much faster than the time scale of measurement. Although these events are rare on the molecular time scale, they are often the key to a complete understanding of the underlying dynamics. For example, intensity traces from individual semiconducting nanocrystals display an intermittency not seen in bulk measurements. This intermittency is thought to result from Auger ionization.[3−6] In biological macromolecules, single-molecule fluorescence studies reveal sudden jumps in detected fluorescence intensity during enzymatic reactions and structural changes.[7−13] Studies of these transitions between relatively stable molecular states have the potential to greatly increase our understanding of enzymatic dynamics and reactivity. Since molecules are studied one at a time, heterogeneous interactions between the molecule and its host environment can be observed. These heterogeneities can be observed in systems as simple as single fluorescent dyes and can be manifested as precipitous changes in optical signal due to electronic, structural, or orientational transitions.[14−23]

The simplest and most commonly used technique for the analysis of intensity change points, binning and thresholding, presents numerous difficulties. In this method, the photon arrival times are binned or filtered so that an average intensity is generated for each bin. This step is necessary to reduce the Poisson counting noise to a level where individual emissive states can be resolved. The choice of a bin width, however, introduces an artificial time scale to the measurement and raises the possibility of missed transitions due to the information lost in each bin. Thresholds are then assigned based on the acquired data, often by visually examining the binned single-molecule trace. Change points are then considered to occur wherever the averaged intensity crosses this threshold. This threshold-based analysis unfortunately requires knowledge of the number of available states for the molecule. A hard threshold is also vulnerable to the detection of false transitions (see Figure 1).

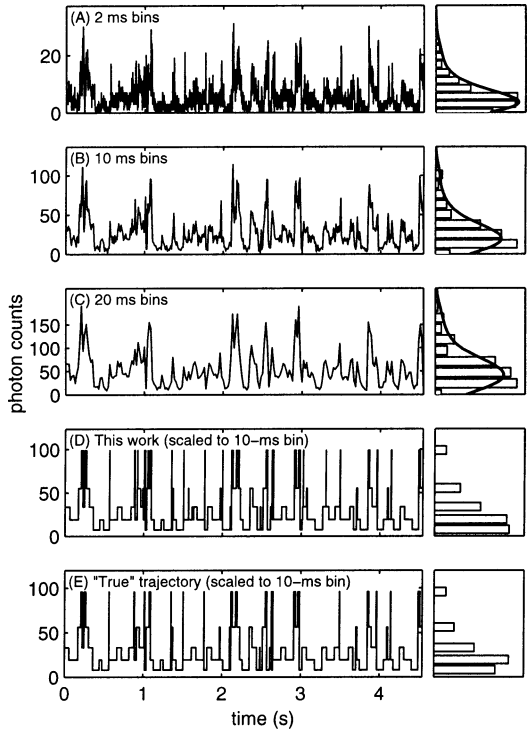

Figure 1. (A) 2-ms binned trajectory. (B) 10-ms binned trajectory. (C) 20-ms binned trajectory. (D) Photonby-photon reconstructed single-molecule intensity states. (E) "True" intensity states used in simulating panels A−C. The intensity distribution histograms are shown to the right of each panel, with the number of bins given by the usual rule $N_{bins} = \log_2 N + 1$. The solid lines overlaid on the histograms are density estimation using Gaussian kernels, with the smoothing parameter determined in the same way. The trajectory was simulated as described in section 3.2.

* To whom correspondence should be addressed.

10.1021/jp0467548 CCC: $30.25
Published on Web 12/14/2004

This problem is exacerbated when the signal-to-background ratio is low. The simulated 5-state single-molecule trace in Figure 1 illustrates these difficulties. The 2-ms binned trajectory in panel (A) appears very noisy; its intensity histogram in no way reflects the existence of 5 states. Further averaging, as shown in the 10-ms and 20-ms binned traces in panels B and C, reduces the extent of noise but does not help in the elucidation of change points. The underlying dynamics are not evident from any of the binned trajectories, and none of the binned trajectories allow for analysis of change points by thresholding. Furthermore, there is the issue of time resolution. Binning the data unnecessarily restricts measurement time resolution. Fast transitions are averaged out, whereas small and slow transitions are still lost in the noise. These concerns emphasize the need for quantitative, statistically robust methods of analysis. Indeed, although much progress has been made in experimental techniques,[24,25] the experimentalist is still faced with a dearth of rigorous methods for extracting information from a data set that is both limited in size and degraded by photon detection noise.

A method based on hidden Markov models has been proposed to extract kinetic parameters directly from unbinned photon by photon single-molecule trajectories.[26] This method achieves high time resolution, but it requires knowledge of the underlying kinetic scheme. For a data reduction procedure to be broadly applicable, it should be free of physical models such as a kinetic scheme. In addition, it should be objective so that no bias is introduced through user-adjustable parameters as in the aforementioned thresholding scheme. Since there are only a limited number of detected photons from a given single molecule, the method should be efficient so that each detected photon is taken into account. Finally, it should be quantitative so that a confidence interval can be associated with derivative parameters such as the lifetime of a particular molecular conformational state. Along this line, we have recently reported a method for the quantitative analysis of continuous dynamics from time-resolved single-molecule emission trajectories.[27] On the basis of information-theoretical considerations, we derived a basic equation that relates measurement uncertainty and time resolution. This general relationship allowed the formulation of an efficient algorithm that reaches the theoretical limit, extracting the maximal amount of information of the underlying dynamics one photon at a time. Continuing our effort to develop general, model-free methods to uncover molecular processes in time-resolved single-molecule experiments, this paper describes a technique for determining transition points and intensity levels from a photon-by-photon emission trajectory.

The main result of this work is presented in Figure 1D. The method developed allows quantitative recovery of the underlying intensity levels and change points in a single-molecule measurement. The first step, identification of intensity change points, may be achieved using a generalized likelihood ratio test. Incoming photons will be Poisson distributed. As such, the probability that the observed data contain an intensity change point may be calculated and compared to the probability that the data contain no such change point. With this treatment, a confidence level for the presence of the change point is also determined, as well as a confidence region for the position of the change point. This methodology is applied to the trajectory as a whole. It is recursively segmented until no more change-points are detected. Thus, both the *number* and the *location* of intensity changes are determined.

To further refine the analysis, we next consider possible groupings of the detected states. The n change points define $n + 1$ different intensity states in the trajectory. Some of these apparently different intensity levels may arise from identical emissive states in the molecule under observation. This hypothesis is quantitatively assessed and the two intensities most likely to be identical are assigned to the same state. This procedure is repeatedly applied until all intensity levels are clustered into the same state, resulting in a hierarchy of clustered intensities. The true number of states in the system is then determined by application of the Bayes information criterion.

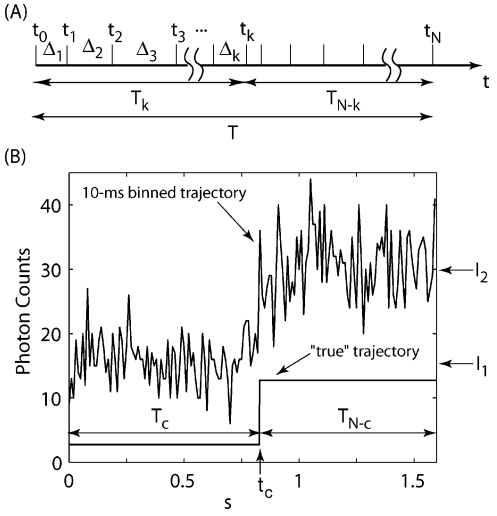

Figure 2. (A) A typical photon-by-photon recording scheme that marks the chronological photon arrival times $\{t_i\}$, the inter-photon time durations $\{\Delta_i\}$, emission intensity change points $\{c_k\}$, and the time durations $\{T_k\}$ for which the single molecule maintains the same emission intensity. (B) A simulated single-molecule trajectory (total 2372 photons) showing a sudden intensity change from $I_1 = 1.5$ kcps to $I_2 = 3.0$ kcps at $t_c = 0.8283$ s (the 1318-th photon). Poissonian counting noise is evident in the 10-ms binned trajectory. As a comparison, the "true" trajectory is also displayed but offset and scaled for clarity.

The result of this analysis is an accurate determination of the true number of states available to the system under study and the timing of transitions between them. This treatment is independent of externally imposed time scales and free of kinetic models. All of the change points are accompanied by a calculated confidence interval. The details of the above-outlined theoretical development are described in section 2. Its performance is evaluated using computer simulations and discussed in section 3.

2. Theory

Thorough analysis of a time-resolved single molecule trajectory with intensity change points involves two critical issues. First, the change points must be located and assigned significance levels and confidence intervals. Second, once the change points have been found, the total number of molecular states and the intensity of each state must be identified. These issues are nontrivial if they are to be resolved quantitatively, especially in high time resolution single-molecule measurements. We utilize ideas from statistical testing theory and from information theory to outline a systematic approach that addresses these issues.

2.1. Basic Considerations. Though the general ideas outlined in this article are applicable to most time-resolved single-molecule measurements, we focus our discussion on the use of photon-by-photon detection/registration methods since they provide the highest time resolution (also see Watkins and Yang[27] for a brief overview). Figure 2A illustrates the recording scheme of a typical photon-by-photon single-molecule experiment. If

the molecule is in a certain state j that exhibits a constant detected emission intensity I_j, photon arrival times will be Poisson distributed. For most detectors capable of registering single photon arrival times (such as avalanche photodiodes and photomultiplier tubes), dark counts will also be Poisson distributed and are thus naturally included in this treatment. The probability density function for recording an inter-photon duration Δ_i is then given by the exponential distribution

$$f(\Delta_i; I_j) = I_j \, e^{-I_j \Delta_i} \tag{1}$$

The change point c_j is defined as the time at which the emission property of the molecule in question changes from I_j to I_{j+1}. The probability of detecting n_j photons within a time period T_j between c_{j-1} and c_j is given by the Poisson function

$$g(n_j; I_j, T_j) = \frac{(I_j T_j)^{n_j} e^{-I_j T_j}}{n_j!} \tag{2}$$

The maximum likelihood estimate (MLE) of the intensity is $\hat{I}_j$ that maximizes the likelihood function eq 2 given n_j and T_j. It can be readily computed to be $\hat{I}_j = n_j/T_j$. Intuitively, the longer the observation (T_j) is made or the more photons (n_j) are collected, the better $\hat{I}_j$ can be measured. These intuitive ideas are quantitatively expressed using principles from information theory to compute the Fisher information[28] of I_j, the amount of information about I_j that is available from the given data set. This will allow estimation of the uncertainties associated with $\hat{I}_j$.

The Fisher information quantifies the knowledge of a physical parameter (I_j) that can be drawn from experimentally measured quantities (n_j). The distribution of experimentally measurable n_j is given by the likelihood function $g(n; I, T)$, the probability that the observed number of photons observed in a time interval T will be n, given that the molecule is in a state with detected intensity I. The Fisher information is given by

$$J_F(I) = \left\langle \left(\frac{\partial}{\partial I} \ln[g(n_j; I_j, T_j)] \right)^2 \right\rangle_{n_j}$$

where $\langle \cdots \rangle_{n_j}$ denotes the expectation value weighted by the likelihood $g(n_j; I_j, T_j)$ over all possible n_j. Uncertainties in measuring I may be expected to be related to the Fisher information since, intuitively, I can be determined more accurately if more information about I can be obtained. This qualitative understanding is quantitatively expressed by the Cramér−Rao−Fréchet bound[29−31]

$$\mathrm{var}(I) \geq \left(\frac{d\langle \mathrm{Est}(I) \rangle}{dI} \right)^2 J_F(I)^{-1}$$

where $\langle \mathrm{Est}(I) \rangle$ is the expectation value of I from the estimator Est. For an unbiased estimator, with $\langle \mathrm{Est}(I) \rangle = I$, the Cramér−Rao−Fréchet inequality states that the variance of the best possible estimator of I is given by the inverse of its Fisher information matrix. In general, the maximum likelihood estimator (MLE) is a good starting point because it is asymptotically normal (Gaussian) under most conditions. The Fisher information of the estimated fluorescence intensity $\hat{I}_j$ is

$$J_F(\hat{I}_j) = \langle (\partial g(n_j; I_j, T_j)/\partial I_j)^2 \rangle_{n_j}|_{I_j \to \hat{I}_j} = T_j/\hat{I}_j = T_j^2/n_j$$

The best attainable variance of I_j is then given by

$$\mathrm{var}(\hat{I}_j) \geq J_F(\hat{I}_j)^{-1} = \frac{n_j}{T_j^2} \tag{3}$$

Equation 3 allows a quantitative assessment of the uncertainties with which a single-molecule intensity time trajectory is measured. For example, application of eq 3 shows that only 100 photons must be collected to measure an intensity to within a relative standard deviation of 10%.

2.2. Detection of Fluorescence Intensity Change Points. Identifying the times or photon indices at which a single molecule changes its emission pattern can be viewed as a variant of the "change point" problem in statistics.[32] Here we treat the detection of intensity change points as a hypothesis test problem and develop a generalized likelihood ratio test that uses all of the available photon arrival time information.[33,34] This test is both powerful and straightforward to implement. The statistical analysis that follows only considers trajectories with single change points. However, an intensity trajectory may contain many change points. The recursive algorithm for detecting all of the change points in a trajectory was constructed with this in mind. Once approximate change point locations and critical regions have been identified, the calculations are repeated using intervals containing only one change point.

This treatment is only applicable to scenarios where the intensity trajectory is recorded in terms of individual photon arrival times. For experimental schemes that measure signals at a fixed integration time, the expressions derived by Boudjellaba et al.[35] may be used to find change points. We emphasize that, when applying this method to experimental data, the usual type of one-standard deviation (or 69% confidence interval) is not applicable because the underlying error distribution is not Gaussian. Nevertheless, we include in some of the figures plots that correspond to 31% error rate simply as a reference point.

2.2.1. Photon-by-Photon Generalized Likelihood Ratio Test. Suppose that, in a trajectory of N photons and time duration T, a molecule suddenly changes its state such that the detected emission intensity experiences a jump at time t_c (cf. Figure 2B). The presence of a change point is clearly indicated when the data is binned, but the exact timing of the change is not clear. To test for the existence of a change point at any proposed photon k, the trajectory is divided at k into segments of length T_k and T_{N-k}, containing k and $N - k$ photons, respectively. Two hypotheses, H_A and H_o, must then be considered. Hypothesis H_A states that there was a change point at time t_c. The likelihood L_A of hypothesis H_A is given by the Poisson distribution, eq 2

$$H_A: \; I(t_1) = \cdots = I(t_k) = I_1 \neq I(t_{k+1}) = \cdots = I(t_N) = I_2$$

$$L_A = g(k; \hat{I}_1, T_k) g(N - k; \hat{I}_2, T_{N-k}) = \frac{(\hat{I}_1 T_k)^k e^{-\hat{I}_1 T_k}}{k!} \frac{(\hat{I}_2 T_{N-k})^{N-k} e^{-\hat{I}_2 T_{N-k}}}{(N\text{-}k)!}$$

The intensities $\hat{I}_1$ and $\hat{I}_2$ are estimated from the trajectory as discussed in section 2.1, since they are not known a priori. Hypothesis H_A must be compared with the null hypothesis H_o, which states that there is no change point at time t_c

$$H_o: \; I(t_1) = I(t_2) = \cdots = I(t_N) = I_0$$

$$L_o = g(n_1; \hat{I}_0, T_k) g(n_2; \hat{I}_0, T_{N-k}) = \frac{(\hat{I}_0 T_k)^k e^{-\hat{I}_0 T_k}}{k!} \frac{(\hat{I}_0 T_{N-k})^{N-k} e^{-\hat{I}_0 T_{N-k}}}{(N - k)!}$$

where I_0 is the average intensity of the entire trajectory.

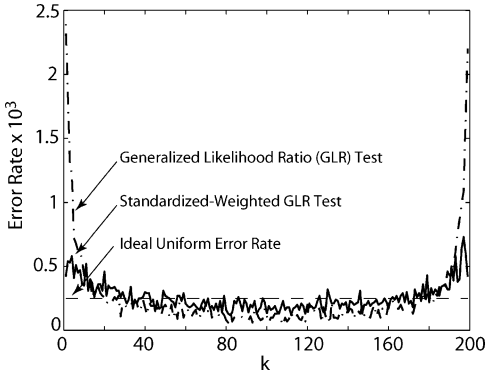

Figure 3. Error rates as a function of the test point k under the null H_o hypothesis in which there is no change point. In this simulation, a total of 100 000 traces of 200 exponentially distributed random number were generated and analyzed. The type-I error rate α is set to 0.05 in the analysis.

To compare the likelihood of H_A over H_o, we compute the log likelihood ratio of the two hypotheses

$$\mathscr{L}_k^\circ = \ln\frac{(\hat{I}_1 T_k)^k \, e^{-\hat{I}_1 T_k}(\hat{I}_2 T_{N-k})^{N-k} \, e^{-\hat{I}_2 T_{N-k}}}{(\hat{I}_0 T_k)^k \, e^{-\hat{I}_0 T_k}(\hat{I}_0 T_{N-k})^{N-k} \, e^{-\hat{I}_0 T_{N-k}}}$$

Upon substitution of the estimates $\hat{I}_1 = k/T_1$, $\hat{I}_2 = (N-k)/T_{N-k}$, and $I_0 = N/T$, this yields the simplified expression

$$\mathscr{L}_k^\circ = 2k\ln\frac{k}{V_k} + 2(N-k)\ln\frac{N-k}{1-V_k} - 2N\ln N \qquad (4)$$

where $V_k = T_k/T$. Note that this ratio is a function of the index k. This ratio is a measure of the likelihood that there was a change point at k. The most likely location of the change point is at the maximum of the log-likelihood ratio as a function of k

$$Z_N^\circ = \max_{1\le k\le N}\{\mathscr{L}_k^\circ\} \qquad (5)$$

This test, while identifying the most likely k for a change point, does not prove that a change point is present. To assess whether a change point occurred, the critical value $\tau_{1-\alpha}^\circ$ must be calculated. Here α is the probability of type I error (a false positive). When Z_N°, the maximum value of $\mathscr{L}_k$, is greater than $\tau_{1-\alpha}^\circ$, a change point is considered to occur with a probability α that this is a false positive. On the other hand, no change point occurred if $Z_N^\circ < \tau_{1-\alpha}^\circ$. The critical values $\tau_{1-\alpha}^\circ$ of Z_N° are not known analytically, but they can be calculated using a recursive algorithm due to Noé.[36] Computational details for the present application are described in Appendix A.

Equation 5 does not have a uniform error rate across the entire trajectory because MLE values of emission intensities are used in the likelihood ratio tests.[37,38] Very few photons are available for estimation of I_1 (I_2) when k is close to the beginning (end) of the data sequence. The uncertainty associated with such an estimate is therefore greater, resulting in higher error rates at both end points. This is illustrated by a simulation shown in Figure 3. The nonuniformity in error rate can be alleviated by standardization and weighting of the likelihood ratio function, as originally proposed by Henderson.[38] In this implementation, which still relies on the fundamental Poisson statistics discussed above, the null hypothesis is rejected when the test statistic becomes

$$Z_N = \max_{1\le k\le N}\{\bar{\mathscr{L}}_k + W_k\} \equiv \max_{1\le k\le N}\{\mathscr{L}_k\} \ge \tau_{1-\alpha} \qquad (6)$$

where $W_k = (1/2)\ln[4k(N-k)/k^2]$ is Henderson's weighting function and $\bar{\mathscr{L}}_k = (\mathscr{L}_k^\circ - E[\mathscr{L}_k^\circ])/\sigma_k$ is the standardized log likelihood ratio with $E[\mathscr{L}_k^\circ]$ and σ_k being the expectation value and standard deviation, respectively. Following Henderson

$$\mathscr{L}_k^\circ - E[\mathscr{L}_k^\circ] = -2k\ln V_k + 2k\mu_k - 2(N-k)\ln(1-V_k) + 2(N-k)\mu_{N-k}$$

where $\mu_k = E[\ln V_k] = -\sum_{j=k}^{n-1}(1/j)$ and $\mu_{N-k} = E[\ln(1-V_k)] = -\sum_{j=N-k}^{n-1}(1/j)$. The standard deviation is

$$\sigma^2 = 4k^2 v_k^{\,2} + 4(N-k)^2 v_{N-k}^{\,2} - 8k(N-k)\xi$$

where $\xi = \pi/6 - \sum_{j=1}^{N-1}(1/j^2)$, $v_k^2 = \sum_{j=k}^{n-1}(1/j^2)$, and $v_{N-k}^2 = \sum_{j=N-k}^{n-1}(1/j^2)$. The critical regions $\tau_{1-\alpha}$, defined by the null probability error rate

$$\Pr(\mathscr{L}_k < \tau_{1-\alpha};\, k = 1, ..., N|T) = 1-\alpha \qquad (7)$$

can also be computed using Noé's algorithm as outlined in appendix A and are listed in Table 1. Equation 6 gives a more uniform empirical error rate as shown in Figure 3. For this reason, eq 6 will be used as the test statistic for the remainder of this article. The critical region defined by $\tau_{1-\alpha}$ gives type I error with probability α. For instance, if $\alpha = 0.05$, the change point selected by the likelihood ratio test eq 5 has a 5% probability of being a false change point. A preliminary evaluation of the accuracy of the likelihood ratio test in selecting the change point is presented in Figure 4.

Type I error alone does not completely characterize intensity change point detection. Nothing is said about the probability of missing a change point or about the accuracy with which the intensity change point is located. The probability of missing a change point is related to the power of a test. For instance, if the power of a test is 0.9, then the probability of missing a change point is only 10%. The power of change-point detection in our implementation was characterized by simulation of 200-photon trajectories with one change point occurring at $c = 100$. The results are summarized in Figure 5. Our simulation indicates that for a sample size of 200 data points with one change point, at least 95% of the intensity jumps with size $I_2/I_1 \simeq 2$ will be detected. Furthermore, the τ_{99} curve indicates that as the jump size increases to $I_2/I_1 > 2.5$ nearly 100% of the intensity jumps are detectable, with a very small 1% probability of assigning an erroneous change point. This is reasonable. A change point from I_1 to I_2 should be easier to detect if the relative size of the intensity jump is greater.

It is also possible to determine the confidence interval for a change point from the data. The change point was found by determining where in the trajectory the null hypothesis could be rejected. The confidence region $C_{1-\beta}$ is defined as the region around the change point where the change point hypothesis H_A cannot be rejected with probability β.[34] That is, every photon for which hypothesis H_A is more than 31% likely to be true is within the 69% confidence region C_{69}

$$C_{1-\beta} = \{k:\ \Pr(\mathscr{L}_m < \tau'_{1-\beta}|T_k)\Pr(\mathscr{L}_{m'} < \tau'_{1-\beta}|T-T_k) \le 1-\beta\} \qquad (8)$$

where $m \in \{1...k\}$ and $m' \in \{k+1...N\}$. In principle, the confidence region can be found by direct evaluation of the above equation. The computational cost for this approach, however, is too high to be practical. Worsley[37] conducted numerical simulations to evaluate eq 8 for exponential distributions over all k, T_k, and T at fixed N and found that the maximum occurred

TABLE 1: Level α and *β* Points for Intensity Change Detection Using Eqs 6 and 9

N	τ_{69}	τ'_{69}	τ_{90}	τ'_{90}	τ_{95}	τ'_{95}	τ_{99}	τ'_{99}
10	6.266	5.710	4.075	3.539	3.112	2.602	1.462	1.052
20	6.903	6.434	4.625	4.191	3.620	3.214	1.881	1.561
30	7.208	6.791	4.890	4.511	3.866	3.512	2.087	1.809
40	7.398	7.017	5.057	4.713	4.021	3.701	2.218	1.966
50	7.533	7.179	5.176	4.857	4.131	3.836	2.311	2.079
60	7.635	7.303	5.266	4.968	4.216	3.939	2.383	2.165
70	7.717	7.402	5.338	5.056	4.283	4.021	2.441	2.234
80	7.784	7.484	5.397	5.128	4.339	4.089	2.488	2.291
90	7.841	7.553	5.448	5.190	4.386	4.147	2.529	2.339
100	7.889	7.612	5.491	5.243	4.426	4.196	2.564	2.381
250	8.246	8.049	5.810	5.634	4.726	4.562	2.823	2.691
500	8.451	8.300	5.996	5.859	4.902	4.774	2.976	2.872
750	8.551	8.422	6.086	5.969	4.988	4.878	3.052	2.961
1000	8.614	8.498	6.144	6.039	5.043	4.944	3.100	3.018

at $k = 1$ and $T_1 \simeq T/N$. This observation allows the calculation of a conservative $1 - \beta$ confidence region around a change point c by

$$\tilde{C}_{1-\beta} = \{k:\ Z_N - \mathscr{L}_k \leq \tau'_{1-\beta}\} \tag{9}$$

That is, plotting $\mathscr{L}_k$ as a function of k, the β confidence region consists of photon indices $\{k\}$ that give $\mathscr{L}_k$ values that are greater than $Z_N - \tau'_{1-\beta}$. Calculation of $\tau'_{1-\beta}$ (cf. Table 1) is also carried out using Noé's algorithm, detailed in Appendix A.

To characterize the performance of the approximation, eq 9, we have used computer simulations to evaluate its accuracy. The results, summarized in Figure 6, suggest that for high-power change point detections (power > 0.9), the critical region defined by eq 9 is indeed a conservative overestimate of the true confidence interval. For lower-power change point detection, however, eq 9 underestimates the confidence interval.

Equations 6 and 9 allow us to find the single-molecule intensity change points by analyzing the available data photon by photon and compute the statistical significance of the change points. They are summarized in Figure 7 using the time trace in Figure 2B as an example. As illustrated in Figure 7, our detection scheme finds the correct intensity transition point to within 9 photons (or 2.5 ms) for this particular data set. Moreover, it allows the assignment of statistical significance to the change point, allowing the experimentalist to *quantitatively* assess the validity of the analysis.

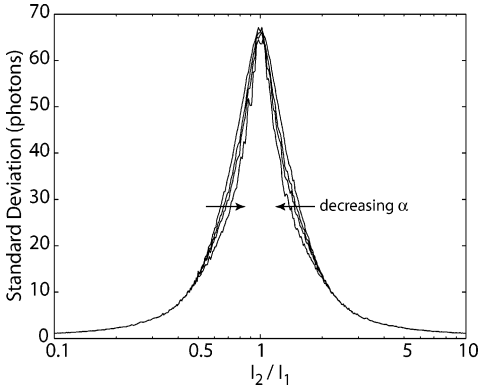

Figure 4. Empirically determined standard deviation of detected intensity change point as a function of the size of intensity change under the alternative H_A hypothesis in which there is exactly one change point. In this simulation, a total of 100 000 traces of 200 exponentially distributed random number were generated and analyzed with the change point occurring at the 100-th point. Four different type-I error rates were used in this analysis: $\alpha = 0.31$, 0.1, 0.05, and 0.01. The standard deviation is also independent of the direction of an intensity jump.

2.2.2. Implementation. In general, a single-molecule time trajectory may contain multiple intensity change points. To find all of the significant change points, we use a recursive binary segmentation algorithm. First, a change point is found by applying eq 5 to the entire single-molecule time trajectory. The maximum of the log-likelihood ratio, eq 6, is considered as a change point and its confidence interval is found using eq 9. The left (right) confidence bound is then held as one end point, with the start (end) of the trajectory serving as the other end point. The search for change points is continued in the left (right) daughter set (cf. Figure 11A). This procedure is repeated recursively until no further change points are found.

Equations 6 and 9 assume that only one change point is present between the two end points. Therefore, a new interval is defined for each change point, using the previous and subsequent change points as end points. The location, and confidence interval of each change point are then recalculated using this interval. Spurious change points that do not meet the selection criteria are eliminated. Change points thus determined and their statistical properties are stored for the next stage of calculations.

The trajectory shown in Figure 1 containing multiple change points is used to illustrate the application of the binary segmentation algorithm. Once the change points are located,

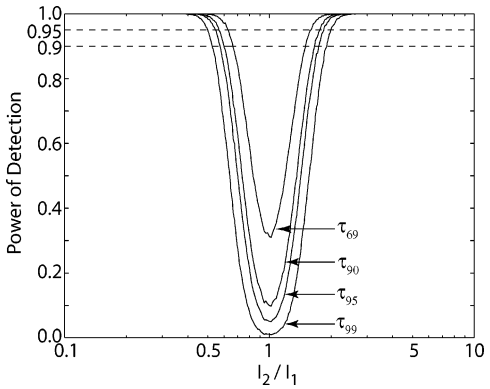

Figure 5. Detection power as a function of the size of intensity change under the alternative H_A hypothesis in which there is exactly one change point. In this simulation, a total of 100 000 traces of 200 exponentially distributed random number were generated and analyzed with the change point occurring at the 100-th point. The critical regions for type-I error used in the analysis are indicated by $\tau_{1-\alpha}$, where $1 - \alpha = 0.69$, 0.90, 0.95, and 0.99. 90% and 95% detection power are indicated by dashed lines. For example, to achieve 90% detection power, the minimum I_2/I_1 ratios are ~1.5 for $\alpha = 0.31$, ~1.66 for $\alpha = 0.1$, ~1.72 for $\alpha = 0.05$, and ~1.9 for $\alpha = 0{:}01$. As expected, the detection power $= 1 - \alpha$ at $I_2/I_1 = 1$. Furthermore, the detection power is independent of the direction of an intensity jump. That is, power(I_1/I_2) = power-(I_2/I_1).

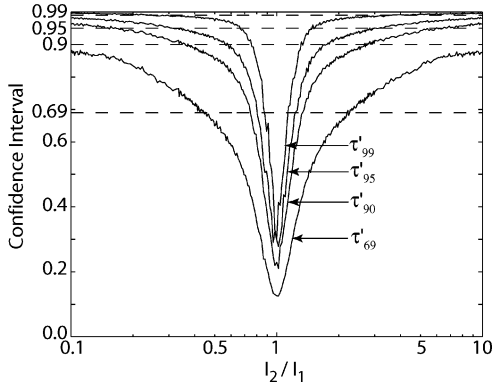

Figure 6. Empirical evaluation of the accuracy of confidence intervals calculated from eq 9. The solid curves represent the probability that the detected change point lies within the calculated confidence interval. In this simulation, a total of 100,000 traces were generated of 200 exponentially distributed random numbers with the change point occurring at the 100th point. An $\alpha = 0.05$ critical region for type I error used in the analysis. For reference, horizontal dashed lines are drawn at the true confidence levels 0.69, 0.90, 0.95, and 0.99. The confidence region is also independent of the direction of an intensity jump.

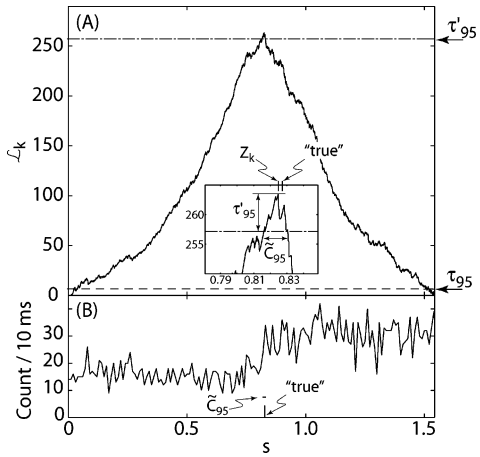

Figure 7. (A) Plot of $\mathcal{Z}_k$ as a function of the chronological photon arrival time of each detected photon using the data from Figure (2). The critical region was found to be $\tau_{95} = 6.343$ (− −) given $\alpha = 0.05$. The change point was found to be $c = 1301$ ($t_c = 0.8247$ s), compared to the true location of $c^{\text{true}} = 1310$ ($t_{\text{true}} = 0.8272$ s). The detection power was found to be 1.00. (B) Reproduction of Figure 2B for comparison. The "true" location of intensity jump and inferred 95% confidence interval $\tilde{C}_{1-\beta}$ are indicated as vertical and horizontal bars, respectively.

the intensity levels can be determined using the maximum likelihood estimator ($\hat{I}_j = n_j/T_j$) discussed in section 2.1. Figure 8 displays a reconstructed intensity trajectory for the data shown in Figure 1. A total of 125 statistically significant change points are found to give 126 different apparent intensity levels. The number of intensity states that give rise to the observed intensity levels still remains to be found, as well as accurate values for the emission intensities of those states.

2.3. Determination of the Number of States and Intensity Levels. A popular approach to determining the number of emission states and intensity levels is to build a histogram of the intensities within a single molecule emission time trajectory. This approach may work for systems that contain 2 or 3 emission states whose intensities are well separated, but it loses value if more states are involved or if the intensity levels are not well separated, as in the example of Figure 1. The above-discussed photon-by-photon change point determination greatly

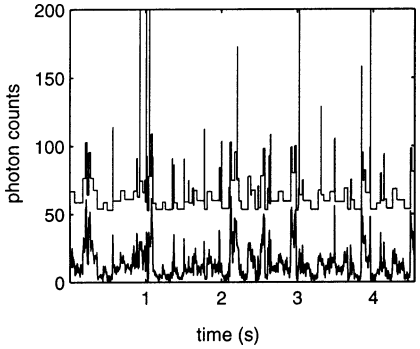

Figure 8. Single-molecule emission time trace (offset by 50 and scaled to a 5-ms bin time) inferred from the simulated data in Figure 1 using generalized likelihood ratio test. The 5-ms binned raw data are also shown as a comparison. The intensity bursts seen in the inferred emission time trace are examples of "shot" counting noise arising from short resident time at a given state.

simplifies the task, but difficulties remain. The Poisson statistics of photon detection and the unknown underlying dynamics that govern the emission states make it difficult to ascertain the number of molecular states and their corresponding intensity levels. For example, sudden intensity bursts may arise from states of short residence time as seen in Figure 8; they will need to be treated in a statistically robust way. Here we develop a model-based clustering analysis method that simultaneously determines the number of molecular states and their emission levels. Following the ideas of Fraley and Raftery,[39] photon-by-photon determined intensity levels are classified into groups using an agglomerative hierarchical clustering algorithm. This clustering algorithm is highly sensitive to initial conditions, so the results can only serve as an initial guess for more advanced analysis. The results are further refined using an expectation-maximization procedure in which each intensity level I_j is assigned a weighting coefficient p_{mj} that describes the probability that it belongs to the mth state out of a total of n states. These intensities and probabilities are calculated for all possible numbers of states, ranging from one state to the total number of detected change points. Finally, we use the Bayesian Information Criterion to quantitatively determine the minimum number of states required to fit describe the data.

2.3.1. Initial Classification by Agglomerative Hierarchical Grouping. Consider an exponential mixture model that allows classification of single-molecule observations, photon by photon, into different intensity levels. The likelihood function for the entire single-molecule time trace is

$$L(\{\Delta_i\}; I_1, ..., I_G, \gamma_1, ..., \gamma_N) = \prod_{i=1}^{N} f_{\gamma_i}(\Delta_i; I_{\gamma_i})$$

where G is the number of distinguishable intensity levels and γ_i is the classification, or grouping, of the ith photon into the γ_ith intensity group. The change-point detection procedure discussed in section 2.2 provides a means with which the initial photon-by-photon intensity classification is accomplished. For J detected intensity change points, this initial grouping yields $(J + 1)$ possible intensity levels $\{\hat{I}_j\}$ with variance $\{n_j/T_j^2\}$, $j = 1...J + 1$ (cf. eq 3). In this classification scheme, the data set $\{\Delta_i\}$ is formed by sampling n_j observations (photons) separately from the jth intensity component. Under this scheme, the classification log-likelihood function takes the form

$$\mathcal{L}_c(\{\Delta_i\}; I_1, ..., I_G) = \sum_{j=1}^{G} \sum_{i \in J_j} \ln f(\Delta_i; I_j) \qquad (10)$$

where G is the number of groups and J_j is the set of photon indices that belong to the j-th intensity level. Based on this likelihood expression, the G intensity levels can be classified into G_{max} groups via an agglomerative hierarchical algorithm first proposed by Ward.[40]

At each stage of the agglomerative hierarchical clustering procedure, the number of groups contracts from G to $(G - 1)$ by merging two groups of similar properties. No conclusions are drawn about the actual number of states in the system. The only result of this procedure is a hierarchical list of the intensity levels that are most likely to have arisen from the same emissive state of the experimental system. Many approaches to classification and clustering have been proposed and studied as recently reviewed by Fraley and Raftery.[41] The established formulations, however, are not directly applicable to the analysis of single-molecule data.

To treat the problem in a manner consistent with that of change-point detection (cf. section 2.2), we consider a merit function based on maximizing the likelihood ratio of grouping. Our treatment closely follows that of Scott and Symons who derived likelihood ratio criteria for Gaussian models.[42] Here the exponential distribution of eq 1 is used to model photon-by-photon detection. Let the merit function $M_{\langle j,m\rangle}$ be the likelihood ratio of contracting from G groups to $(G - 1)$ groups by merging the mth and the jth groups. Using eq 10, the log-likelihood ratio merit function is

$$M_{\langle j,m\rangle} = \mathscr{L}_c(\{\Delta_i\}; I_1, ..., I_G) - \mathscr{L}_c(\{\Delta_i\}; I_1, ..., I_{\langle j,m\rangle}, ..., I_{G-1})$$
$$= (n_m + n_j)\ln\left[\frac{n_m + n_j}{T_m + T_j}\right] - n_m\ln\left[\frac{n_m}{T_m}\right] - n_j\ln\left[\frac{n_j}{T_j}\right] \quad (11)$$

Thus, for each stage a merit matrix is built, composed of the elements $M_{\langle j,m\rangle}$. The two groups to be merged are the j and m that give the maximum $M_{\langle j,m\rangle}$. The number of photons n_{jm} and time duration T_{jm} of the newly merged group are updated according to

$$n_{jm} = n_j + n_m \quad (12)$$

$$T_{jm} = T_j + T_m \quad (13)$$

Equations 11−13 form the basis for fast computer algorithms.[43] Initially, all of the $G = J + 1$ intensity levels are treated as distinct groups. The hierarchical clustering gives a picture of the grouping of intensity levels as a function of the total number of independent intensity states, recording the intensity-level partitionings for all $G \leq G_{max}$ for further refinement by EM clustering described below.

2.3.2. Refinement by Expectation−Maximization Clustering. As previously mentioned, the agglomerative hierarchical clustering procedure performed above is very sensitive to initial conditions, but it serves as an initial guess for more advanced clustering schemes. Further improvement of the partitioning of the $\{I_1, ..., I_{J+1}\}$ intensity levels into $G = 1...G_{max}$ groups, as well as the estimation of $\{\hat{I}_1, ..., \hat{I}_G\}$, is accomplished by the expectation−maximization procedure of Dempster, Laird, and Rubin.[44] This procedure assigns a probability for a certain time interval, defined by the change points previously located, to belong to a certain group. The group of maximum probability is considered to be the correct group for that particular time interval.

Let p_{mj} be the probability of assigning I_j to the mth group. Then the initial guess for the expectation-maximization procedure is based on the result of the agglomerative hierarchical grouping

$$p_{mj} = \begin{cases} 1 & \text{if } I_j \text{ belongs to the } m\text{th group} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

The probability density of observing I_j given p_{mj} is $\prod_{m=1}^{G} g(n_j; I_m, T_j)^{p_{mj}}$. Assuming that $\{p_{mj}\}$ are drawn from independently and identically distributed multinomial distributions, the complete likelihood function for the observation of $\{I_1, ..., I_{J+1}\}$ is

$$L(I_j; p_{mj}, p_m, I_m, T_m) = \prod_{j=1}^{J+1}\prod_{m=1}^{G}[p_m g(n_j; I_m, T_j)]^{p_{mj}}$$

where p_m is the probability of drawing an I_j from the mth intensity level, n_j and T_j are respectively the number of photons and time duration between the $(j - 1)$th and the jth change points. The log-likelihood function to be maximized is

$$\mathscr{L}_{em}(I_j; p_{mj}, p_m, I_m, T_m) = \sum_{j=1}^{J+1}\sum_{m=1}^{G} p_{mj}\ln[p_m g(n_j; I_m, T_j)] \quad (15)$$

For the model in eq 15, p_{mj} is calculated as $\bar{p}_{mj} = E[p_{mj}|\{\hat{I}_j\}; T_m, I_m]$ and can be understood as the conditional expectation value given the observation $\{\hat{I}_j\}$ and the associated parameter values T_m and I_m. The jth intensity segment may come from one of the $m = 1$ to G intensity states. Under the multinomial framework, the probability for the jth segment to come from the mth intensity state is then as displayed in Figure 9.

Initialize $\bar{p}_{mj}$ according to Eq. (14).
repeat

M-step: compute parameter MLEs given $\bar{p}_{mj}$

- $\hat{T}_m \leftarrow \sum_{j=1}^{J+1} \bar{p}_{mj} T_j$
- $\hat{n}_m \leftarrow \sum_{j=1}^{J+1} \bar{p}_{mj} n_j$
- $\hat{p}_m \leftarrow \sum_{j=1}^{J+1} \hat{T}_j / T$
- $\hat{I}_m \leftarrow \sum_{j=1}^{J+1} \bar{p}_{mj} n_j / \hat{T}_m$

E-step: compute $\bar{p}_{mj}$ given the M-step MLEs

$$\bar{p}_{mj} \leftarrow \frac{\hat{p}_m g\left(n_j; \hat{I}_m, \hat{T}_m\right)}{\sum_{m=1}^{G} \hat{p}_m g\left(n_j; \hat{I}_m, \hat{T}_m\right)}$$

until $\bar{p}_{mj}$ is converged.

Figure 9. EM algorithm for single-molecule intensity level clustering.

The EM procedure iterates between the E step, where the expectation values $\{\bar{p}_{mj}\}$ are computed from the data using estimated parameters, and the M step, where the parameters are estimated by maximizing the likelihood function eq 15. An EM algorithm for clustering single-molecule intensity levels is outlined in Figure 9.

2.3.3. Determination of Number of States Using the Bayesian Information Criterion (Schwarz's Criterion). Although the grouping of intensity levels has been optimized using the EM algorithm, the number of molecular states remains unknown. Generally, one is interested in the minimal set of parameters required to describe the data. Unconstrained maximization of the likelihood function usually results in models that contain more groups (one of the parameters in this treatment) than the true number, over fitting the data. This is clearly undesirable. To quantitatively assess the minimum number of parameters

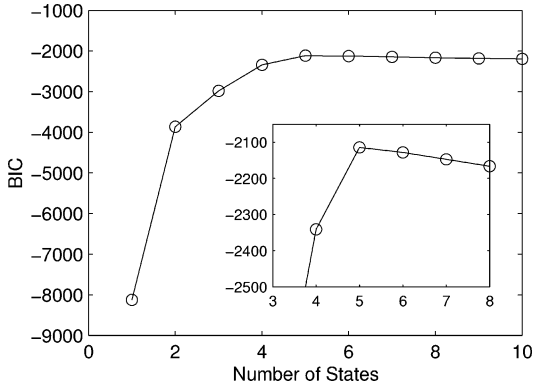

Figure 10. BIC as a function of number of groups, showing that there are most likely 5 distinct intensity states. The numerical example in Figure 1 is used.

required to accurately fit the data, we use a Bayesian approach. This method has deep connections to other information theoretical concepts such as minimum description length and Kolmogorov complexity.[45] Casting these ideas in a framework consistent with the above-discussed likelihood ratio test results in a penalized maximum likelihood estimator. The maximum likelihood estimator is penalized by an amount proportional to the information added when the number of adjustable parameters is increased. This approach has been shown to predict a number of states that is at least as large as the correct one.[46] Furthermore, as the number of data point increases, it asymptotically converges to the correct number of states. Formally, the integral to be maximized is

$$p(\{\Delta_i\}, m) = p_m(m)\int_{\Theta_{em}}\int_{\Theta_{cp}} p_l(\{\Delta_i\}|\theta_{cp}, \theta_{em}, m)p_p \times (\theta_{cp}|\theta_{em}, m)p_p(\theta_{em}|m)\, d\theta_{em}\, d\theta_{cp}$$

Here Θ_{cp} and Θ_{em} represent the parameter spaces for the number of change points and the number of distinct intensity levels, respectively, θ_{cp} and θ_{em} are specific realizations of those parameters, and the integrals are performed over the entire parameter spaces, which may change as a function of the model m (the number of intensity levels). The integrals may be evaluated for large N and N_{cp}, first over Θ_{cp} and then over Θ_{em}, using an approximation due to Schwarz,[45,47] giving

$$\ln p(\{\Delta_i\}, m) \approx 2\mathcal{L}_G - (2n_G - 1)\ln N_{cp} - N_{cp}\ln N \quad (16)$$

where N_{cp} is the number of change points detected, n_G is the number of groups, and $\mathcal{L}_G$ is the log-likelihood given in eq 15. Equation 16, often called the Bayesian information criterion (BIC), is a measure of how much experimental evidence favors the model. The first and second terms are clearly dependent on the model chosen. The third term is also dependent on the model because the choice of a lower number of intensity states will reduce the number of detected change points, as consecutive states are assigned to identical intensity levels. Using the numerical data in Figure 1 as an example, Figure 10 shows BIC values plotted as a function of the number of groups G. There is a maximum at $G = 5$, correctly selecting the number of groups. This, as well as the computational studies detailed below, indicate that both the number of change points and the number of photons are large enough to satisfy the requirement of large N for the Schwarz approximation.

Once the number of intensity levels is determined using BIC, the time evolution of a single-molecule intensity trajectory can be reconstructed photon by photon. Shown in Figure 1D is a reconstructed single-molecule trace using the numerical data in Figure 1. The distribution of molecular states shown to the right of the trajectory is quantitatively reproduced. The reconstructed trajectory also reproduced features seen in the true trace in Figure 1E. As clearly demonstrated in Figure 1, our method based on quantitative statistical analysis is superior to the heuristic, but popular, binning-thresholding approach.

3. Discussion

3.1. Resolving Power of Change Point Determination. As implied in Figures 5 and 6, the accurate inference of an intensity change point critically depends on the quality of data. For example, if the magnitude of an intensity change $|I_2 - I_1|$ is small, a change point is less likely to be detected. Furthermore, if an intensity change point occurs very close to the preceding one, say 10 photons away, it is less likely to be resolved from the previous one. Broadly, this is because such an intensity change is difficult to distinguish from Poisson counting noise. This can also be understood in terms of the information available to estimate the true intensities $\hat{I}_1$ and $\hat{I}_2$, as discussed in section 2.1. That is, the accuracy with which I_1 and I_2 can be estimated in applying the generalized likelihood ratio test depends on the number of photons available.

To further characterize our method, we define the resolving power as the probability of detecting two change points which are well enough separated that their one standard deviation confidence intervals do not intersect. For the binary segmentation algorithm used in our method, the resolving power may be computed as the probability of detecting a change point with a 69% confidence interval that does not include the end points in a one-change point trajectory (cf. Figure 11A).

Intuitively, one would expect that more photons (greater k and n) will be needed to resolve change points with small intensity jumps, and only few photons will be needed to resolve large intensity jumps. In other words, very fast state-switching dynamics can be resolved if their emission characteristics are markedly different. To put these ideas in a more quantitative framework, we compute the resolving power using computer simulations. Representative results are displayed in Figure 11B−E. As an example, for very small intensity jumps ($I_2/I_1 = 1.5$, Figure 11B) in a 400-photon segment, an intensity change point must be at least 114 photons away from either end point to be detected with a probability greater than 90%. On the other hand, for greater intensity jump steps say, $I_2/I_1 = 5$ in Figure 11E, only a total of 20 photons are needed to resolve change points that are 10 photons apart with a >90% certainty. Therefore, the accuracy with which change points are located depends on the quality of data and on the dynamics that underlie the time-dependent emission characteristics.

3.2. Performance of Our Method. To assess the performance of our method, simulations were performed as follows. Intensity levels were assigned based on a uniform distribution between 500 and 10 000 photons per second, with a ratio of at least 1.5 between successive intensity levels. The duration of each state was constrained to be at least enough to achieve 95% resolving power. This constraint is reasonable since, for example, given two states with an intensity ratio of 5.0, any transitions between them would be separated by only 13 photons. This number increases with decreasing intensity ratio. In the most extreme case, two states with an intensity ratio of 1.5 require that their transitions to be separated by 200 photons. Based on these conditions, a photon by photon intensity trajectory was produced by generation of inter photon timings based on the exponential distribution $t \sim \exp\{-It\}$. Trajectories were generated with 2−7

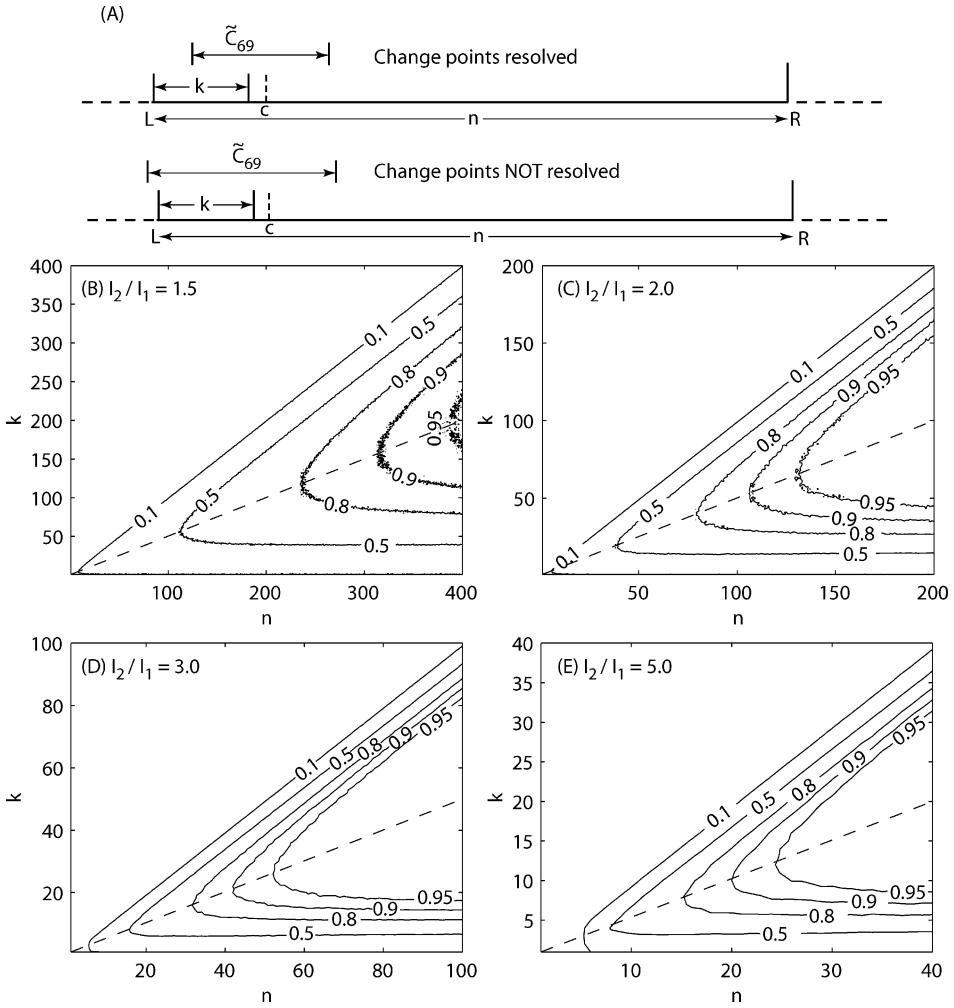

Figure 11. Characterization of resolving power by simulation. (A) Illustration of resolving power with the use of the binary segmentation algorithm. L and R are respectively the left and right end points of a data segment of length n which contains a true change point k photons from the left end. A detected change point c is considered resolved if the conservative 69% confidence region $\tilde{C}_{69}$ does not include L or R. (B)−(E) Contour diagrams of resolving power as functions of length k and total number of photons n at various intensity jump magnitudes. The dashed lines indicate cases when the change point occurs at the middle of a data segment. In these simulations, a data segment composed of k (with parameter $I_1 = 1$) and $n - k$ (with parameter I_2 calculated for each case B−E) exponentially distributed random numbers are generated. The n point segment is then subjected to analysis using our method at a type-I error rate $\alpha = 0.1$. For each (n, k) pair, the occurrence of a successful change-point detection and resolution are accumulated. The resolving powers are computed by dividing the occurrences by the total number of simulations, 10,000, for each (n, k) pair.

states and 50, 100, 250, or 500 change points. A total of 1000 simulations were run for each configuration, and each trajectory was analyzed by the above method, using a type I error rate of 0.05.

To make a useful comparison, the bias in measurement of parameters (intensity and occupancy) from the simulated trajectories was scaled by the true value, $\text{bias}(p_m) = (p_t - p_m)/p_t$, where p_t and p_m are the true and measured parameters, respectively. Bias in intensity and occupancy measurements are shown in Figure 12. The bias is consistently positive for the highest state in the trajectory and negative for the lowest state in the trajectory. Bias for the intermediate states falls in order between these two extremes. This is due to the stochastic nature of photon counting data. Although the mean intensity of a state is predetermined, the mean intensity of a particular instance of that state will vary. The magnitude of the intensity change between states will also vary, and, as shown earlier, larger changes are more likely to be detected. Indeed, the change point algorithm is intended to maximize differences in intensities between the detected states. This introduces a natural extremism in estimation of intensity levels, causing the bias observed.

The absolute value of the relative bias is very low (less than 3%) for trajectories with two states, even when the trajectory only includes 50 change points. With more states in a trajectory with the same number of change points, there will be fewer instances of any particular state, thus reducing the sample size and the quality of the statistics. This is reflected in the bias measurements. As the length of the trajectory increases, the bias decreases. In fact, given 500 change points, even a six state trajectory has a relative bias less than 5% in intensity measurements and less than 9% in occupancy measurements. These errors represent a vast improvement over binning and thresholding. Finding change points in a six state trajectory, let alone accurate determination of the intensities of the various states, would be completely impractical with a binning and thresholding scheme.

Results from BIC determination of the number of states are shown in Figure 13. As expected, the BIC estimate predicts at least the true number of states, predicting the correct number of states in the majority of cases. Occasionally, the BIC predicts fewer states than the correct number. This is because the conditions for the proof by Leroux[46] are not completely satisfied

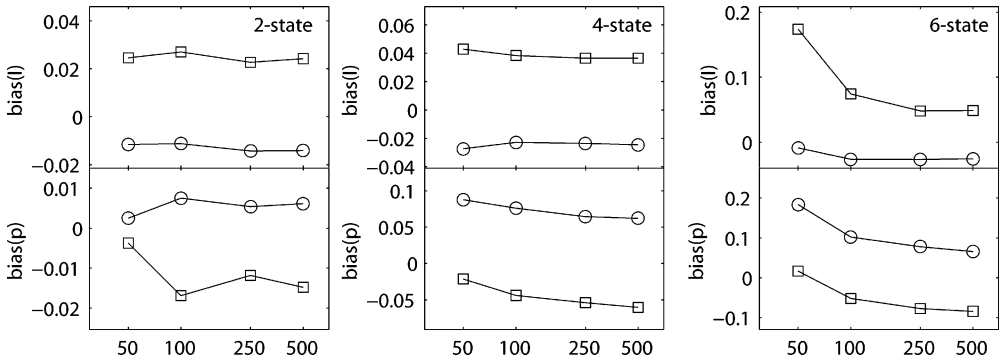

Figure 12. Relative bias in estimation of intensity (I) and occupancy (p) for simulations involving 2, 4, and 6 states as discussed in the text. Traces are shown for the states of highest (□) and lowest (○) intensity in each set, representing the extremes in bias.

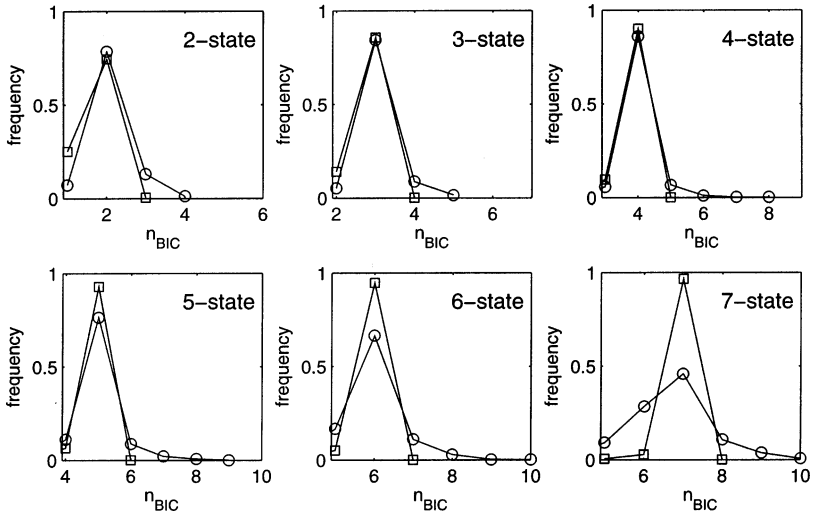

Figure 13. Histogram of BIC predictions of numbers of states. Each panel summarized results from simulations with a set number of intensity states. The true number of states is found in the top-right corner of each plot. Predictions of the number of states were made on the basis of trajectories including 50 change points (○) and 500 change points (□). Results for 100 and 250 change points were omitted in the interest of clarity and are included in the supplemental information.

by the change point analysis detailed above, mainly due to the uncertainties involved in change-point determination. Also, the two-step analysis involved in change point detection means that there is an inconsistency in computation of the BIC for one state as opposed to its computation for more than one state. This is the cause of the relative inaccuracy in applying this procedure to a two state system. Our simulation studies also suggest that, in applying to experimental data, it is important to examine BIC values over many single molecules and report on the most likely case. Despite these caveats, the BIC estimate is remarkably effective. The accuracy of the estimate increases consistently as the number of change points increases. Furthermore, estimation of the number of intensity levels exceeds 90% accuracy in every case but one when given 500 change points. The BIC procedure is thus a powerful method for accurately determining the true number of states in a system.

4. Concluding Remarks

Many dynamic systems exhibit intermediate states that are undetectable on the bulk level. These states and the transitions between them are essential to a fundamental understanding of the physical principles underlying these systems. Single-molecule spectroscopy can be an effective tool for the detection of these states; the ensemble average is undone and the existence of these intermediates is made clear and compelling. However, single-molecule measurements are made at the limits of optical detection, relying on very few photons to draw conclusions. Thus the possible advantages stemming from such experiments are offset by the ubiquitous Poisson photon counting noise, often making the result confusing and vague.

Our method provides a powerful and well characterized procedure for the location of intensity change points in a time-resolved trajectory. The Poisson statistics of photon detection are an integral part of our data treatment, so there is no ambiguity in deciding whether a transition occurs. Thus the subjectivity often seen in thresholding approaches is removed. Due to the photon-by-photon nature of the tests used, no extraneous time scales are introduced and the data quality is not arbitrarily reduced. It is therefore possible to achieve very high time resolution using our method. Empirical binning of photon counting trajectories is thus rendered superfluous. In addition, no kinetic models are imposed. Thus the experimentalist may deduce from the data which models are justified. Each intensity change point is provided with a significance level and a confidence interval, so derivative parameters of physical import can be determined quantitatively. Furthermore, this approach generates a sequential analysis of the entire trajectory, allowing a detailed treatment of nonstationary cases. This makes our method complementary to correlation function approaches,[48−50] which often require the assumption of stationarity.

A direct application of our change point detection scheme is the inference of the number and intensity of emissive states present in a single molecule trajectory [A software package inplementing this functionality can be found on our group

website, http://picasso.cchem.berkeley.edu/software.html]. Our study shows that the correct emission states can be *quantitatively* recovered from a single molecule trajectory. Finally, we emphasize the generality of these ideas. Our new approach provides a framework for photon by photon analysis of change points that can be extended to the treatment of any signal, two important examples being wavelength and polarization. Once the statistical distributions underlying the data are known, the tests discussed above are directly applicable. It is hoped that with this new method, which dramatically increases the amount of information that can be extracted from a single molecule trajectory, new and unexpected phenomena can be observed and studied.

Acknowledgment. This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy contract No. DE-AC03-76SF00098, by the University of California at Berkeley, and by the National Science Foundation. L.P.W. acknowledges a graduate fellowship from the National Science Foundation.

Supporting Information Available: A complete listing of numerical results for the critical regions up to $n = 1000$. This material is available free of charge via the Internet at http://pubs.acs.org.

Appendix A. Computation of Critical Regions

Under the H_0 hypothesis in which the single-molecule emission intensity does not change for the N sequentially recorded photons with an exponentially distributed inter-photon duration $\{\Delta_i\}$, the probability that eq 5 picks up a change point is $\Pr(\mathscr{T}_k^\circ \geq \tau_{1-\alpha}^\circ(N);\ k = 1, ..., N-1|T) = \alpha$, where α is the type I error rate and $\tau_{1-\alpha}^\circ(N)$ is the critical region that depends on both N and α. Worsley has shown that this probability is equivalent to having N ordered random variables $V_1^N \leq V_2^N \leq ... \leq V_{N-1}^N$ in which each $V_k^N \equiv T_k/T$, as defined earlier, is enclosed by its lower a_k and upper b_k bounds[34]

$$\Pr(\mathscr{T}_k^\circ \leq \tau_{1-\alpha}^\circ;\ k = 1, ..., N-1|T) = \Pr(a_k \leq V_k \leq b_k;\ k = 1, ..., N-1) \quad (17)$$

Both a_k and b_k are functions of $\tau_{1-\alpha}^\circ$, with k running up to $N-1$ since k denotes the location of a change point. Noé's algorithm was originally developed to compute the distribution of Kolmogorov-Smirnov type order statistics,[36] but if a_k and b_k can be found, it can be easily modified to calculate such probabilities as eq 17. Owen has described an efficient way to implement Noé's algorithm.[51] In our implementation, a_k and b_k were found by numerically solving $\mathscr{T}_k^\circ = \tau_{1-\alpha}^\circ(N)$ (cf. eq 4) at given α and N using the *zbrent* subroutine from Numerical Recipes in C modified for use with floating point operations that are 80 bit or greater.[52] Both lower and upper bounds for the solution are required for *zbrent*. In the present case, a_k is bound by 0 and k/N, and b_k is bound by k/N and 1.

Once a_k and b_k are determined eq 17 can be computed using Noé's algorithm. The critical region $\tau_{1-\alpha}^\circ(N)$ then can be found by numerically solving eq 17 using the modified *zbrent* subroutine. Since $\tau_{1-\alpha}^\circ(N)$ is a monotonically increasing function of N, we set the lower bound for $\tau_{1-\alpha}^\circ(N)$ to $\tau_{1-\alpha}^\circ(N-1)$ for use in *zbrent*. For $N = 2$, the lower bound is found by trial and error.

For the upper bounds, we use the asymptotic expression for the critical region for generalized likelihood ratio test due to Gombay and Horváth,[53] who found that the limiting distribution of $Z_N^{1/2}$

$$\lim_{N\to\infty} \Pr\{Z_N^{1/2} > \sqrt{\tau_{1-\alpha}^\infty}\} = \alpha$$

can be approximated by the distribution of another random variable $U(y)$ based on Ornstein-Uhlenbeck processes such that

$$\Pr\{\sup_{0\leq y\leq Y} U(y) > \sqrt{\tau_{1-\alpha}^\infty}\} = \alpha \quad (18)$$

where $0 \leq y \leq Y = \ln[(1-h)(1-l)/hl]$, $h(N) = l(N) = (\ln N)^{3/2}/N$, and the supremum, $\sup_{0\leq y\leq Y} U(y)$, is the least upper bound of $U(y)$. The analytical expression of the distribution of $U(y)$ in eq 18 has been obtained by Vostrikova to give[54]

$$\alpha = \frac{(\tau_{1-\alpha}^\infty)^{d/2} e^{-\tau_{1-\alpha}^\infty/2}}{2^{d/2}\Gamma(d/2)}\left[Y - \frac{d}{\tau_{1-\alpha}^\infty}Y + \frac{4}{\tau_{1-\alpha}^\infty} + O\left(\frac{1}{(\tau_{1-\alpha}^\infty)^2}\right)\right] \quad (19)$$

where d is the dimension of the parameter to be tested and is unity in the current case (one change point in intensity), and $\Gamma(d) = \int_0^\infty t^{d-1} e^{-t}\, dt$ is the Gamma function. The asymptotic critical region $\tau_{1-\alpha}^\infty$ can be found by numerically solving eq 19 and serves as the upper bound for finding $\tau_{1-\alpha}^\circ(N)$. We have computed the critical regions of eq 5 for $N = 1...1000$ at $\alpha = 0.31$ (note that this does not correspond to one standard deviation but is included here only as a reference), $\alpha = 0.1$, 0.05, and $\alpha = 0.01$. The results are plotted in Figure 14 as a function of N, a portion of which is listed in Table 1.

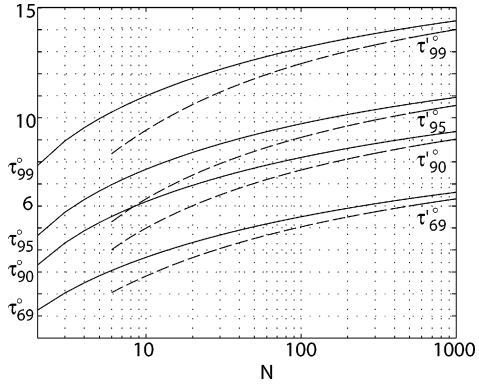

Figure 14. Numerical values of critical regions $\tau_{1-\alpha}^\circ$ (—) as functions of N for eq 5. Critical regions $\tau_{1-\beta}'^\circ$ (— —) for conservative confidence intervals analogous to eq 9 are also plotted.

Analogous procedures can be used to find the critical regions $\tau_{1-\alpha}$ for eq 6 by numerically solving eq 7. The results are plotted in Figure 15 and partially listed in Table 1.

We next turn to computation of the critical regions $\tau_{1-\beta}'$ (or $\tau_{1-\beta}'^\circ$) for β-level confidence intervals once the intensity change point is found. The conservative confidence interval in eq 9 at various N and β can be found by solving

$$\Pr(Z_N - \mathscr{T}_k \leq \tau_{1-\beta}') = 1 - \beta$$

$$= \Pr\left(a_k\frac{N}{N-1} \leq V_k \leq b_k\frac{N}{N-1};\ k = 1...N-2\right)$$

where the scaling factor $N/(N-1)$ results from Worsley's approximation to conditionality of T and T_k in eq 8. Except for the scaling factor, the numerical procedures for finding a_k, b_k, and $\tau_{1-\beta}$ are identical to those described earlier and are not repeated here. The results are also included in Figures 14 and 15.

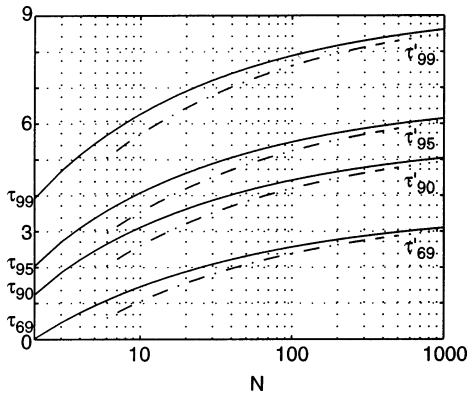

Figure 15. Numerical values of critical regions $\tau_{1-\alpha}$ (—) and $\tau'_{1-\beta}$ (— —) as functions of N for eqs 6 and 9, respectively.

Finally, a note on the application of these results. Although the standardized and weighted likelihood ratio test in eq 6 offers a better performance than the classic test in eq 5, deviation of type-I error rate from ideality (cf. Figure 3) still exists and becomes more deleterious as the number of data points n increases. Through our simulation studies, we found that there is no observable degradation of the performance of our method when $n < 1000$. We therefore recommend that data sets be processed in segments of less than 1000 photons. The break point can be at a detected change point or, when there is no change point within the 1000-photon segment, at a photon index that allows at least a 200-photon overlap between consecutive segments. The change points thus detected can then be combined for the hierarchical clustering, expectation maximization, and BIC analyses. There is thus no limit to the length of trajectories that may be analyzed. A complete listing of numerical results for the critical regions up to $n = 1000$ are provided in the Supporting Information.

References and Notes

(1) Moerner, W.; Orrit, M. *Science* **1999**, *283*, 1670−1676.
(2) Mandel, L.; Wolf, E. *Optical Coherence and Quantum Optics*; Cambridge University Press: New York, 1995.
(3) Nirmal, M.; Dabbousi, B. O.; Bawendi, M. G.; Macklin, J. J.; Trautman, J. K.; Harris, T. D.; Brus, L. E. *Nature* **1996**, *383*, 802−1106.
(4) Neuhauser, R. G.; Shimizu, K. T.; Woo, W. K.; Empedocles, S. A.; Bawendi, M. G. *Phys. Rev. Lett.* **2000**, *85*, 3301−3304.
(5) Kuno, M.; Fromm, D.; Hamann, H.; Gallgher, A.; Nesbitt, D. *J. Chem. Phys.* **2001**, *115*, 1028−1040.
(6) Krauss, T. D.; O'Brian, S.; Brus, L. E. *J. Phys. Chem. B* **2001**, *105*, 1725−1733.
(7) Dickson, R. M.; Cubitt, A. B.; Tsien, R. Y.; Moerner, W. E. *Nature* **1997**, *388*, 355−358.
(8) Lu, H. P.; Xun, L.-Y.; Xie, X. S. *Science* **1998**, *282*, 1877−1882.
(9) Zhuang, X. W.; Bartley, L. E.; Babcock, H. P.; Russell, R.; Ha, T. J.; Herschlag, D.; Chu, S. *Science* **2000**, *288*, 2048−2051.
(10) Tan, E.; Wilson, T. J.; Nahas, M. K.; Clegg, R. M.; Lilley, D. M. J.; Ha, T. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9308−9313.
(11) Yasuda, R.; Masaike, T.; Adachi, K.; Noji, H.; Itoh, H.; Kinosita, K., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9314−9318.
(12) Yildiz, A.; Forkey, J. N.; McKinney, S. A.; Ha, T.; Goldman, Y. E.; Selvin, P. R. *Science* **2003**, *300*, 2061−2065.
(13) Diez, M.; Zimmermann, B.; Börsch, M.; König, M.; Schweinberger, E.; Steigmiller, S.; Reuter, R.; Felekyan, S.; Kudryavtsev, V.; Seidel, C.; Gräber, P. *Nat. Struct. Mol. Biol.* **2004**, *11*, 135−141.
(14) Ha, T.; Glass, J.; Enderle, T.; Chemla, D. S.; Weiss, S. *Phys. Rev. Lett.* **1998**, *80*, 2093−2096.
(15) Tinnefeld, P.; Herten, D. P.; Sauer, M. *J. Phys. Chem. A* **2001**, *105*, 7989−8003.
(16) Mei, E.; Vinogradow, S.; Hochstrasser, R. M. *J. Am. Chem. Soc.* **2003**, *125*, 13199−13204.
(17) Biju, V.; Micic, M.; Hu, D.; Lu, H. P. *J. Am. Chem. Soc.* **2004**, *126*, 9374−9381.
(18) van den Bout, D.; Yip, W.; Fu, D.; Swager, T.; Barbara, P. *Science* **1997**, *277*, 1074−1077.
(19) Bartko, A. P.; Dickson, R. M. *J. Phys. Chem. B* **1999**, *103*, 3053−3056.
(20) Weston, K. D.; Goldner, L. S. *J. Phys. Chem. B* **2001**, *105*, 3453−3462.
(21) Bartko, A. P.; Xu, K.; Dickson, R. M. *Phys. Rev. Lett.* **2002**, *89*, 026101−1−026101−4.
(22) Stevens, B. C.; Ha, T. *J. Chem. Phys.* **2004**, *120*, 3030−3039.
(23) Lippitz, M.; Hübner, C. G.; Christ, T.; Eichner, H.; Bordat, P.; Herrmann, A.; Müllen, K.; Basché, T. *Phys. Rev. Lett.* **2004**, *92*, 103001−1−103001−4.
(24) Moerner, W.; Fromm, D. *Rev. Sci. Instrum.* **2003**, *74*, 3597−3619.
(25) Kulzer, F.; Orrit, M. *Annu. Rev. Phys. Chem.* **2004**, *55*, 585−611.
(26) Andrec, M.; Levy, R. M.; Talaga, D. S. *J. Phys. Chem. A* **2003**, *107*, 7454−7464.
(27) Watkins, L. P.; Yang, H. *Biophys. J.* **2004**, *86*, 4015−4029.
(28) Fisher, R. A. *Proc. Cam. Philos. Soc.* **1925**, *22*, 700−725.
(29) Fréchet, M. *Metron* **1930**, *8*, 3.
(30) Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, 1946.
(31) Rao, C. R. *Proc. Cambridge Philos. Soc.* **1949**, *45*, 213−218.
(32) Chen, J.; Gupta, A. K. *Commun. Statist.-Simula.* **2001**, *30*, 665−697.
(33) Worsley, K. J. *Biometrika* **1983**, *70*, 455−464.
(34) Worsley, K. J. *Biometrika* **1986**, *73*, 91−104.
(35) Boudjellaba, H.; MacGibbon, B.; Sawyer, P. *Commun. Statist.−Theory Methods* **2001**, *30*, 407−434.
(36) Noé, M. *Ann. Math. Stat.* **1972**, *43*, 58−64.
(37) Worsley, K. J. *Biometrics* **1988**, *44*, 259−263.
(38) Henderson, R. *Biometrika* **1990**, *77*, 835−43.
(39) Fraley, C.; Raftery, A. E. *Comput. J.* **1998**, *41*, 578−588.
(40) Ward, J. H., Jr. *J. Am. Statist. Assoc.* **1963**, *58*, 234−244.
(41) Fraley, C.; Raftery, A. E. *J. Am. Statist. Assoc.* **2002**, *97*, 611−631.
(42) Scott, A. J.; Symons, M. J. *Biometrics* **1971**, *27*, 387−397.
(43) Farley, C. *SIAM J. Sci. Comput.* **1998**, *20*, 270−281.
(44) Dempster, A. P.; Laird, N. M.; Rubin, D. B. J. R. *Stat. Soc. B* **1977**, *39*, 1−38.
(45) Lanterman, A. D. *Int. Stat. Rev.* **2001**, *69*, 185−212.
(46) Leroux, B. G. *Ann. Stat.* **1992**, *20*, 1350−1360.
(47) Schwarz, G. *Ann. Stat.* **1978**, *6*, 461−464.
(48) Wang, J.; Wolynes, P. *Phys. Rev. Lett.* **1995**, *74*, 4317−4320.
(49) Yang, H.; Xie, X. S. *J. Chem. Phys.* **2002**, *117*, 10965−10979.
(50) Barsegov, V.; Chernyak, V.; Mukamel, S. *J. Chem. Phys.* **2002**, *116*, 4240−4251.
(51) Owen, A. B. *J. Am. Statist. Assoc.* **1995**, *90*, 516−521.
(52) Press: W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: New York, 1988.
(53) Gombay, E.; Horvath, L. J. *Multivariate Anal.* **1996**, *56*, 120−152.
(54) Vostrikova, L. Y. *Theory Probab. Appl.* **1981**, *26*, 356−362.