UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Optimal Statistical Methods for Analysis of Single-Molecule Data

Permalink
https://escholarship.org/uc/item/4gx17974

Author
Watkins, Lucas

Publication Date
2009

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Optimal Statistical Methods for Analysis of Single-Molecule Data

by

Lucas Paul Watkins

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Chemistry

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Graham Fleming, Chair
Professor Carlos Bustamante
Professor Daniel Fletcher

Fall 2009

Optimal Statistical Methods for Analysis of Single-Molecule Data

Copyright 2009
by
Lucas Paul Watkins

Abstract

Optimal Statistical Methods for Analysis of Single-Molecule Data

by

Lucas Paul Watkins

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Graham Fleming, Chair

Time-resolved single molecule fluorescence measurements are important probes for the conformational dynamics of biological macromolecules. The best time resolution in such techniques will only be achieved by measuring the arrival times of individual photons at the detector and by making a detailed statistical analysis of the photon arrival time data. This work presents several general approaches to the estimation of molecular parameters based on individual photon arrival times. In the first, the amount of information present in a data set involving continuous intramolecular motion is quantified by the Fisher information, allowing an algorithm to be constructed which achieves the theoretical limits on time and distance resolution. The second approach is tailored to the analysis of discrete intensity jumps in optical measurements of single molecules or single particles. It uses a recursive generalized likelihood ratio test to determine the location of intensity change points directly from individual photon arrival times. Third, a maximum entropy approach is taken to molecular probability density function estimation. This approach takes advantage of the information measurements previously derived, and avoids the over-fitting produced by straightforward deconvolution methods. All of these analysis methods are applied to measure the persistence length of a series of polyproline oligopeptides.

_____

Professor Graham Fleming
Dissertation Committee Chair

To Abbas, who was there “every time.”

Contents

List of Figures **v**

List of Tables **xiii**

1 Introduction **1**

I An Information-Based Approach to Monitoring Continuous Processes **4**

2 General Information Theory **7**
2.1 Fisher information 8
2.2 Covariance matrix 9
2.3 Maximum likelihood estimators 9

3 Application to Förster Resonance Energy Transfer **11**
3.1 Fisher information and maximum likelihood estimators 12
3.2 Cross talk and cross-excitation between donor and acceptor channels 13
3.3 Distance and time resolution 14
3.3.1 One channel 14
3.3.2 Two channels 15

4 Application to Electron Transfer **18**
4.1 Fisher information and maximum likelihood estimators 18
4.2 Distance and time resolution 19

5 Application to Single Molecule Trajectories **21**
5.1 Maximum information data analysis algorithm 21
5.2 Simulation details 22
5.3 Example trajectories 22

6 Possible Experimental Complications: Triplets and Bias **25**
6.1 Bias in distance measurements and effects of fluorophore intermittency . . . 25
6.2 Experimental calibration 26

II Detecting Intensity Changepoints **28**

7 Intensity Estimation **32**

8 Detection of Fluorescence Intensity Change Points **35**
8.1 Photon-by-photon generalized likelihood ratio test 35
8.2 Implementation 42

9 Determination of the Number of States and their Intensity Levels **44**
9.1 Initial classification by agglomerative hierarchical grouping 44
9.2 Refinement by expectation maximization clustering 46
9.3 Determination of number of states using the Bayesian information criterion (Schwarz's criterion) 47

10 Characterization **50**
10.1 Resolving power of change point determination 50
10.2 Performance of our method 52

III Miscellaneous Statistical Methods **55**

11 Recovering a Probability Density Function from Single Molecule Trajectories **58**
11.1 Gaussian kernel density estimation 60
11.2 Covariance of the raw density 60
11.3 Maximum entropy method 62
11.4 Covariance of the deconvolved density 62
11.5 Computational validation 63

12 Photon-by-photon Calculation of Intensity Correlation **65**

IV Application of the Statistical Methods to a Model System: Polyproline **67**

13 Methods and Materials **69**
13.1 A server-based single-molecule microscope 69
13.2 Sample preparation and characterization 73
13.3 Calibration of the maximum information method for the experimental system 75

14 Static Poly(l-Proline) End-to-End Distance on Single-Molecule Experiment Time Scales **78**

15 Towards Quantitative FRET Measurement **81**
15.1 Sufficient sampling of relative donor-acceptor orientation using MIM 81

15.2 Molecule-to-molecule variations are dominated by parameter calibration uncertainty. 82

16 Worm-Like Chain Model for Poly(l-Proline) Molecules with a Short Persistence Length 85

V Conclusion 88

Bibliography 93

A FRET information 105

B Computation of Critical Regions for Changepoint Detection 108

C Efron's bootstrap 111

List of Figures

1.1 Simulated single-molecule trajectories where the number of detected photons within certain bin times (5, 10, and 30 ms) is recorded as a function of chronological time. The simulation assumes a FRET configuration in which the donor-acceptor distance follows Langevin dynamics evolving on a parabolic potential (see main text for details). Only the donor intensity is shown. The simulation also assumes a confocal optical detection scheme with which the number of detected fluorescence photons from a single molecule is recorded as a function of time. The signal level is set to 3000 counts per second (kcps), and the background level is 0.4 kcps. The molecule undergoes an irreversible photochemical reaction (photobleaching) at ~ 1 s such that it no longer fluoresces. 5

3.1 Energy transfer efficiency as a function of normalized donor (D) acceptor (A) distance x, defined by $x = R/R_0$. 12

3.2 Observation time, in units of $(I_d^\beta)^{-1}$ required to achieve a relative measurement uncertainty $\alpha = 0.1$ as a function of normalized donor-quencher distance $\delta R/R_0 < \alpha$. Shown in the figure are expected time resolutions under various signal/background ratios in the donor channel: $\beta_d = 2$ $(-\cdot-)$, $\beta_d = 5$ $(\circ)$, $\beta_d = 20$ $(\triangle)$, which are compared to that under background-free conditions, $\beta_d \to \infty$ $(\cdot)$. Overlaid is a FRET efficiency curve for comparison $(\cdots)$, referenced to the ordinate on the right. 15

3.3 Observation time T, in units of $(I_d^\beta)^{-1}$, as a function of normalized donor-acceptor distance $x = R/R_0$, required to achieve a relative measurement error of $\sigma(x) < \alpha = 0.1$ in two-channel detection. (A) Background-free scenario when $\rho = 1.5$ ($\triangledown$), $\rho = 1.0$ (—), and $\rho = 0.5$ ($\square$). Background-free, single-channel detection ($\bullet$) is also included for comparison. (B) Background emission is present, but no cross talk between the donor and acceptor channels. The background levels are $\beta = 2$ ($-\cdot$), $\beta = 5$ ($\circ$), and $\beta = 20$ ($\triangle$). The background-free case is also plotted ($-$) for comparison. In all plots on this panel, the background levels are the same for both the donor and acceptor channels. (C) Both background, $\beta_d = \beta_a = 5$, and and cross talk, $\chi_d = \chi_a = 0$ ($*$), $\chi_d = \chi_a = 0.25$ ($+$), and $\chi_d = \chi_a = 0.5$, are present for two-channel detection. The background- and cross-talk-free, two-channel detection curve (—) is also plotted for comparison. (D) Comparison of background-free single-channel detection ($\bullet$), single-channel with a signal/background ratio of $\beta_d = \beta_a = 5$ ($\circ$), background- and cross-talk-free two-channel detection (—), and two-channel detection with signal/background ratio of $\beta_d = \beta_a = 5$ and cross-talk coefficient of $\chi_d = \chi_a = 0.25$ ($+$). On A-D FRET efficiency E as a function of x is overlaid ($\cdots$) and referenced to the ordinate to the right. 16

4.1 Observation time, in units of $(I_0)^{-1}$ under background-free conditions, or in units of $(I_0^\beta)^{-1}$ in the presence of background, required to achieve a relative measurement uncertainty $\alpha = 0.7$ as a function of normalized donor-quencher distance $x_e = R_e\beta_e$. This confidence interval corresponds to an absolute error of ~ 0.5 Åif the distance dependence of electron transfer β_e is 1.4 Å. The value ξ in these plots is set to 10,000, corresponding to $k_e = 10^{13}$ and $k_0 = 10^9$. Shown in the figure are expected time resolutions under various signal/background ratios: $\beta = 2(-\cdot)$, $\beta = 5(\circ)$, and $\beta = 20(\triangle)$, which are compared to that under the background-free conditions, $\beta \to \infty$ (—). Overlaid is excited-state lifetime relative to the quencher-free case for comparison ($\cdots$), referenced to the ordinate on the right. 19

5.1 Flowchart for the maximum information algorithm. 22

5.2 Sample FRET trajectories analyzed according to the maximum information algorithm. The top half of each panel shows the simulated intensities on the donor (–) and acceptor (−−) channels as a function of time. The bottom half of each panel compares the analysis of the given FRET trajectory with the simulated trace corresponding to the true trajectory (–). The dashed black line (−·) is the trajectory recovered by the maximum information method, and shaded areas outline the standard deviations calculated from the information analysis. All trajectories were generated according to Eq. 43 on the potential $\bar{V}(x) = 20(x - 0.9)^2$ at a temperature of $\Theta = 1/k_B$. Trajectories A and B were simulated with $\gamma = 10$ and $\gamma = 1$, respectively, $I_d^0 = I_a^0 = 3000$cps and $B_d = B_a = 400$cps, and analyzed with $\alpha = 0.07$. Their intensity trajectories were calculated with 15-ms bins. C was simulated with $\gamma = 0.3$, $I_d^0 = I_a^0 = 10,000$cps, and $B_d = B_a = 1200$cps, and analyzed with $\alpha = 0.1$. Its intensity trajectory was calculated with 5-ms bins. 23

5.3 The theoretical limit of the time resolution as a function of x, calculated using Eq. 37 with parameters determined by the simulation parameters for the trajectory in Fig. 7 B (–) is compared with the δt values from the maximum information algorithm (□). 24

6.1 (A–C) Mean error and (D–F) root mean-square error as a function of triplet lifetime (Eq. 44). Only donor triplet states are allowed in A and D; only acceptor triplet states in B and E; and both acceptor and donor triplet states are allowed in C and F. Trajectories were simulated at constant x-values of 0.8 (□), 1.0 (◁), and 1.2 (◇)) with effective $\phi_{isc} = 1 \times 10^{-2}$. The trajectories were analyzed by the maximum information algorithm with a $\alpha = 0.05$. Under typical experimental conditions the triplet lifetime t will not exceed 1-2 ms. At these lifetimes, there is no significant effect on the accuracy of the algorithm. 26

6.2 (A) 2-ms binned trajectory. (B) 10-ms binned trajectory. (C) 20-ms binned trajectory. (D) Photon-by-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 Chapter 10.2. 30

7.1 (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. 33

8.1 Error rates as a function of the test point k under the null $H_\circ$ 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. 37

8.2 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. 37

8.3 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 $1-\alpha = 0.69, 0.90, 0.95, 0.99$. 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 $\sim$1.5 for $\alpha = 0.31$, $\sim$1.66 for $\alpha = 0.1$, $\sim$1.72 for $\alpha = 0.05$, and $\sim$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). 39

8.4 Empirical evaluation of the accuracy of confidence intervals calculated from Eq. 8.6. 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 100-th 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. 41

8.5 (A) Plot of $\mathcal{L}_k$ as a function of the chronological photon arrival time of each detected photon using the data from Fig. (7.1). 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 Fig. 7.1B for comparison. The "true"location of intensity jump and inferred 95% confidence interval $\widetilde{C}_{1-\beta}$ are indicated as vertical and horizontal bars, respectively. 42

8.6 Single-molecule emission time trace (offset by 50 and scaled to a 5-ms bin time) inferred from the simulated data in Fig. 6.2 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. . . . 43

9.1 An EM algorithm for single-molecule intensity level clustering. 46

9.2 BIC as a function of number of groups, showing that there are most likely 5 distinct intensity states. The numerical example in Fig. 6.2 is used. 48

10.1 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 $\widetilde{C}_{69}$ 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. 51

10.2 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. 52

10.3 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. 53

10.4 Comparison of histograms constructed from a simulated trajectory by constant-time binning with the true probability density (solid lines), illustrating the challenge in determining conformational distributions from single-molecule measurements. Here $x \equiv R/R_0$ is the normalized donor-acceptor distance in a Förster-type resonance energy transfer measurement (cf. Eq. 13.3). The left column shows the true x trajectory, as well as the donor (black) and acceptor (gray) intensity trajectories, binned at 10, 50, and 100 ms. The other three columns compare the underlying probability density (—) with histograms computed from the equal-time binned intensity trajectories, using N_b bins in the x coordinate. The numbers of bins for the set of histograms in the third column are generated according to Scott's formula for the optimal bin width [121]. This formula, like most non-parametric density estimation [122], assumes that there is no error associated with each datum. This assumption is not valid for single-molecule measurements. The columns to the left and right use bin numbers half and twice the optimal value, respectively. Note that different regions of the histogram are broadened differently because of the changes in the variance of the distance estimator as a function of distance. [137] 57

11.1 Comparison of true (—) and raw (– –) PDFs for trajectories simulated (A) at constant x, (B) on a harmonic potential, and (C) on a bimodal potential. 59

11.2 Results from deconvolution of test trajectories: (A) True (black) and deconvolved (red) probability densities for constant trajectories at $x = 0.6$, 0.8, 1.0, 1.2, and 1.4. (B) True (black) and deconvolved (red) probability densities for a trajectory simulated on a harmonic potential centered at $x = 1.0$. The 95% confidence interval for the deconvolved density is indicated by dashed lines. (C) True (black) and deconvolved (red) probability densities for a trajectory simulated on a bimodal potential centered at $x = 1.0$. The 95% confidence interval for the deconvolved density is indicated by broken lines. (D) Covariance matrix for the deconvolved density shown in A at $x = 1.0$. (E) Covariance matrix of the deconvolved density shown in B. (F) Covariance matrix of the deconvolved density shown in C. 63

13.1 Configuration of a cw-excitation photon-by-photon microscope. Illumination is provided by a continuous-wave diode pumped solid-state laser. Telescope lens assembly L1 expands the beam to 8 mm in order to fill the back aperture of the objective O. A shutter S is installed in the beam path to minimize unnecessary illumination. A polarization element P is placed immediately before the entrance of the home-built microscope body. Dichroic mirror D1 reflects the excitation light into the back aperture of an infinity-corrected objective O, which focuses it onto the sample mounted on a piezo stage. Excitation light that leaks through D1 is detected by a photodiode PM to monitor laser excitation power. Fluorescence from the sample is collected and collimated by objective O and passes through dichroic D1 before being focused by tube lens L2 and split into donor and acceptor channels by dichroic mirror D2. The emitted photons finally pass through emission filters E1 and E2 before being recorded by avalanche photodiodes APD1 and APD2. The stage and the APDs are each controlled by a separate computer server, all of which are connected via a high-speed network backbone via the TCP/IP protocol to a client computer which runs the user interface. The modular design of the system is such that if more parameters such as polarization are to be measured, one simply drops in additional APD/server modules. . . . 70

13.2 (A) Raster-scanned image of doubly labeled $P_{12}CG_3K$(biotin), where the donor and acceptor emission are represented in blue and red false color, respectively. The image was binned at 40 nm and filtered with a 5-pixel by 5-pixel Gaussian averager. (B) A representative intensity trajectory acquired from the spot indicated by the arrow in (A). The segmentation of different regions for data analysis is shown by the vertical dashed lines. 72

13.3 Chemical structure of doubly labeled P_nCG_3K(biotin), where $n = 8$, 12, 15, 18, 24. 74

14.1 Intensity (left), FRET efficiency (center), and calculated donor-acceptor distance (right) as a function of time. Intensity trajectories are binned at 100 ms on both the donor (black) and acceptor (gray) channels. Distance and efficiency trajectories trajectories are calculated using the maximum information method and assume $R_0 = 51$ Å. Dimensions of the gray boxes on the distance and efficiency trajectories indicate the time resolution (horizontal dimension) and expected 95% confidence interval in energy transfer efficiency or donor-acceptor distance (vertical dimension). 79

14.2 Raw (- -) and deconvolved (—) distribution functions from single P_nCG_3K(biotin) trajectories shown in Fig. 14.1. 80

15.1 Probability density $p(\langle\kappa^2\rangle_m)$ as a function of the m-measurement mean value $\langle\kappa^2\rangle_m$. The bold line highlights the $m = 1$ function. As m increases, the mean value rapidly peaks at the ensemble-averaged value, $\langle\kappa^2\rangle_{m\to\infty} = 2/3$. The inset shows the relative standard deviation of $\langle\kappa^2\rangle_m$ as a function of m, which quickly decreases to less than 10% (at $m \simeq 15$). 82

15.2 Distributions of donor-acceptor distances and FRET efficiency of P_nCG_3K(biotin), ($n = 8, 12, 15, 18, 24$). The solid line is a Gaussian distribution with a variance that is the mean of expected variance of individual molecules by propagating uncertainties in parameters calibration. This indicates that the molecule-to-molecule distributions are dominated by uncertainties in parameter calibration. 83

16.1 Comparison of experimental results with various models for poly(L-proline). As a reference, ± 2 Å ranges for the model are also displayed. Experimental error bars represent 95% confidence interval of the experimental mean. . . . 86

A.1 Bias in estimators of x based on simulations of (A) one-channel FRET and (B) two-channel FRET at distances of 0.8, 1.0, and 1.2 and electron transfer at distances of (C) x= 9, 12, and (D) x 6 (- - -). Wedges are placed to indicate the values of the information (J) which satisfy the cutoff values of a 0.1 and a 0.07, as discussed in the main text. 106

B.1 Numerical values of critical regions $\tau^{\circ}_{1-\alpha}$ (—) as functions of N for Eq. 8.2. Critical regions $\tau'^{\circ}_{1-\beta}$ $(--)$ for conservative confidence intervals analogous to Eq. 8.6 are also plotted. 109

B.2 Numerical values of critical regions $\tau_{1-\alpha}$ (—) and $\tau'_{1-\beta}$ $(--)$ as functions of N for Eqs. (8.3) and (8.6), respectively. 110

List of Tables

8.1 Level α and β points for intensity change detection using Eqs. (8.3) and (8.6). 40

13.1 Steady-state fluorescence anisotropy of doubly labeled poly-prolines, excited at 532nm. 74

14.1 Mean ($\bar{\tau}$) and standard deviation (σ_τ of donor (d) and acceptor (a) bleaching times for P_nCG_3K(biotin). 78

Acknowledgments

There are many people who helped make my time at Berkeley a wonderful experience.

Although my interactions with Professor Graham Fleming were few and far between, he provided a constant, sterling example of how a research group should be led. Without his support and personal regard—especially at the end of the process—this dissertation would never have been completed. Abbas Rizvi was always there, both as a friend and as a co-worker, but his brilliance never outshone his compassion. My lab mates were valuable sources of aid and inspiration, especially my undergraduate assistants Sucharita Bhattacharyya and Irina Draganovic. The laboratories of Charles Harris, A. Paul Alivisatos, Graham Fleming, and Carlos Bustamante were unfailingly generous in sharing their equipment and their expertise. I could not have built my microscope without assistance and training from Eric and his crew in the machine shop. Heidi Bond reminded me every day that joy can be found anywhere, and Kimberly Pierce was a constant cheerleader. My family never lost faith in me, even when I decided that law school was in my future. Andrea Tao showed me that it was possible to be an insightful researcher, a dedicated friend, and an adventurous soul, all at the same time. Finally, Jacqueline Hom gave me a safe space in which to work, as well as frequent reminders that this dissertation was not going to write itself.

The work reported in this thesis was performed in the laboratory and with the assistance of Haw Yang at the University of California at Berkeley. It received financial support from the National Science Foundation, the donors of the Petroleum Research Fund of the American Chemical Society, the University of California at Berkeley, and the Laboratory Directed Research and Development Program at the Lawrence Berkeley National Laboratory, under U.S. Department of Energy contract No. DE-AC03-76SF00098. I was the fortunate recipient of a graduate research fellowship from the National Science Foundation. I would also like to thank David King for assistance with synthesis of the poly(L-proline) samples used in this work, A.P. Alivisatos for the generous loan of a fluorometer, and C.C. Hayden and C.-Y. Wang for helpful discussions.

Chapter 1

Introduction

The greatest advantage of optical single-molecule spectroscopy—the elimination of the ensemble average—is also its greatest fault. Elimination of the ensemble average allows direct observation of conformational fluctuations. From these observations it may be possible to draw meaningful connections between protein structure, molecular movement, and enzymatic function. But these measurements also approach the fundamental limitations of optical detection. The effort to achieve higher resolution confronts the fundamental limitation of photon statistics.

The basic problem is that, while the system under observation is constantly changing, information about the system is parceled out photon by photon. Any serious effort to extract the most possible information from an optical microscope must confront the nondeterministic nature of photon emission. Raw experimental data are inundated with Poissonian photon counting noise. For instance, as pointed out by Kollner and Wolfrum [75], at least 185 photons are required to measure with 10% accuracy a static, monoexponential fluorescence lifetime from a single molecule. The potential benefits of single molecule measurements justify a concerted effort to achieve optimal treatment of single molecule data.

Since the pioneering experiments by Moerner and Kador [95] and Orrit and Bernard [106], optical single-molecule spectroscopy has gained great momentum both in technology development [96] and in applications [100, 140, 147]. It is particularly suited for the investigation of biological systems because it probes dynamics on the enzymatically relevant submicrosecond-to-second timescales. Using this technique, for example, enzymatic reaction rates of cholesterol oxidase [89] and horse radish peroxidase [35] were found to fluctuate with time; previously unreported folding intermediates were directly observed in RNA molecules [129, 157] and their transition states characterized [16]; the detailed dynamics of F1-ATPase rotation were revealed [1, 154]; and the timescales of protein conformational fluctuations were quantitatively characterized and modeled [150]. Not only have single-molecule experiments contributed to our fundamental understanding of biomolecular function, they have also stimulated much theoretical work that provides physical insights into such processes as dynamic disorder, conformational fluctuations, and photon statistics [73].

Monitoring biochemical events in real-time utilizing optical single-molecule spec-

troscopy can in principle establish a quantitative relationship between the static structure and the dynamic function of a biomolecule. Structural changes in a single molecule can be probed using Förster-type resonance energy transfer (FRET) [59]. Although FRET allows studies of structural changes on the length scale of an entire biomolecule (20–80 Å) [126], the minutiae of conformational fluctuations that accompany or facilitate the functioning of a biomolecule can be examined by utilizing excited-state electron transfer (ET) quenching of fluorescence. ET is sensitive to distance variations on the Ångström length scale and is a probe of conformational fluctuations at the catalytically active site [150]. In addition, fluorescence polarization experiments can yield information about the orientational dynamics of a molecule [1, 7, 154].

A disadvantage of single molecule spectroscopy is that the organic dyes commonly used as fluorescent probes eventually undergo irreversible photodegradation, limiting the length of recordable single-molecule trajectories [31, 40]. Photobleaching causes two problem. First, it limits the amount of information that can be obtained in a single trajectory. Second, it is not always guaranteed that the molecular system under investigation explores all possible configurations during the measurement period, as the ergodic principle would have dictated. Such non-ergodic conditions are expected to be encountered experimentally in a reactive system such as a single enzyme molecule. These practical matters inevitably hamper the experimentalist's ability to quantitatively characterize the fast conformational motions that contribute to the function of a biomolecule. The challenge thus lies in the efficient extraction of the maximal amount of dynamic information from short, noisy single molecule traces.

Many theoretical tools for the analysis of single molecule systems already exist. In the context of room-temperature time-resolved studies, for example, the correlation method [105, 135], which is a very sensitive probe of the memory of a system, has been used to analyze the dynamics of a single enzyme molecule [2, 113] and conformational fluctuations [25, 35, 139, 158]. In principle, more features can be revealed using higher order correlations [152] or event echo analysis [23, 153]. These methods are applicable to systems that exhibit stationarity and ergodicity. In cases where the measurement period is commensurate with the interconversion timescale between states [36, 72], kinetic parameters can be deduced by applying the motional narrowing concept originally developed for lineshape analysis [10, 11, 9, 51]. However, due to nonideal experimental conditions—namely, short trajectories and the non-ergodic conditions typically seen in a reactive setting—it may prove difficult to use these powerful theoretical tools on experimental data.

Recent advances in experimental data registering, originally developed for time-correlated single photon counting (TCSPC) [8, 15], allow the chronological arrival time of each detected photon to be recorded. This has stimulated new experimental schemes such as multiparameter fluorescence spectroscopy [77] and photon-by-photon correlation [150]. For ergodic systems, the latter method allows detailed examination of conformational dynamics that covers a wide range of timescales from submicroseconds to tens of seconds. Advanced statistical methods that rely on stationarity and ergodicity have also been developed to elicit physical parameters from such time-stamped data streams [103, 151, 152]. Photon by photon approaches assuming Bayesian prior models have been proposed for time-dependent ET (S. C. Kou, X. S. Xie, and J. S. Liu, unpublished) and FRET distance measurements

[116].

Despite these exciting new developments, a general, nonparametric method that allows in-depth studies of a reactive, non-ergodic, single molecule system to relate its dynamics to its biochemical function is still lacking. In particular, such methods should allow one to accurately determine the conformational state of a single enzyme molecule with a temporal resolution that is better than its catalytic timescale (microseconds to seconds), while simultaneously addressing the problems of background photons, cross talk between multiple data acquisition channels, and error analysis of the results obtained. These model-free analyses allow the construction of a quantitative model that extracts the dynamics underlying the motions of a complex biological macromolecule.

This work describes a theoretical effort to achieve the limits of information extraction in three different contexts and the subsequent experimental validation of these methods. Part I develops a model-free method to analyze single molecule trajectories where the observable (perhaps intensity or distance) varies continuously. Part II takes a model-free approach to systems where the parameter under investigation causes sudden intensity jumps (i.e., the timescale of the change is much faster than the rate of change). Part III describes two techniques for estimation of molecular probability density functions. Finally, Part IV validates these methods experimentally in a model system, poly(L-proline).

Part I

An Information-Based Approach to Monitoring Continuous Processes

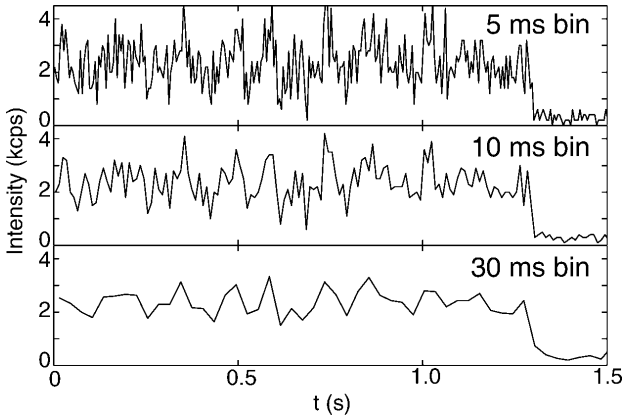


Figure 1.1: Simulated single-molecule trajectories where the number of detected photons within certain bin times (5, 10, and 30 ms) is recorded as a function of chronological time. The simulation assumes a FRET configuration in which the donor-acceptor distance follows Langevin dynamics evolving on a parabolic potential (see main text for details). Only the donor intensity is shown. The simulation also assumes a confocal optical detection scheme with which the number of detected fluorescence photons from a single molecule is recorded as a function of time. The signal level is set to 3000 counts per second (kcps), and the background level is 0.4 kcps. The molecule undergoes an irreversible photochemical reaction (photobleaching) at ~ 1 s such that it no longer fluoresces.

Obtaining quantitative information about continuous processes from single molecule trajectories is one of the major challenges in single-molecule spectroscopy. The problem is illustrated by the simulated single-molecule traces displayed in Fig. 1. The trace that is binned at 5 ms is very noisy. As the bin time is increased from 5 to 30 ms (lowering the time resolution), the noise subsides, but the dynamics are obscured as well. On the other hand, very small bin widths (high time resolution) lead to very large statistical errors. What is the correct balance between time resolution and precision? This part presents a framework for relating time resolution to measurement uncertainty. Once this relationship is understood, maximum likelihood estimators–which achieve the theoretical limits of accuracy–are used in a data analysis algorithm to achieve the maximum possible time resolution given a desired measurement accuracy.

With the ultimate goal of developing general methods to build a quantitative dynamic structure-function relationship in biological macromolecules, this part address the theoretical limits of time and distance resolution in time-resolved single-molecule measurements of continuous processes. Principles from information theory [26], specifically the Fisher information [44], are essential in quantifying the knowledge that can be extracted from experimental data.

As a basic introduction to the technique, consider the task of estimating an experimental parameter q from a measurable quantity λ. Note that the only restriction on the parameter q is that the value of λ must depend on q in some way. Otherwise q may represent any property of the experimental system, including distance, orientation, and

oxidation state. The distribution of experimentally observed λ is given by the likelihood function $f(\lambda; q)$: the probability, given that the value of the parameter is q, that the observable will be λ. The Fisher information about q is given by

$$J(q) = \left\langle \left(\frac{\partial}{\partial q} \ln[f(\lambda; q)] \right)^2 \right\rangle_\lambda , \tag{1.1}$$

where $\langle \cdots \rangle_\lambda$ denotes the expectation value weighted by the likelihood $f(\lambda; q)$ over all possible λ. One may intuitively expect uncertainties in measuring q to be related to the Fisher information, because q can be determined more accurately if more information about q can be obtained. This qualitative understanding can be quantitatively expressed by the Cramér-Rao-Fréchet inequality [28, 49, 110], $\mathrm{var}(q) \geq (d \langle F(q) \rangle / dq)^2 J(q)^{-1}$, where $\langle F(q) \rangle$ is the expectation value of q from the estimator F. For an unbiased estimator, $\langle F(q) \rangle = q$. The Cramér-Rao-Fréchet inequality states that the variance of the best possible estimator of q is given by the inverse of its Fisher information matrix. In general, the maximum likelihood estimator (MLE)—which is determined by maximizing $f(\lambda; q)$ as a function of q given the experimental observation λ—is a good starting point for estimating q because it is asymptotically normal (Gaussian) under most conditions.

Based on these considerations, Chapter 2 derives the basic equations that determine the best achievable time resolution in a single molecule fluorescence photon-counting experiment. In this context, the measurable quantity (above, λ) will be the arrival times of individual photons. The general case is then particularized for Förster Resonance Energy Transfer in Chapter 3 and for Electron Transfer in Chapter 4. These equations allow Chapter 5 to propose and test a data reduction algorithm to extract, photon by photon, the maximum amount of information in distance measurements as a function of time. Finally, Chapter 6 considers potential complications, such as fluorophore blinking and estimator bias.

Chapter 2

General Information Theory

The experimentalist conducting a time-resolved single-molecule fluorescence measurement wishes to measure some parameter q as a function of time. In general, q is a dynamic variable that changes with time as the molecular conformation undergoes thermal fluctuations. If fluctuations in q cause corresponding fluctuations in the emitted fluorescence intensity of the molecule, then the dynamics of q can in principle be followed in real time by recording the arrival times of the emitted photons. If the photons can be meaningfully separated based on wavelength, polarization, or some other property, they may be detected and analyzed on separate channels.

It is not immediately clear how to analyze the data thus acquired, nor is it clear exactly how much that data will mean. For example, an observer may measure the donor-acceptor distance of a single-molecule to be 4 nm with a time resolution of 100 μs but with a 68% confidence interval $\sigma(x(t)) = 10$nm. This datum, although measured at a very high time resolution, is not very meaningful. The 10 nm uncertainty is most likely greater than the size of the molecule. Some averaging will thus be required before a meaningful value can be obtained:

$$\bar{q} = \frac{1}{T}\int_{t}^{t+T} q(t')dt'. \tag{2.1}$$

The time interval T is chosen so that the uncertainty associated with this measurement, $\sigma(\bar{q})$, is less than some predefined value. Further time averaging, thus reducing $\sigma(\bar{q}(t))$, will improve the accuracy in q; but at the expense of time resolution. The following discussions are based on a coarse-grained picture in which the parameter q is assumed to remain constant during the time required to reduce the standard deviation below a certain threshold. The rationale behind this assumption is that an observer has no knowledge, a priori, of the true value $\tilde{q}(t)$ until an accurate measurement can be made.

To determine the proper averaging time T, the Fisher information matrix is calculated and then inverted to find the covariance matrix for the parameters of interest. This gives us the Cramér-Rao-Fréchet bound for the variance of an estimator. The MLE, which approaches the Cramér-Rao-Fréchet bound, is then derived.

2.1 Fisher information

Suppose the variables of interest, $\mathbf{q} = \{q_i\}$, are being measured on m independent channels. Typically, $\mathbf{q}$ are chosen such that they are relevant as an indicator of the molecular state on the single-molecule level, for example, the FRET efficiency or the distance between a fluorescent donor and acceptor. Since the exact arrival times of the photons on these channels will be uncorrelated from one another, the Fisher information of these independent channels is additive. The Fisher information can be computed for each of these channels individually. The observed intensity at a detector can be written as $I(\mathbf{q})$. The intensity on the channel is generally measured relative to some constant reference intensity I^0. We can write $I(\mathbf{q})$ as $I^0\zeta(\mathbf{q})$, with the dimensionless scaling factor $\zeta(\mathbf{q})$ containing all of the $\mathbf{q}$-dependence of the detected intensity. The probability density function for observing n photons at the detector for this channel in time T is Poisson,

$$f(n;\mathbf{q},T) = \frac{\left[I^0\zeta(\mathbf{q})T\right]^n}{n!}e^{-I^0\zeta(\mathbf{q})T}. \tag{2.2}$$

Based on this probability density, the Fisher information matrix elements for a single channel are

$$j_{ij}(\mathbf{q}) = \frac{I^0T}{\zeta(\mathbf{q})}\left(\frac{\partial\zeta}{\partial q_i}\right)\left(\frac{\partial\zeta}{\partial q_j}\right). \tag{2.3}$$

This form makes it clear that information is acquired at a rate proportional to $I^0/\zeta(\mathbf{q})$ through the course of the measurement.

The information from independent channels may be combined. The total information matrix is just the sum of the information matrices for each of the m channels:

$$J_{ij} = \sum_{k=1}^{m}\frac{I_k^0T}{\zeta_k(\mathbf{q})}\left(\frac{\partial\zeta_k}{\partial q_i}\right)\left(\frac{\partial\zeta_k}{\partial q_j}\right). \tag{2.4}$$

Adding the effects of a detected background intensity B, the total detected intensity is

$$I(\mathbf{q}) = I^0\zeta(\mathbf{q}) + B \tag{2.5}$$

$$= I^\beta\left[\left(1-\beta^{-1}\right)\zeta(\mathbf{q}) + \beta^{-1}\right], \tag{2.6}$$

where the signal/background ratio $(I^0 + B)/B$ has been written as β and the maximum observed intensity in the presence of background is $I^\beta = I^0/(1-\beta^{-1})$. Finally, then, the total m-channel information is

$$J_{ij} = \sum_{k=1}^{m}\frac{I_k^\beta T\left(1-\beta^{-1}\right)^2}{\left(1-\beta^{-1}\right)\zeta_k(\mathbf{q}) + \beta^{-1}}\left(\frac{\partial\zeta_k}{\partial q_i}\right)\left(\frac{\partial\zeta_k}{\partial q_j}\right). \tag{2.7}$$

The inclusion of background photons on a detection channel thus degrades the information that can be collected.

2.2 Covariance matrix

The Cramér-Rao-Fréchet bound states that the covariance (σ_{ij}) between the estimated parameters q_i and q_j is bounded by the inverse of the information matrix,

$$\sigma_{ij}^2 \geq \left(\frac{\partial \langle F(\mathbf{q})\rangle}{\partial q_i}\right)\left(\frac{\partial \langle F(\mathbf{q})\rangle}{\partial q_j}\right)\left(J^{-1}\right)_{ij} \approx \left(J^{-1}\right)_{ij}, \tag{2.8}$$

where the approximation is true when the bias of the estimator $F(\mathbf{q})$ approaches 0. The bias of an estimator depends on the probability density function of the parameter to be measured. The estimators proposed in this work can be shown to meet the consistency condition $\lim_{t\to\infty} F(\mathbf{q}) \to \mathbf{q}$ [114]. Therefore, the subsequent derivations will assume the use of unbiased estimators. Bias in the short-time limit will be discussed case by case in the Appendix.

Given p parameters to estimate, the information matrix will be of order p. The form of the matrix given in Eq. 2.7 makes it clear that p channels are required to form an invertible matrix. If multiple parameters are to be estimated simultaneously, the entire information matrix must be inverted to find the variances of the individual parameters and their covariances, so at least p independent sources of information are required to estimate p different parameters. Concentrating on the estimation of one variable, q_i, the variance of that measurement is simply $(J_{ij})^{-1}$. If the variable q is to be measured to a relative accuracy of $\alpha \equiv \partial q/q$, the requirement is $\sigma(q) \leq \alpha$. The best possible time resolution will be

$$T = \left(\alpha^2 \sum_{k=1}^{m} \frac{I_k^0}{\zeta_k(q)}\left[\frac{\partial \zeta_k}{\partial q}\right]^2\right)^{-1}. \tag{2.9}$$

2.3 Maximum likelihood estimators

Again concentrating on the estimation of one variable, given expressions for $\zeta_1(q)$ through $\zeta_m(q)$, the total probability density for observing $n_1 \ldots n_m$ photons on channels $1 \ldots m$ is

$$f(n_1, \ldots, n_m; q, T) = \prod_{k=1}^{m} \frac{[I_k(q)T]^{n_k}}{n_k!} e^{-I_k(q)T}. \tag{2.10}$$

The maximum likelihood estimator is the value of q for the observed T and $n_1, \ldots, n_m$ that maximizes $f(n_1, \ldots, n_m; q, T)$. This value of q is given by

$$\frac{\partial}{\partial q} \ln f = \sum_{k=1}^{m} \frac{\partial I_k(q)}{\partial q}\left[\frac{n}{I_k(q)} - T\right] = 0 \tag{2.11}$$

The solution to this equation gives the maximum likelihood estimator in terms of T and $n_1, \ldots, n_m$.

From another point of view, each photon can be regarded as an instantaneous measurement of the state of the system under observation. Then a photon will be detected on channel k with probability

$$P_k = \frac{I_k(q)}{\sum_{k=1}^{m} I_k(q)}. \tag{2.12}$$

The probability distribution for observing n_k photons on channel k, with N being the total number of photons, is

$$f(n_1, \ldots, n_m; q) = N! \prod_{k=1}^{m} \frac{P_k(q)_k^n}{n_k!}. \tag{2.13}$$

The maximum likelihood estimator for q will thus be given by the solution to this equation:

$$\sum_{k=1}^{m} \frac{\partial I_k(q)}{\partial q} \left[\frac{n}{I_k(q)} - \frac{N}{\sum_{k=1}^{m} I_k(q)} \right] = 0 \tag{2.14}$$

The Poisson (Eq. 2.11) and the multinomial (Eq. 14 2.14) approaches are equivalent since the former can be derived as a limiting case of the latter. The only differences are practical. First, the multinomial approach cannot be used with a single-channel measurement. Second, the multinomial approach generally yields simpler maximum likelihood estimators for multiple channel measurements.

Chapter 3

Application to Förster Resonance Energy Transfer

A variety of measures have been proposed and implemented to quantify the extent to which resonance energy transfer occurs from an energy donor to an acceptor [12]. Here, the energy transfer efficiency, E, is used because it has been widely adopted in single-molecule experiments [62]. It is defined as the fraction of photon energy absorbed by the donor that is transferred to the acceptor. In cases where the acceptor is a non-fluorescent quencher, the transfer efficiency is measured by the donor fluorescence intensity alone (single-channel detection) and is expressed as

$$E \equiv 1 - I_d^0(x)/I_d^0, \tag{3.1}$$

where $I_d^0(x)$ is the detected background-free donor intensity and I_d^0 is the detected background-free donor intensity in the absence of the quencher. I_d can be measured in a separate control experiment. For simultaneous detection of donor and acceptor fluorescence, the transfer efficiency is given by

$$E \equiv \frac{1}{1 + \rho I_d^0(x)/I_a^0(x)}. \tag{3.2}$$

I_a and I_d are the detected background-free fluorescence intensities (number of photons per second) of the acceptor and donor channels, respectively. $\rho \equiv \phi_a \eta_a / \phi_d \eta_d$, to be determined experimentally, is a scaling factor that corrects for differences in fluorescence quantum yields of the donor (ϕ_d) and acceptor (ϕ_a) probes, as well as those in detection (η_d for the donor channel and η_a for the acceptor channel).

Within the framework of orientation-randomized dipole-dipole coupling between the donor and acceptor probes [46], the energy transfer efficiency can be related to the distance between the donor and acceptor probes,

$$E \approx \frac{1}{1 + (R/R_0)^6} \equiv \frac{1}{1 + x^6}, \tag{3.3}$$

where x is the normalized donor-acceptor distance, R/R_0, R is the center-to-center distance of the donor and acceptor probes, and R0 is the Förster radius–the distance at which energy

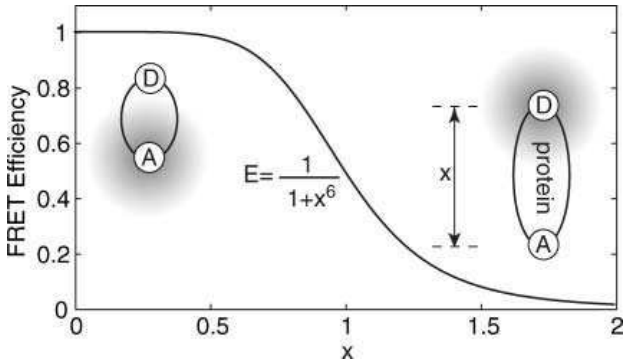


Figure 3.1: Energy transfer efficiency as a function of normalized donor (D) acceptor (A) distance x, defined by $x = R/R_0$.

transfer efficiency is 0.5 (see Fig. 2). For a given donor-acceptor pair, the corresponding Forster radius can be calculated from the donor fluorescence and acceptor absorption spectra, and the orientation factor can be calculated from fluorescence anisotropy measurements [155]. Alternatively, one may construct a series of polypeptides of different length to calibrate the effective R_0 for tethered, gyrating fluorescent probes [117]. Here, it is assumed that both fluorescent probes gyrate around the tethered point on a timescale much shorter than the achievable experimental time resolution T. This assertion can be verified experimentally if necessary for the system under investigation. Information on slow orientation-dependent dynamics can be acquired by considering additional polarization-dependent channels.

3.1 Fisher information and maximum likelihood estimators

A FRET measurement consists of observation of fluorescence from a donor fluorophore and/or from an acceptor chromophore. The E- and x-dependence of the intensities on these channels is

$$\zeta_d(E) = 1 - E, \tag{3.4}$$

$$\zeta_a(E) = E, \tag{3.5}$$

$$\zeta_d(x) = \frac{x^6}{1 + x^6}, \tag{3.6}$$

$$\zeta_a(x) = \frac{1}{1 + x^6}. \tag{3.7}$$

Note that with these definitions, $I_d^0(x) = I_d^0\zeta_d(x)$, $I_a^0(x) = I_a^0\zeta_a(x)$, and $\rho = I_a^0/I_d^0$. Using Eq. 7 to calculate the information, one has

$$J_d(E) = I_d^\beta T \frac{\left(1-\beta_d^{-1}\right)^2}{E\left(1-\beta_d^{-1}\right)-1}, \tag{3.8}$$

$$J_a(E) = I_a^\beta T \frac{\left(1-\beta_a^{-1}\right)^2}{E\left(1-\beta_a^{-1}\right)+\beta_a^{-1}}, \tag{3.9}$$

$$J_d(x) = I_d^\beta T \frac{36x^{10}\left(1-\beta_d^{-1}\right)^2}{\left(1+x^6\right)^3\left(x^6+\beta_d^{-1}\right)}, \tag{3.10}$$

$$J_a(x) = I_a^\beta T \frac{36x^{10}\left(1-\beta_a^{-1}\right)^2}{\left(1+x^6\right)^3\left(1+x^6\beta_d^{-1}\right)}. \tag{3.11}$$

For one-channel measurements, the only possible MLE is that given by the Poisson distribution, Eq. 2.11,

$$\hat{E} = \frac{I_d^\beta T - n_d}{I_d^\beta T\left(1-\beta_d^{-1}\right)}, \tag{3.12}$$

$$\hat{x} = \left(\frac{n_d - I_d^\beta T\beta_d^{-1}}{I_d^\beta T - n_d}\right)^{1/6}. \tag{3.13}$$

For two-channel measurements, the multinomial estimator given by Eq. 2.14 is used. The equations for the maximum likelihood estimators are

$$\hat{E} = \frac{I_d^\beta n_a - I_a^\beta n_d\beta_a^{-1}}{I_d^\beta n_a\left(1-\beta_d^{-1}\right) + I_a^\beta n_d\left(1-\beta_a^{-1}\right)}, \tag{3.14}$$

$$\hat{x} = \left(\frac{\beta_a}{\beta_d} \times \frac{I_d^\beta n_a - I_a^\beta n_d\beta_d}{I_a^\beta n_d - I_d^\beta n_a\beta_a}\right)^{1/6}. \tag{3.15}$$

3.2 Cross talk and cross-excitation between donor and acceptor channels

Due to spectral overlap and other experimental considerations, there is often cross talk between the donor and acceptor channels. Also, if the absorbance spectrum of the acceptor overlaps with that of the donor, the acceptor may be excited directly. As will be shown below, cross talk and cross-excitation simply change the effective signal/background ratio. As such, Eqs. 3.8–3.15 still apply.

Cross-talk coefficients from the donor and acceptor channels are denoted χ_d and χ_a, respectively. The cross-excitation coefficient is denoted χ_x. All three of these coefficients may be measured experimentally. For instance, χ_d can be measured by recording the acceptor channel intensity at different excitation power levels for donor probes whereas χ_a can be measured by recording the donor channel intensity at different excitation power

levels for a control system where donor and acceptor probes are in juxtaposition so that $E \to 1$. χ_x may be measured by recording the acceptor channel intensity in the absence of the donor. With these notations, the distance-dependent photon intensities become

$$I_d^{\beta\chi}(x) = I_d^0\zeta_d(x) + \chi_a I_a^0\zeta_a(x) + B_d, \tag{3.16}$$

$$I_a^{\beta\chi}(x) = I_a^0\zeta_a(x) + \chi_d I_d^0\zeta_d(x) + \chi_x I_d^0 + B_a. \tag{3.17}$$

Expanding the intensity terms, one has

$$I_d^{\beta\chi}(x) = I_d^{\beta}\left[\left(1-\beta_{d\chi}^{-1}\right)\zeta_d(\mathbf{q}) + \beta_{d\chi}^{-1}\right], \tag{3.18}$$

$$I_a^{\beta\chi}(x) = I_a^{\beta}\left[\left(1-\beta_{a\chi}^{-1}\right)\zeta_a(\mathbf{q}) + \beta_{a\chi}^{-1}\right], \tag{3.19}$$

$$\tag{3.20}$$

with

$$\beta_d^{\chi} = \frac{I_d^{\beta}}{B_d + \chi_a I_a^0}, \tag{3.21}$$

$$\beta_a^{\chi} = \frac{I_a^{\beta}}{B_a + (\chi_x + \chi_a) I_d^0}. \tag{3.22}$$

The crests and troughs in one channel correspond with the valleys and peaks in the other, so the only effect of the cross talk is to decrease the apparent signal/background ratio. The expressions given for $J(x)$ and $\hat{x}$ in the previous section still hold, using the new $\beta_{d\chi}$ and $\beta_{a\chi}$ in place of β_d and β_a, respectively.

3.3 Distance and time resolution

3.3.1 One channel

When measuring FRET efficiency on only the donor channel, the total information is

$$J(x) = I_d^{\beta}\frac{36x^{10}\left(1-\beta_d^{-1}\right)^2}{\left(1+x^6\right)^3\left(x^6+\beta_d^{-1}\right)}. \tag{3.23}$$

According to this equation, time resolution is a function of donor-quencher distance. Fig. 3 shows the theoretical minimum 0 observation time period (T, in units of $1/I_d^0$) required to achieve a relative measurement error $\partial R/R_0$ less than a preset value $\alpha = 0.1$. Given α, the theoretically achievable time resolution at various donor-quencher distances can be found under the ideal condition that there are no background photon counts (solid circle in Fig. 3). The time resolution worsens sharply at both large and small x. This is not surprising: the energy transfer efficiency E does not vary much with x for donor-quencher distances that are significantly larger or smaller than R_0 (compare to overlaid FRET efficiency curve in Fig. 3). Consequently, it will take a large number of photons to measure x at these distances to within this error tolerance. At distances closer to R_0 the efficiency is very sensitive to changes in x, so fewer photons are required to obtain the desired tolerance.

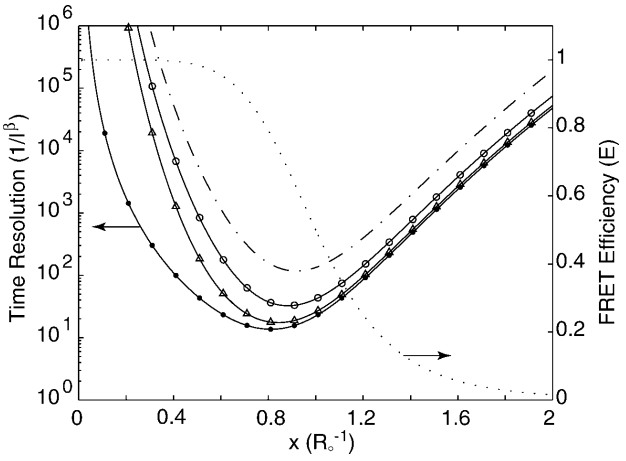


Figure 3.2: Observation time, in units of $(I_d^\beta)^{-1}$ required to achieve a relative measurement uncertainty $\alpha = 0.1$ as a function of normalized donor-quencher distance $\delta R/R_0 < \alpha$. Shown in the figure are expected time resolutions under various signal/background ratios in the donor channel: $\beta_d = 2$ $(- \cdot -)$, $\beta_d = 5$ $(\circ)$, $\beta_d = 20$ $(\triangle)$, which are compared to that under background-free conditions, $\beta_d \to \infty$ $(\cdot)$. Overlaid is a FRET efficiency curve for comparison $(\cdots)$, referenced to the ordinate on the right.

In practice, one cannot avoid recording background photons. In these cases, the curve remains U-shaped, but is shifted to longer observation times. This is because the information of x is degraded by a factor of $x^6(1-\beta^{-1})^2/(x^6+\beta^{-1})$ in the presence of background photons.

These equations can be used to understand the time and distance resolution limits of a single molecule experiment. For example, for a single molecule labeled with a donor-quencher pair that exhibits a Forster radius of 50 Å and whose fluorescence can be measured with a signal/background ratio of 10, the highest time resolution achievable for measuring the donor-quencher distance is achieved at $R \sim 43$Å with $I_d^\beta T = 0.22/\alpha^2$ photons required to achieve the desired accuracy of α. To measure R within a standard deviation of 5 Å, then, one must collect 22 photons.

3.3.2 Two channels

Time resolution in the two-channel detection scheme is also a function of normalized donor-acceptor distance. The total information for this scheme is

$$J(x) = \frac{36x^{10}}{(1+x^6)^3}\left[I_d^\beta T\frac{\left(1-\beta_d^{-1}\right)^2}{\left(x^6+\beta_d^{-1}\right)} + I_a^\beta T\frac{\left(1-\beta_a^{-1}\right)^2}{\left(1+x^6\beta_a^{-1}\right)}\right]. \tag{3.24}$$

A comparison of background-free single-channel and two-channel detection schemes is displayed in Fig. 3.3.2 A. Although both detection schemes behave similarly at short distances, the two-channel detection scheme delivers better performance. This is expected since more

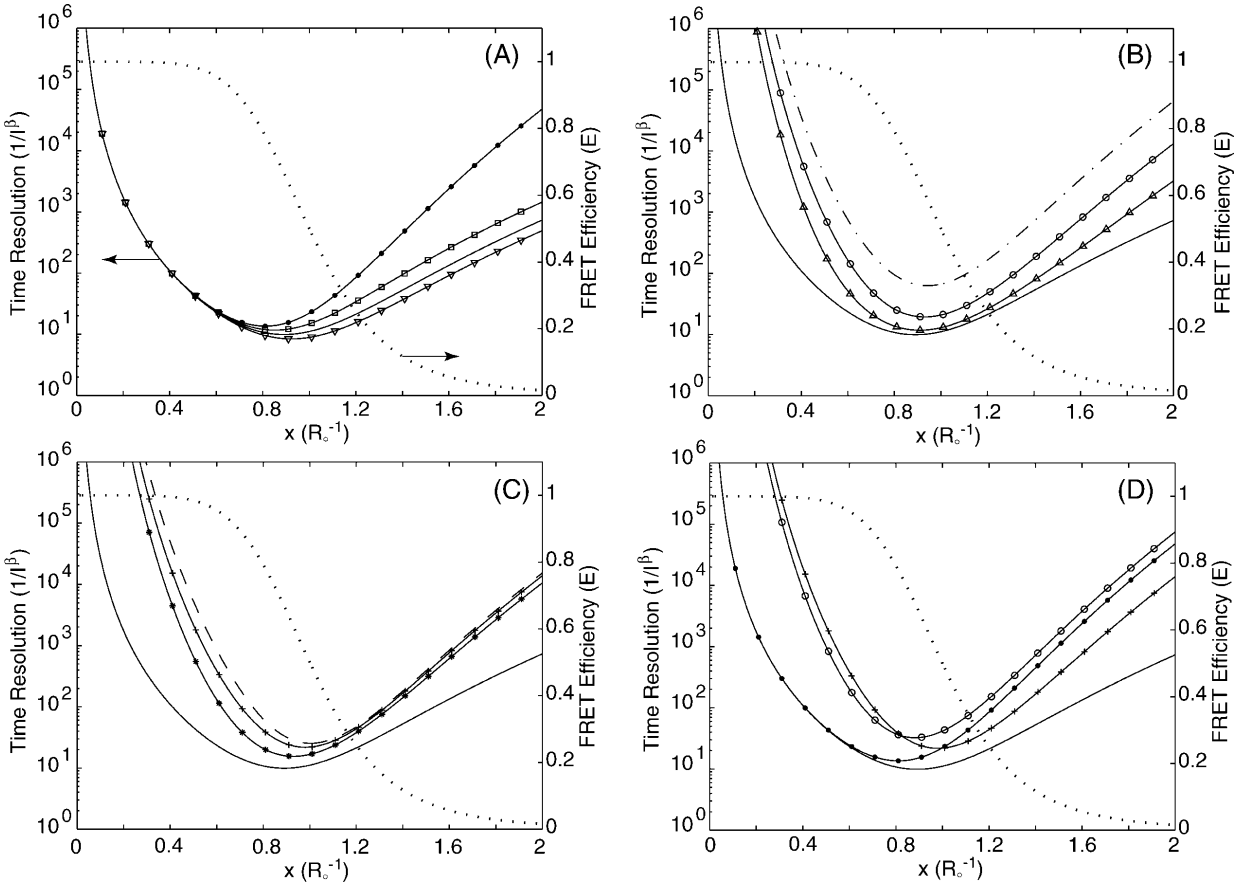


Figure 3.3: Observation time T, in units of $(I_d^\beta)^{-1}$, as a function of normalized donor-acceptor distance $x = R/R_0$, required to achieve a relative measurement error of $\sigma(x) < \alpha = 0.1$ in two-channel detection. (A) Background-free scenario when $\rho = 1.5$ ($\triangledown$), $\rho = 1.0$ (—), and $\rho = 0.5$ ($\square$). Background-free, single-channel detection ($\bullet$) is also included for comparison. (B) Background emission is present, but no cross talk between the donor and acceptor channels. The background levels are $\beta = 2$ ($-\cdot$), $\beta = 5$ ($\circ$), and $\beta = 20$ ($\triangle$). The background-free case is also plotted ($-$) for comparison. In all plots on this panel, the background levels are the same for both the donor and acceptor channels. (C) Both background, $\beta_d = \beta_a = 5$, and and cross talk, $\chi_d = \chi_a = 0$ ($*$), $\chi_d = \chi_a = 0.25$ ($+$), and $\chi_d = \chi_a = 0.5$, are present for two-channel detection. The background- and cross-talk-free, two-channel detection curve (—) is also plotted for comparison. (D) Comparison of background-free single-channel detection ($\bullet$), single-channel with a signal/background ratio of $\beta_d = \beta_a = 5$ ($\circ$), background- and cross-talk-free two-channel detection (—), and two-channel detection with signal/background ratio of $\beta_d = \beta_a = 5$ and cross-talk coefficient of $\chi_d = \chi_a = 0.25$ ($+$). On A-D FRET efficiency E as a function of x is overlaid ($\cdots$) and referenced to the ordinate to the right.

information about x is gathered with two channels. Furthermore, better time resolution can be achieved for larger x in cases where $\rho > 1$, compared to the $\rho = 1$ case where the emission/detection are the same for both donor and acceptor channels. This is because in the $\rho > 1$ cases, the acceptor probe emits more photons than it would have if $\rho = 1$, to give more information about x. Information degradation due to various degrees of background and cross talk in the two-channel detection scenario is depicted in Fig. 4, B and C. Fig. 4 D illustrates a more realistic situation in which the signal/background ratio is 5. The performance of the two-channel scheme is generally better than the single-channel scheme for large x. Using the same example as in the last section, for a single molecule labeled with a donor-quencher pair that exhibits a Förster radius of 50 Åand whose fluorescence can be measured with a signal/background ratio of 10, the highest time resolution achievable for measuring this donor-quencher distance is achieved at $R =\approx 43$Å with $I_d^\beta T = 0.15/\alpha^2$ photons required to achieve the desired accuracy of α. To measure R within a standard deviation of 5 Å, then, one must collect a total of 15 photons.

Chapter 4

Application to Electron Transfer

Excited-state electron transfer has also been used as a probe for investigating conformational changes in individual molecules [39, 72, 112]. In most cases, emission intensity or florescence lifetime of the probe is quenched via electron transfer to or from a nearby quencher. Due to the exponential distance dependence, ET can be used as a spectroscopic ruler to measure distances on the Ångström scale under such conditions that chromophore-quencher distance variation is the sole source for changes in ET rate. These conditions include, for instance, barrier-less excited-state ET so that thermal fluctuation in the relative free-energy levels $\Delta\Delta G$ is negligible, rapidly randomized, or fixed relative to orientation of chromophore and quencher. The timescale of conformational motions in the protein must also be separated from that of probe rotation and facile electron back transfer so that repetitive excitation of a single molecule is achievable. Therefore, ET allows investigation of minute changes of biomolecular conformation [150] and serves as a complementary method to FRET which, as discussed in earlier sections, is sensitive to distance changes on the scale of 20–80 Å. In the following discussion, we assume that ET is primarily dominated by chromophore-quencher distance. That is, the quenching rate k_q is

$$k_q = k_e e^{-\beta_e R_e} = k_e e^{-x_e} \tag{4.1}$$

where k_e is the ET rate when chromophore and quencher are in van der Waals contact, β_e is the distance parameter in ET and varies from 1.0 to 1.4 Å^{-1} for proteins [57, 98], R_e is the edge-to-edge distance between chromophore and quencher, and $x_e = \beta_e R_e$ is the normalized chromophore-quencher distance.

4.1 Fisher information and maximum likelihood estimators

Let the radiative and non-radiative decay rates of a fluorescent probe in its excited state be k_r and k_{nr}, respectively. The total decay rate and emission intensity of the probe in the absence of quenchers is $k_0 \equiv k_r + k_{nr} < k_e$ and I_0, respectively. In the presence of a quencher, the excited-state decay rate becomes $k_x = k_0 + k_q$. If ET is the sole mechanism that increases the excited-state decay rate of a chromophore, the emission intensity of the chromophore is inversely proportional to its excited-state decay rate: $I_x \propto k_x^{-1}$. The total

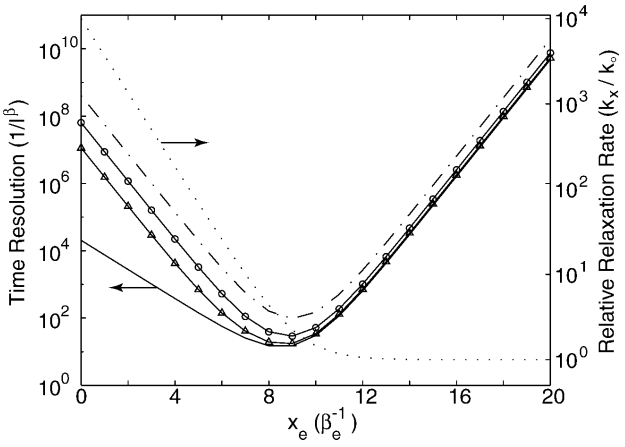


Figure 4.1: Observation time, in units of $(I_0)^{-1}$ under background-free conditions, or in units of $(I_0^\beta)^{-1}$ in the presence of background, required to achieve a relative measurement uncertainty $\alpha = 0.7$ as a function of normalized donor-quencher distance $x_e = R_e\beta_e$. This confidence interval corresponds to an absolute error of ~ 0.5 Åif the distance dependence of electron transfer β_e is 1.4 Å. The value ξ in these plots is set to 10,000, corresponding to $k_e = 10^{13}$ and $k_0 = 10^9$. Shown in the figure are expected time resolutions under various signal/background ratios: $\beta = 2(-\cdot)$, $\beta = 5(\circ)$, and $\beta = 20(\triangle)$, which are compared to that under the background-free conditions, $\beta \to \infty$ (—). Overlaid is excited-state lifetime relative to the quencher-free case for comparison $(\cdots)$, referenced to the ordinate on the right.

detected emission intensity becomes $I_x = I_0(1 + \xi e^{-x_e})^{-1}$, where $\xi = k_e/k_0 > 1$. The x-dependent part of the intensity is thus

$$\zeta(x) = (1 + \xi e^{-x})^{-1} \tag{4.2}$$

Using the Poissonian formula, Eq. 2.11, the Fisher information is calculated to be

$$J(x) = \frac{\left(1 - \beta^{-1}\right)^2 \xi^2 e^{-2x}}{\left(1 + \xi e^{-x}\beta^{-1}\right)\left(1 + \xi e^{-x}\right)^3} I^\beta T \tag{4.3}$$

The MLE for x can be also derived from the Poissonian formula, Eq. 2.14:

$$\hat{x} = \ln \xi + \ln \left[\frac{n - I^\beta T \beta^{-1}}{I^\beta T - n} \right]. \tag{4.4}$$

4.2 Distance and time resolution

The expression for information in an electron transfer experiment, Eq. 40, makes it clear that time resolution T at a given detected photon flux is a function of chromophore-quencher distance x and background level β. The information about x degrades by a factor

of $(1-\beta^{-1})^2/(1+\xi e^{-x}\beta^{-1}) \leq 1$ in the presence of background photons. The condition for measuring x_e to a relative error α is

$$\alpha \geq \frac{\left(e^{x_e}+\xi\right)^{3/2}\left(e^{x_e}+\xi\beta^{-1}\right)^{1/2}}{\xi e^{x_e}\left(1-\beta^{-1}\right)\sqrt{I_0^{\beta}T}}. \tag{4.5}$$

Note that whereas time resolution in general is related to ξ, the ratio of maximum ET rate (k_e) to the excited-state decay rate of the chromophore (k_0), the best possible time resolution is independent of ξ. In fact, the best possible time resolution becomes $I_0T^{opt} = 27/4\alpha^2$ at $x_e^{opt} = \log[\xi/2]$ under the ideal background-free condition when $\beta \to \infty$. In other words, under such ideal conditions, $\sim 6.8/\alpha^2$ photons are needed on average to measure x_e to a relative error α. For example, if one is interested in measuring R_e to an absolute error of 0.5 Åwithin a protein having $\beta_e = 1.4\text{Å}^{-1}$, which corresponds to a relative error $\alpha = 0.5$, $\beta_e = 0.7$, at least ~ 14 photons will be needed in the ideal background-free condition. (See Fig. 5.)

In addition, the optimal chromophore-quencher distance becomes greater in the presence of background photons, but decreases asymptotically to $\log[\xi/2] = \log[k_e/2k_0]$ as $\beta \to \infty$. This suggests that one may choose probes of different fluorescent lifetime k_0 for different systems so the best time resolution will be achieved at an experimentally relevant distance. This is analogous to choosing donor and acceptor pairs by their spectral overlaps for optimal measurements in FRET applications.

Chapter 5

Application to Single Molecule Trajectories

The above analysis focuses on information-theoretic results about photon counting in general. Now the practical application of these results is considered. In this section, an algorithm is given to convert a list of measured photon arrival times to a distance trajectory. This algorithm is general for the experimental methods described and reaches the information theoretic limit. Since it applies equally to all approaches, it is discussed only in the context of two-channel FRET. Similar considerations apply for the measurement of any other parameter that can be understood in the context of the previous theoretical discussion, including single-channel FRET and ET measurements.

5.1 Maximum information data analysis algorithm

The data analysis algorithm that follows is predicated on accurate detection of the arrival times of individual photons. Experimentally, this is typically accomplished by a single photon avalanche photodiode [86]. It is also assumed that the fluorophores used as probes are excited by a light source that provides constant illumination on the timescale of photon detection. The molecules under observation must be well enough separated that any electronic interactions between fluorophores on different molecules may be neglected. Any experimental configuration that satisfies these criteria may be used to generate the single molecule trajectories whose analysis is described.

An algorithm for obtaining distance measurements of predefined precision α is prescribed as follows (see Fig. 5.1). Each measurement consists of a chronological time t, a time uncertainty δt, a distance $\hat{x}$, and a distance uncertainty σ. With the first measurement starting at time $T = 0$, find the minimum block length that will give $\sigma(\hat{x}) \leq \alpha$. Set the chronological time t for that data point to the middle of the time block, time resolution $T(t)$ to its length, and calculate x and s, according to the formulae given above. This algorithm achieves the limit of maximum information. Other termination conditions are possible as well, but only the constant $\sigma(\hat{x})$ method is treated here, as it is the most practically applicable.

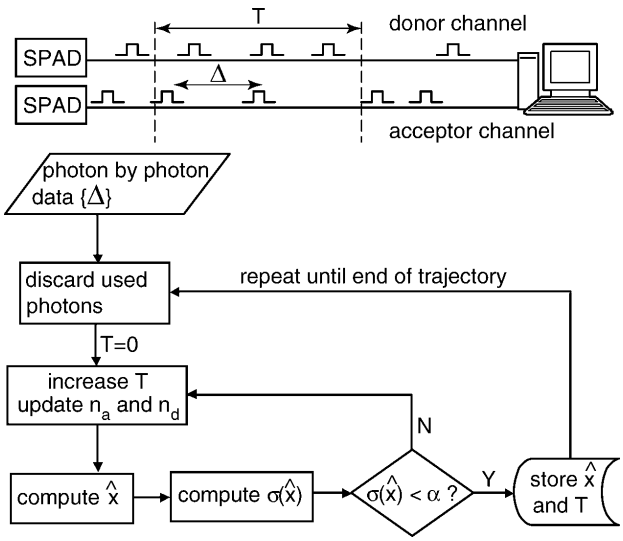


Figure 5.1: Flowchart for the maximum information algorithm.

5.2 Simulation details

This algorithm was validated using simulated single molecule trajectories for which the donor-acceptor coordinate is exactly known as a function of time. Motion on the x-coordinate is modeled according to a discretized Langevin equation in the limit of large, fast friction with potential of mean force $\bar{V}(x)$,

$$x(t+\Delta t) - x(t) = \Delta t \left[-\frac{1}{\gamma}\frac{\partial \bar{V}}{\partial x(t)} + \frac{1}{\gamma}\delta f(t) \right], \tag{5.1}$$

where Δt is the propagation time of the simulation, and γ is a friction coefficient representing velocity-dependent dissipation. $\delta f(t)$ is a Gaussian-distributed random force with $\langle \delta f(t)\delta f(t')\rangle = 2\delta(t-t')\gamma\Delta t k_B\Theta$ in which $\delta(t-t')$ is the Dirac δ-function, k_B is Boltzmann's constant, and Θ is the absolute temperature.

At each time step the emission and inter-system crossing rates are calculated and the system tested to see if either fluorophore has emitted a photon or entered the triplet state. Both photon emission and inter-system crossing are distance-dependent according to the FRET efficiency relation, Eq. 17. Relaxation from the triplet state is treated as exponential in time. If a photon is emitted on either channel, the x-coordinate of the simulation is recorded. Photon data is recorded as inter-photon timings on donor and acceptor channels and subsequently analyzed according to the prescribed algorithm.

5.3 Example trajectories

Evolution of the x-coordinate was simulated according to Eq. 43 on a parabolic potential (see Fig. 5.2). The performance of this algorithm for a sample $x(t)$ trajectory

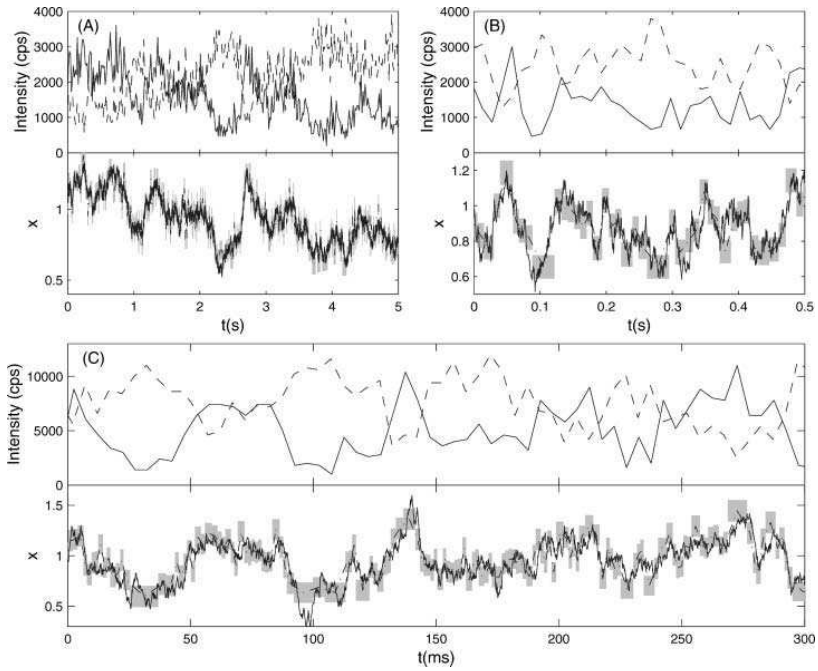


Figure 5.2: Sample FRET trajectories analyzed according to the maximum information algorithm. The top half of each panel shows the simulated intensities on the donor (–) and acceptor (−−) channels as a function of time. The bottom half of each panel compares the analysis of the given FRET trajectory with the simulated trace corresponding to the true trajectory (–). The dashed black line (−·) is the trajectory recovered by the maximum information method, and shaded areas outline the standard deviations calculated from the information analysis. All trajectories were generated according to Eq. 43 on the potential $\bar{V}(x) = 20(x - 0.9)^2$ at a temperature of $\Theta = 1/k_B$. Trajectories A and B were simulated with $\gamma = 10$ and $\gamma = 1$, respectively, $I_d^0 = I_a^0 = 3000$cps and $B_d = B_a = 400$cps, and analyzed with $\alpha = 0.07$. Their intensity trajectories were calculated with 15-ms bins. C was simulated with $\gamma = 0.3$, $I_d^0 = I_a^0 = 10,000$cps, and $B_d = B_a = 1200$cps, and analyzed with $\alpha = 0.1$. Its intensity trajectory was calculated with 5-ms bins.

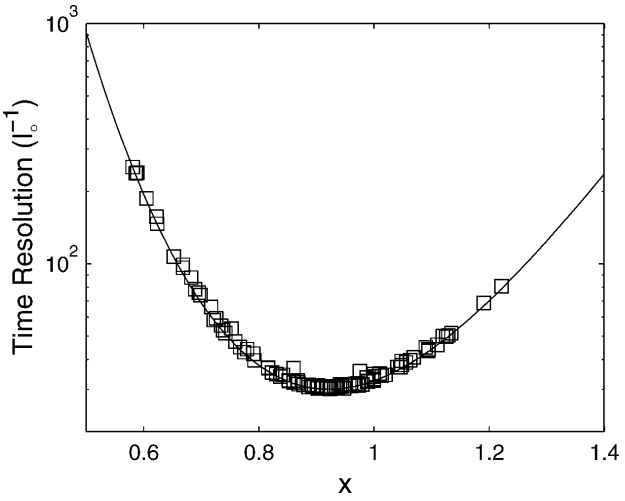


Figure 5.3: The theoretical limit of the time resolution as a function of x, calculated using Eq. 37 with parameters determined by the simulation parameters for the trajectory in Fig. 7 B (–) is compared with the δt values from the maximum information algorithm (□).

is analyzed in Fig. 5.3. Eq. 3.24 gives a lower bound for the time resolution $T(t)$. The maximum information algorithm achieves this lower bound and is thus the optimal data analysis algorithm for extraction of x-trajectories from this kind of data.

Analysis with the maximum information algorithm yields the trajectories shown in Fig. 5.2. Fig. 5.2, A and B, were simulated with $I_d^\beta = I_a^\beta = 3.4$kcps. The maximum time resolution is 8.8 ms (see Fig. 5.3). The maximum information algorithm can detect any conformational changes happening on this timescale or longer.

Since the analysis is primarily based on the number of photons detected, this resolution scales exactly as the inverse of the average count rate. If one were to measure a single molecule FRET trajectory at an experimentally realizable average count rate of 10 kcps, the maximum time resolution would improve to 2.6 ms. Also, the value of the accuracy cutoff, α, makes a significant difference. Analysis of the trajectory in Fig. 5.2 C with $\alpha = 0.1$ improves the maximum time resolution to 1.2 ms.

If donor-acceptor distance fluctuations in the experimental system are large and fast, measured distance trajectories may not represent the full conformational flexibility of the experimental system. In this case, spatial resolution may be sacrificed for time resolution, allowing the full conformational distribution to be observed. In this way the analysis can be tailored to the experimental system and conditions.

Chapter 6

Possible Experimental Complications: Triplets and Bias

6.1 Bias in distance measurements and effects of fluorophore intermittency

The maximum likelihood estimator is asymptotically unbiased when measuring a single distance in the limit of long measurement time. It is, however, slightly biased at small time intervals. In the cases we have studied, this bias is always much lower than the standard deviation given by the Fisher information (see Fig. 9 and numerical studies of bias in the Appendix). If the inherent bias becomes significant, many methods exist to generate estimators that correct for the bias while simultaneously approaching the Cramér-Rao-Fréchet bound [133], or the bias can be corrected empirically by numeric simulations such as those presented in Appendix.

Intensity blinking due to triplet state trapping or other mechanisms has the potential to cause inaccuracies in distance measurements. The algorithm given above does not take these intensity intermittency effects into account. Here we use simulations to show that our algorithm is robust against such blinking behavior up to a few microseconds of non-fluorescent state lifetime. Without loss of generality, we use triplet-state blinking as an example and consider typical dye molecules that exhibit $S_1 \rightarrow T_1$ intersystem crossing quantum yields on the order of $\phi_{isc} = 5 \times 10^{-4}$ and triplet lifetimes of 500 μs [67]. Using these typical values, and assuming a collection efficiency of 5% (giving an effective ϕ_{isc} of 10^{-2}), simulations were carried out at constant x-values.

To evaluate the accuracy of the analysis under these conditions, we use the error parameter $\langle(\delta r)^s\rangle$ as a measure of the closeness of a particular analysis $\{\hat{x}_i\}$ to the true data $x(t)$,

$$\langle(\delta r)^s\rangle = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{x}_i - \frac{1}{\delta t_i}\int_{t_i-\delta t_i/2}^{t_i+\delta t_i/2} dt' x(t')\right]^s, \tag{6.1}$$

where $\hat{x}_i$ is the maximum-information estimate of $x(t)$ at time t_i and N the total number of estimates at given α. Therefore, bias (mean error) is represented by $s = 1$ and mean-square error by $s = 2$. Fig. 6.1 shows the results of this analysis.

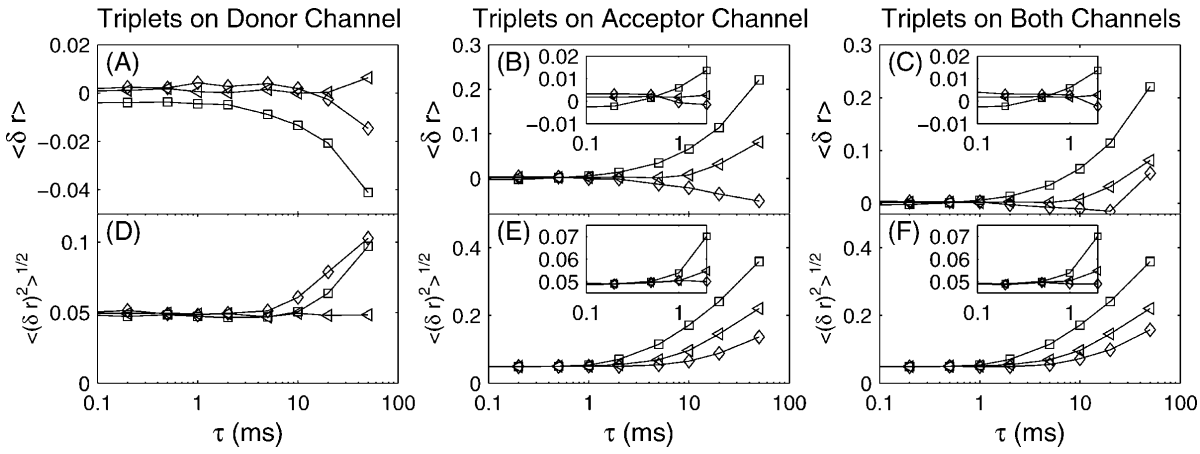


Figure 6.1: (A–C) Mean error and (D–F) root mean-square error as a function of triplet lifetime (Eq. 44). Only donor triplet states are allowed in A and D; only acceptor triplet states in B and E; and both acceptor and donor triplet states are allowed in C and F. Trajectories were simulated at constant x-values of 0.8 (□), 1.0 (◁), and 1.2 (◇)) with effective $\phi_{isc} = 1 \times 10^{-2}$. The trajectories were analyzed by the maximum information algorithm with a $\alpha = 0.05$. Under typical experimental conditions the triplet lifetime t will not exceed 1-2 ms. At these lifetimes, there is no significant effect on the accuracy of the algorithm.

For experimentally relevant triplet lifetimes, there is no significant error. This is because, although the analysis is done on a photon-by-photon basis, generally 10 or more photons are included in each box. As long as the triplet lifetime and intersystem crossing quantum yield are such that the total time spent in the triplet state is not a large fraction of the width of the box, the analysis will not be adversely affected.

In the case that only donor triplets are allowed, there is no significant increase in bias due to the triplet state even at very long triplet lifetimes. This is due to the multinomial nature of the analysis–the time taken to acquire photons is not important, only the channel they arrive on. Since donor triplets states prevent photon emission from both the donor and the acceptor, the only effect is that the time resolution will be decreased.

This is not the case with acceptor triplet states. Bias due to acceptor triplet states is very distance-dependent. At large x, even with very long acceptor triplet lifetimes, there will be no effect on the accuracy of distance measurements: the acceptor simply will not enter the triplet state. As x decreases, the probability that the acceptor will enter the triplet state increases, and the bias becomes more significant.

6.2 Experimental calibration

The formulas presented here are ready for immediate use in many experimental setups that have already been reported in the literature. In addition, these results are

applicable to other, time-independent measurements. For example, the maximum likelihood estimators for determining distance from the numbers of photons measured on the donor and acceptor channels can be used in any situation where all of the calibration numbers (B_d, B_a, I_d^0, I_a^0, χ_d, and χ_a) are known. It is important to note that, although formulas regarding FRET efficiency are also given in this part, the use of FRET efficiency as an indicator of molecular state may be misleading as it is not a linear function of donor-acceptor distance; small distance changes may be amplified as a result.

The number of photons that may be collected from a single molecule is heavily dependent on experimental conditions and the particular fluorophores used. On the order of 10^8 photons may be collected from a single molecule of Rhodamine 6G under vacuum in a poly (methylacrylate) film at low excitation intensities [31]. This means that on the order of 10^6 independent distance measurements may be made if the molecule is part of a FRET pair between $0.4R_0$ and $1.6R_0$. To study fast dynamics, one may wish to excite the single molecules at higher intensities, but the detected time trace will also be shorter. Currently, for example, using the Alexa-555/Alexa-647 dyes (Molecular Probes), photon arrival rates on the order of 5×10^4 photons per second are experimentally feasible in water solution, giving time resolutions better than 1 ms (L.P.W. and H.Y., unpublished data).

Several detection methods exist that are compatible with the maximum information method. Avalanche photodiodes, photomultiplier tubes, and multichannel plates may all be used to detect and count single photons. The dark counts on high quantum efficiency (QE $> 60\%$) Peltier-cooled, avalanche photodiodes range from 25-500 cps. The dark counts on uncooled, single photon-counting photomultiplier tubes and multichannel plates (QE $< 20\%$) are similar, ranging from 10-1500 cps. At high excitation intensities, the detection device is not the primary source of background. Dark counts from the detector are minuscule compared to other sources including Raman scattering and autofluorescence in cells, both of which are difficult to suppress using spectral filters. These background contributions, however, do play a role in determining the lowest possible excitation intensities and thus the longest possible trajectories. The maximum information method will work at almost any signal/background ratio; but as this ratio approaches 1, the required number of photons to make a particular distance measurement increases without bound. It should be stressed that, although the proposed method has in mind the use of detectors capable of single photon counting, the concepts and ideas that underlie the development of maximum information method is general and should be applicable to any measurements that are information-limited.

The preceding discussions suggest that, when choosing a FRET dye pair to measure a distance of $\sim R$, the most effective dyes will be those that exhibit an R_0 of $\sim 1.1 - 1.3R$, instead of the commonly used $R \approx R_0$ condition. The options available in dye and filter selection are much broader. Since cross talk and cross-excitation are now recognized as merely contributions to the background, the filter set may be chosen with this in mind. Bandpass filters may be made as wide as possible to collect as many photons as possible. Also, the excitation wavelength may be chosen to be the maximum absorbance of the donor, even if that would generate some direct acceptor excitation.

Part II

Detecting Intensity Changepoints

Intensity change points, discrete jumps in detected intensity between periods of constant intensity, occur in many systems and arise from processes that occur 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 [101, 99, 79, 76]. In biological macromolecules, single-molecule fluorescence studies reveal sudden jumps in detected fluorescence intensity during enzymatic reactions and structural changes [32, 89, 157, 129, 155, 156, 33]. 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. [60, 131, 94, 14, 132, 6, 141, 7, 125, 87].

Binning and thresholding is the simplest and most commonly used technique for the analysis of intensity change points, but it is highly problematic. 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 Fig. 6.2). This problem is exacerbated when the signal-to-background ratio is low. The simulated 5-state single-molecule trace in Fig. 6.2 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 while small and slow transitions are still lost in the noise. These concerns emphasize the need for quantitative, statistically robust methods of analysis. Indeed, while much progress has been made in experimental techniques [96, 78], there is still 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[3]. 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

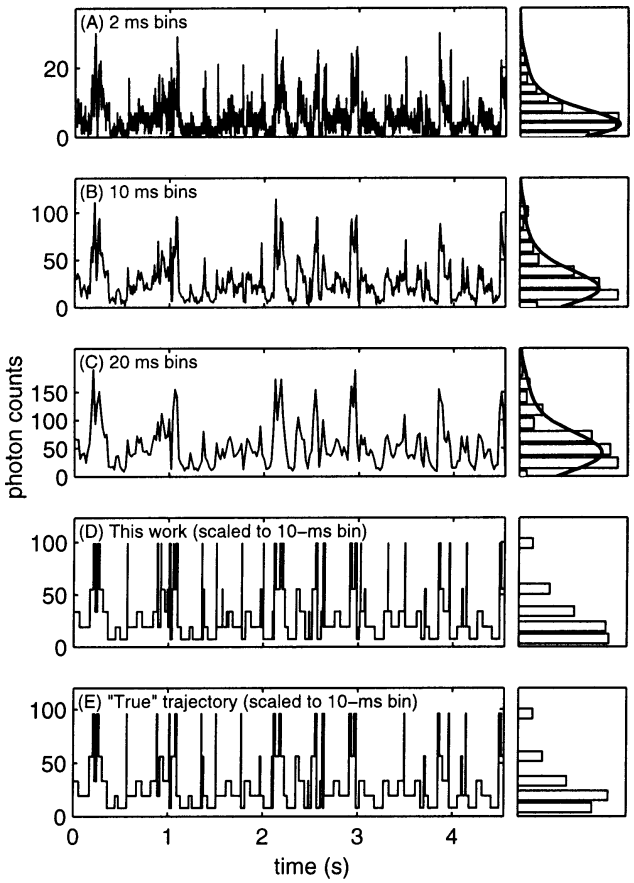


Figure 6.2: (A) 2-ms binned trajectory. (B) 10-ms binned trajectory. (C) 20-ms binned trajectory. (D) Photon-by-photon reconstructed single-molecule intensity states. (E) "True" intensity states used in simulating panels A–C. The intensity distribution histograms are shown to the right of each panel, 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 Chapter 10.2.

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. Based on 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 part 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 Fig. 6.2D. The method developed allows quantitative recovery of the underlying intensity levels and change points in a single-molecule measurement. In the first step, described in Chapter 8, I develop a generalized likelihood ratio test to identify intensity changepoints. 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. The trajectory is recursively segmented until no more change-points are detected. Thus, both the *number* and the *location* of intensity changes are determined.

Chapter 9 develops a method to group the detected states. The n change points define $n + 1$ possible 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 possibility 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 assigned to the same state, resulting in a statistical hierarchy of intensity levels. This over-clustering allows application of the Bayes Information Criterion to determine the true number of intensity states in the system.

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 performance of this method is evaluated using computer simulations and discussed in Chapter 10.

Chapter 7

Intensity Estimation

Though the general ideas outlined in this part are applicable to most time-resolved single-molecule measurements, the discussion focuses on the use of photon-by-photon detection/registration methods since they provide the highest time resolution. We start with a brief review of the information-theoretical concepts introduced in Part I. Fig. 7.1A 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

$$f(\Delta_i; I_j) = I_j e^{-I_j \Delta_i}. \tag{7.1}$$

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,

$$g(n_j; I_j, T_j) = \frac{(I_j T_j)^{n_j}\, e^{-I_j T_j}}{n_j!}. \tag{7.2}$$

The maximum likelihood estimate (MLE) of the intensity is $\hat{I}_j$ that maximizes the likelihood function Eq. 7.2 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 [44] 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

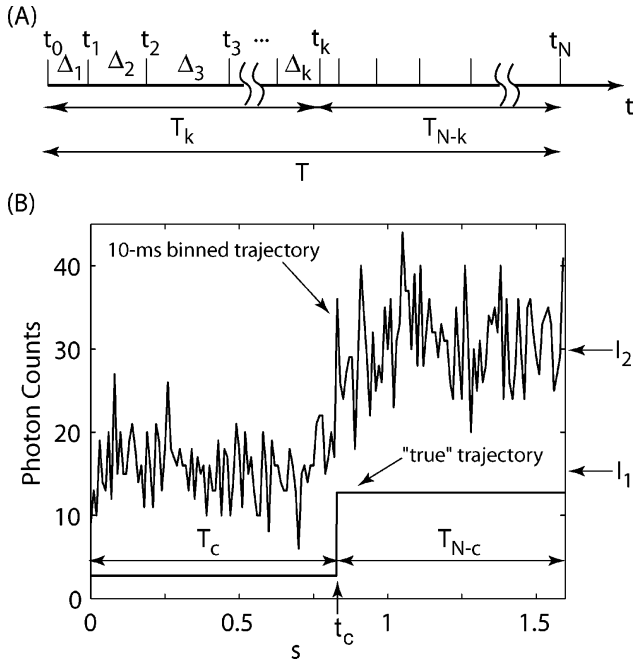


Figure 7.1: (A) A typical photon-by-photon recording scheme that marks the chronological photon arrival times $\{t_i\}$, the inter-photon time durations $\{\Delta_i\}$, emission intensity change points $\{c_k\}$, and the time durations $\{T_k\}$ for which the single molecule maintains the same emission 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.

molecule is in a state with detected intensity I. The Fisher information is given by,

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

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 [49, 28, 110],

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

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

$$J_{\mathrm{Fisher}}(\hat{I}_j) = \left\langle (\partial g(n_j; I_j, T_j)/\partial I_j)^2 \right\rangle_{n_j} \Big|_{I_j \to \hat{I}_j} = T_j/\hat{I}_j = T_j^2/n_j.$$

The best attainable variance of I_j is then given by

$$\mathrm{var}(\hat{I}_j) \geq J_{\mathrm{Fisher}}(\hat{I}_j)^{-1} = \frac{n_j}{T_j^2}. \tag{7.3}$$

Eq. 7.3 allows a quantitative assessment of the uncertainties with which a single-molecule intensity time trajectory is measured. For example, application of Eq. 7.3 shows that only 100 photons must be collected to measure an intensity to within a relative standard deviation of 10%.

Chapter 8

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 [24]. 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 [144, 145]. 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.* [17] 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 corresponding to 31% error rate simply as a reference point.

8.1 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. Fig. 7.1B). 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 $H_\circ$, 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. 7.2.

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

$$\begin{aligned} L_A &= g\left(k; \hat{I}_1, T_k\right) g\left(N-k; \hat{I}_2, T_{N-k}\right) \\ &= \frac{\left(\hat{I}_1 T_k\right)^k e^{-\hat{I}_1 T_k}}{k!} \frac{\left(\hat{I}_2 T_{N-k}\right)^{N-k} e^{-\hat{I}_2 T_{N-k}}}{(N-k)!} \end{aligned}$$

The intensities $\hat{I}_1$ and $\hat{I}_2$ are estimated from the trajectory as discussed in chapter 7, since they are not known *a priori*. Hypothesis H_A must be compared with the null hypothesis $H_\circ$, which states that there is no change point at time t_c.

$$H_\circ : I(t_1) = I(t_2) = \cdots = I(t_N) = I_0$$

$$\begin{aligned} L_\circ &= g\left(n_1; \hat{I}_0, T_k\right) g\left(n_2; \hat{I}_0, T_{N-k}\right) \\ &= \frac{\left(\hat{I}_0 T_k\right)^k e^{-\hat{I}_0 T_k}}{k!} \frac{\left(\hat{I}_0 T_{N-k}\right)^{N-k} e^{-\hat{I}_0 T_{N-k}}}{(N-k)!} \end{aligned}$$

where I_0 is the average intensity of the entire trajectory.

To compare the likelihood of H_A over $H_\circ$, we compute the log likelihood ratio of the two hypotheses,

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

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

$$\mathcal{L}_k^\circ = 2k \ln \frac{k}{V_k} + 2(N-k) \ln \frac{N-k}{1-V_k} - 2N \ln N \tag{8.1}$$

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.

$$Z_N^\circ = \max_{1 \leq k \leq N} \{\mathcal{L}_k^\circ\} \tag{8.2}$$

This test, while identifying the most likely k for a change point, does not prove that a change point is present. To assess whether or not 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 $\mathcal{L}_k$, 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

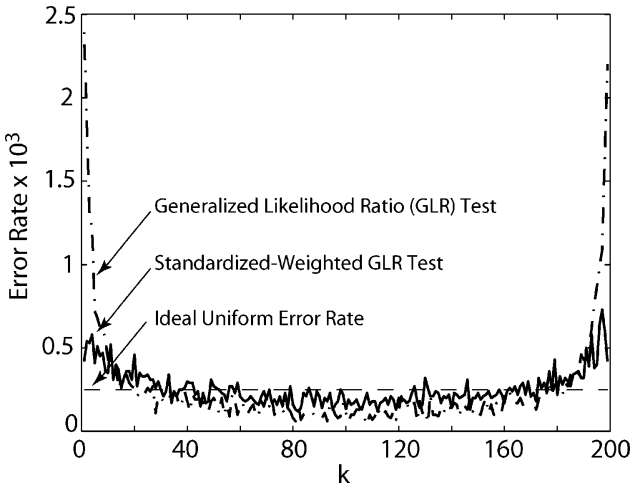


Figure 8.1: Error rates as a function of the test point k under the null $H_\circ$ hypothesis in which there is no change point. In this simulation, a total of 100,000 traces of 200 exponentially distributed random number were generated and analyzed. The type-I error rate α is set to 0.05 in the analysis.

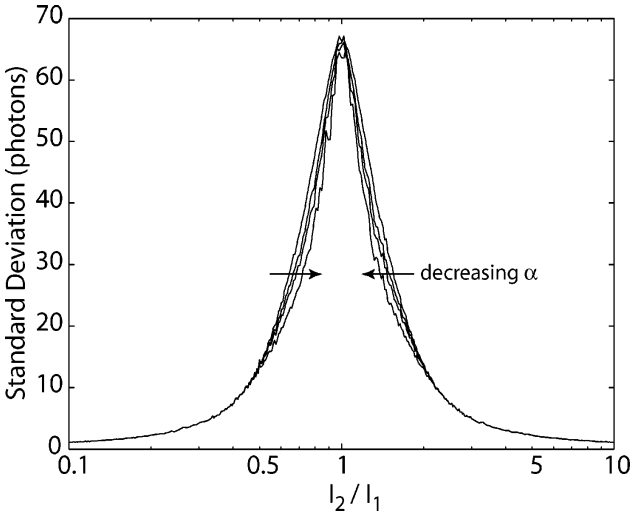


Figure 8.2: Empirically determined standard deviation of detected intensity change point as a function of the size of intensity change under the alternative H_A hypothesis in which there is exactly one change point. In this simulation, a total of 100,000 traces of 200 exponentially distributed random 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.

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 [102]. Computational details for the present application are described in Appendix B.

Eq. 8.2 does not have a uniform error rate across the entire trajectory because MLE values of emission intensities are used in the likelihood ratio tests [146, 66]. 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 Fig. 8.1. The non-uniformity in error rate can be alleviated by standardization and weighting of the likelihood ratio function, as originally proposed by Henderson [66]. In this implementation, which still relies on the fundamental Poisson statistics discussed above, the null hypothesis is rejected when the test statistic becomes

$$Z_N = \max_{1\le k\le N} \left\{\bar{\mathcal{L}}_k + W_k\right\} \equiv \max_{1\le k\le N} \left\{\mathcal{L}_k\right\} \ge \tau_{1-\alpha}, \tag{8.3}$$

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

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

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

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

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

$$\Pr\left(\mathcal{L}_k < \tau_{1-\alpha}; k = 1,\ldots,N|T\right) = 1-\alpha, \tag{8.4}$$

can also be computed using Noé's algorithm as outlined in appendix B and are listed in Table 8.1. Eq. 8.3 gives a more uniform empirical error rate as shown in Fig. 8.1. For this reason, Eq. 8.3 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. 8.2 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 Fig. 8.2.

Type-I error alone does not completely characterize intensity change point detection, because it says nothing 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 Fig. 8.3. Our simulation indicates that for a sample size of 200 data points with one change point, at

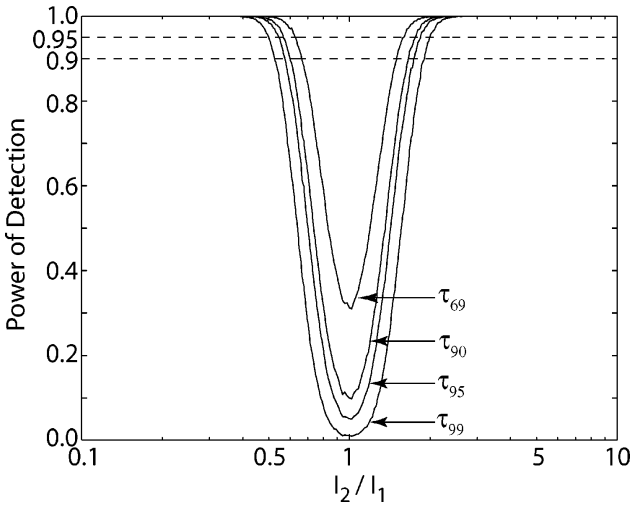


Figure 8.3: Detection power as a function of the size of intensity change under the alternative H_A hypothesis in which there is exactly one change point. In this simulation, a total of 100,000 traces of 200 exponentially distributed random number were generated and analyzed with the change point occurring at the 100-th point. The critical regions for type-I error used in the analysis are indicated by $\tau_{1-\alpha}$, where $1-\alpha = 0.69, 0.90, 0.95, 0.99$. 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 $\sim$1.5 for $\alpha = 0.31$, $\sim$1.66 for $\alpha = 0.1$, $\sim$1.72 for $\alpha = 0.05$, and $\sim$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).

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

Table 8.1: Level α and β points for intensity change detection using Eqs. (8.3) and (8.6).

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 [145]. 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}.

$$C_{1-\beta} = \left\{k : \Pr\left(\mathcal{L}_m < \tau'_{1-\beta} | T_k\right) \Pr\left(\mathcal{L}_{m'} < \tau'_{1-\beta} | T - T_k\right) \leq 1 - \beta\right\}, \tag{8.5}$$

where $m \in \{1 \ldots k\}$ and $m' \in \{k+1 \ldots N\}$. 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 [146] conducted numerical simulations to evaluate Eq. 8.5 for exponential distributions over all k, T_k and T at fixed N and found that the maximum occurred 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,

$$\widetilde{C}_{1-\beta} = \left\{k : Z_N - \mathcal{L}_k \leq \tau'_{1-\beta}\right\}. \tag{8.6}$$

That is, plotting $\mathcal{L}_k$ as a function of k, the β confidence region consists of photon indices $\{k\}$ that give $\mathcal{L}_k$ values that are greater than $Z_N - \tau'_{1-\beta}$. Calculation of $\tau'_{1-\beta}$ (cf. Table 8.1) is also carried out using Noé's algorithm, detailed in appendix B.

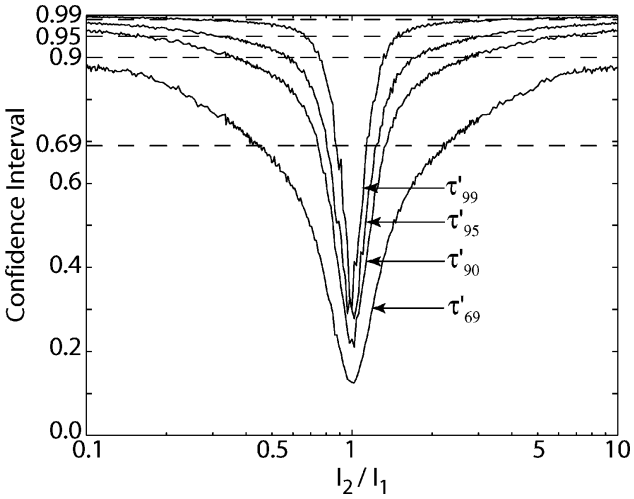


Figure 8.4: Empirical evaluation of the accuracy of confidence intervals calculated from Eq. 8.6. The solid curves represent the probability that the detected change point lies within the calculated confidence interval. In this simulation, a total of 100,000 traces were generated of 200 exponentially distributed random numbers with the change point occurring at the 100-th 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.

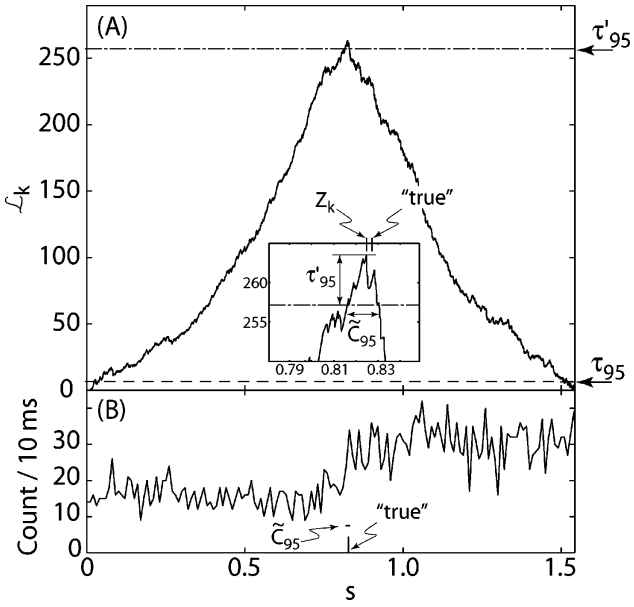


Figure 8.5: (A) Plot of $\mathcal{L}_k$ as a function of the chronological photon arrival time of each detected photon using the data from Fig. (7.1). The critical region was found to be $\tau_{95} = 6.343$ $(--)$ given $\alpha = 0.05$. The change point was found to be $c = 1301$ ($t_c = 0.8247$ s), 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 Fig. 7.1B for comparison. The "true"location of intensity jump and inferred 95% confidence interval $\widetilde{C}_{1-\beta}$ are indicated as vertical and horizontal bars, respectively.

To characterize the performance of the approximation, Eq. 8.6, we have used computer simulations to evaluate its accuracy. The results, summarized in Fig. 8.4, suggest that for high-power change point detections (power > 0.9), the critical region defined by Eq. 8.6 is indeed a conservative overestimate of the true confidence interval. For lower-power change point detection, however, Eq. 8.6 underestimates the confidence interval.

Eqs. (8.3) and (8.6) 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 Fig. 8.5 using the time trace in Fig. 7.1B as example. As illustrated in Fig. 8.5 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.

8.2 Implementation

In general, a single-molecule time trajectory may contain multiple intensity change points. To find all the significant change points, we use a recursive binary segmentation

algorithm. First, a change point is found by applying Eq. 8.2 to the entire single-molecule time trajectory. The maximum of the log-likelihood ratio, Eq. 8.3, is considered as a change point and its confidence interval is found using Eq. 8.6. 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. Fig. 10.1A). This procedure is repeated recursively until no further change points are found.

Eqs. (8.3) and (8.6) 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 Fig. 6.2 containing multiple change points is used to illustrate the application of the binary segmentation algorithm. Once the change points are located, the intensity levels can be determined using the maximum likelihood estimator ($\hat{I}_j = n_j/T_j$) discussed in Chapter 7. Fig. 8.6 displays a reconstructed intensity trajectory for the data shown in Fig. 6.2. A total of 125 statistically significant change points are found, producing 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.

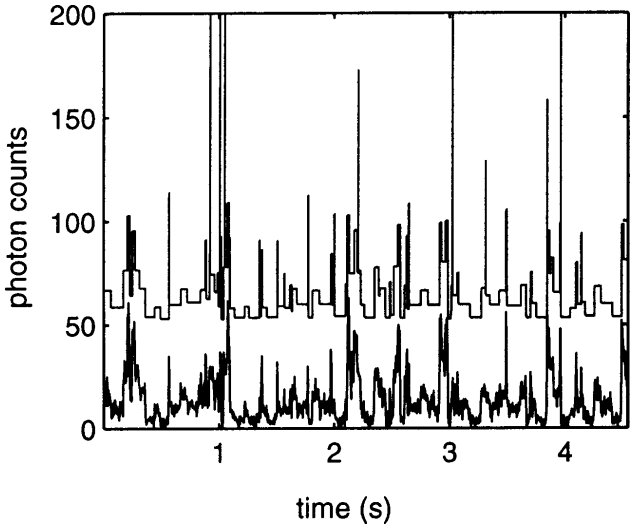


Figure 8.6: Single-molecule emission time trace (offset by 50 and scaled to a 5-ms bin time) inferred from the simulated data in Fig. 6.2 using generalized likelihood ratio test. The 5-ms binned raw data are also shown as a comparison. The intensity bursts seen in the inferred emission time trace are examples of "shot" counting noise arising from short resident time at a given state.

Chapter 9

Determination of the Number of States and their 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 power if more states are involved or if the intensity levels are not well-separated, as in the example of Fig. 6.2. The above-discussed photon-by-photon change point determination greatly 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 Fig. 8.6; 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 [47], intensity levels—determined according to the maximum likelihood estimator—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 m-th 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.

9.1 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

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

where G is the number of distinguishable intensity levels and γ_i is the classification, or grouping, of the i-th photon into the γ_i-th intensity group. The change-point detection procedure discussed in Section 8 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 $\{n_j/T_j^2\}$, $j = 1 \ldots J+1$ (cf. Eq. 7.3). In this classification scheme, the data set $\{\Delta_i\}$ is formed by sampling n_j observations (photons) separately from the j-th intensity component. Under this scheme, the classification log-likelihood function takes the form:

$$\mathcal{L}_c\left(\{\Delta_i\}; I_1, \ldots, I_G\right) = \sum_{j=1}^{G} \sum_{i \in \mathcal{J}_j} \ln f(\Delta_i; I_j), \tag{9.1}$$

where G is the number of groups and $\mathcal{J}_j$ 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 [136].

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 [48]. The established formulations, however, are not directly applicable to the analysis of single-molecule data.

To treat the problem consistently with change-point detection (cf. Chapter 8), 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 [120]. Here the exponential distribution of Eq. 7.1 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 m-th and the j-th groups. Using Eq. 9.1, the log-likelihood ratio merit function is

$$\begin{aligned} M_{\langle j,m \rangle} &= \mathcal{L}_c\left(\{\Delta_i\}; I_1, \ldots, I_G\right) - \mathcal{L}_c\left(\{\Delta_i\}; I_1, \ldots, I_{\langle j,m \rangle}, \ldots, I_{G-1}\right) \\ &= (n_m + n_j) \ln\left[\frac{n_m + n_j}{T_m + T_j}\right] - n_m \ln\left[\frac{n_m}{T_m}\right] - n_j \ln\left[\frac{n_j}{T_j}\right]. \end{aligned} \tag{9.2}$$

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

$$n_{jm} = n_j + n_m, \tag{9.3}$$

$$T_{jm} = T_j + T_m. \tag{9.4}$$

Initialize $\bar{p}_{mj}$ according to Eq. 9.5.
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 9.1: An EM algorithm for single-molecule intensity level clustering.

Eqs. (9.2), (9.3), and (9.4) 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.

9.2 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 $\{I_1, \ldots, I_{J+1}\}$ intensity levels into $G = 1 \ldots G_{\max}$ groups, as well as the estimation of $\{\hat{I}_1, \ldots, \hat{I}_G\}$, is accomplished by the Expectation Maximization procedure of Dempster, Laird, and Rubin [30]. 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 m-th group. Then the initial guess for the expectation-maximization procedure is based on the result of the agglomerative

hierarchical grouping.

$$p_{mj} = \begin{cases} 1 & \text{if } I_j \text{ belongs to the } m\text{-th group,} \\ 0 & \text{otherwise.} \end{cases} \tag{9.5}$$

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 $\{I_1, \ldots, I_{J+1}\}$ is

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

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

$$\mathcal{L}_{\text{em}}\left(I_j; p_{mj}, p_m, I_m, T_m\right) = \sum_{j=1}^{J+1} \sum_{m=1}^{G} p_{mj} \ln\left[p_m g(n_j; I_m, T_j)\right] . \tag{9.6}$$

For the model in Eq. 9.6, p_{mj} is calculated as $\bar{p}_{mj} = E\left[p_{mj} | \{\hat{I}_j\}; T_m, I_m\right]$ 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 j-th intensity segment may come from one of the $m = 1$ to G intensity states. Under the multinomial framework, the probability for the j-th segment to come from the m-th intensity state is then as displayed in Fig. 9.1.

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. 9.6. An EM algorithm for clustering single-molecule intensity levels is outlined in Fig. 9.1.

9.3 Determination of number of states using the Bayesian information criterion (Schwarz's criterion)

While 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 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 [80]. 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

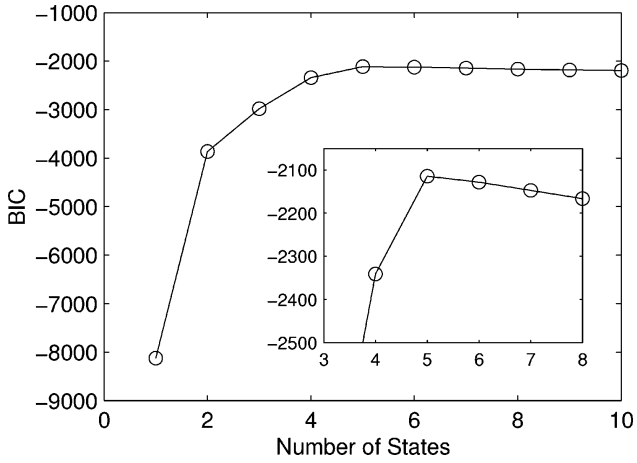


Figure 9.2: BIC as a function of number of groups, showing that there are most likely 5 distinct intensity states. The numerical example in Fig. 6.2 is used.

states that is at least as large as the correct one [85]. Furthermore, as the number of data point increases, it asymptotically converges to the correct number of states. The integral to be maximized is

$$\begin{aligned} p\left(\{\Delta_i\}, m\right) &= p_m\left(m\right) \int_{\Theta_{em}} \int_{\Theta_{cp}} p_l\left(\{\Delta_i\} | \theta_{cp}, \theta_{em}, m\right) \\ &\times p_p\left(\theta_{cp} | \theta_{em}, m\right) p_p\left(\theta_{em} | m\right) d\theta_{em} d\theta_{cp} \end{aligned}$$

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 [80, 119], giving

$$\ln p\left(\{\Delta_i\}, m\right) \approx 2\mathcal{L}_G - \left(2n_G - 1\right) \ln N_{cp} - N_{cp} \ln N, \tag{9.7}$$

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. 9.6. Eq. 9.7, 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 Fig. 6.2 as an example, Fig. 9.2 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 Fig. 6.2D is a reconstructed single-molecule trace using the numerical data in Fig. 6.2. 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 Fig. 6.2E. As clearly demonstrated in Fig. 6.2, our method based on quantitative statistical analysis is superior to the heuristic, but popular, binning-thresholding approach.

Chapter 10

Characterization

10.1 Resolving power of change point determination

As implied in Figs. 8.3 and 8.4, 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 chapter 7. 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. Fig. 10.1A).

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 Fig. 10.1B–E. As an example, for very small intensity jumps ($I_2/I_1 = 1.5$, Fig. 10.1B) 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 Fig. 10.1E, 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.

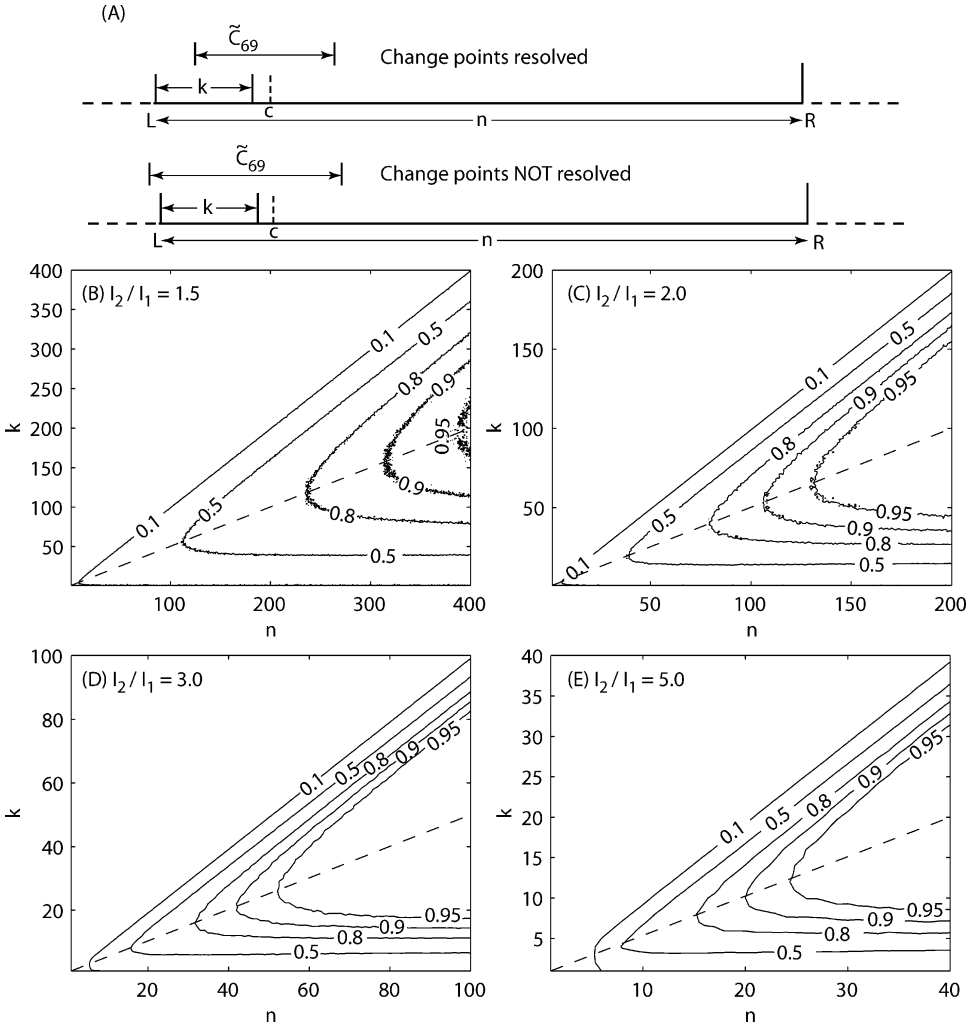


Figure 10.1: Characterization of resolving power by simulation. (A) Illustration of resolving power with the use of the binary segmentation algorithm. L and R are respectively the left and right end points of a data segment of length n which contains a true change point k photons from the left end. A detected change point c is considered resolved if the conservative 69% confidence region $\widetilde{C}_{69}$ 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.

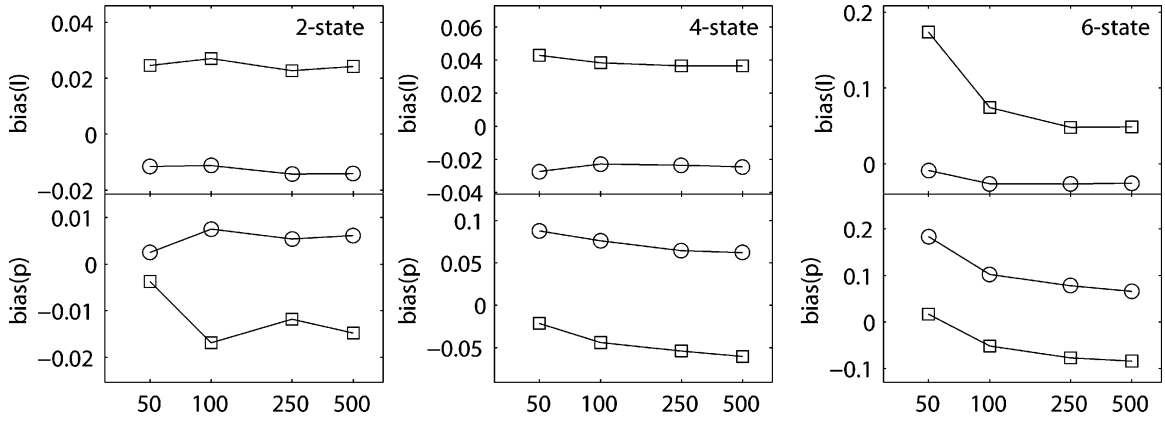


Figure 10.2: Relative bias in estimation of intensity (I) and occupancy (p) for simulations involving 2, 4, and 6 states as discussed in the text. Traces are shown for the states of highest (□) and lowest (∘) intensity in each set, representing the extremes in bias.

10.2 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 10000 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 states and 50, 100, 250, or 500 change points. 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 Fig. 10.2. 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. While 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.

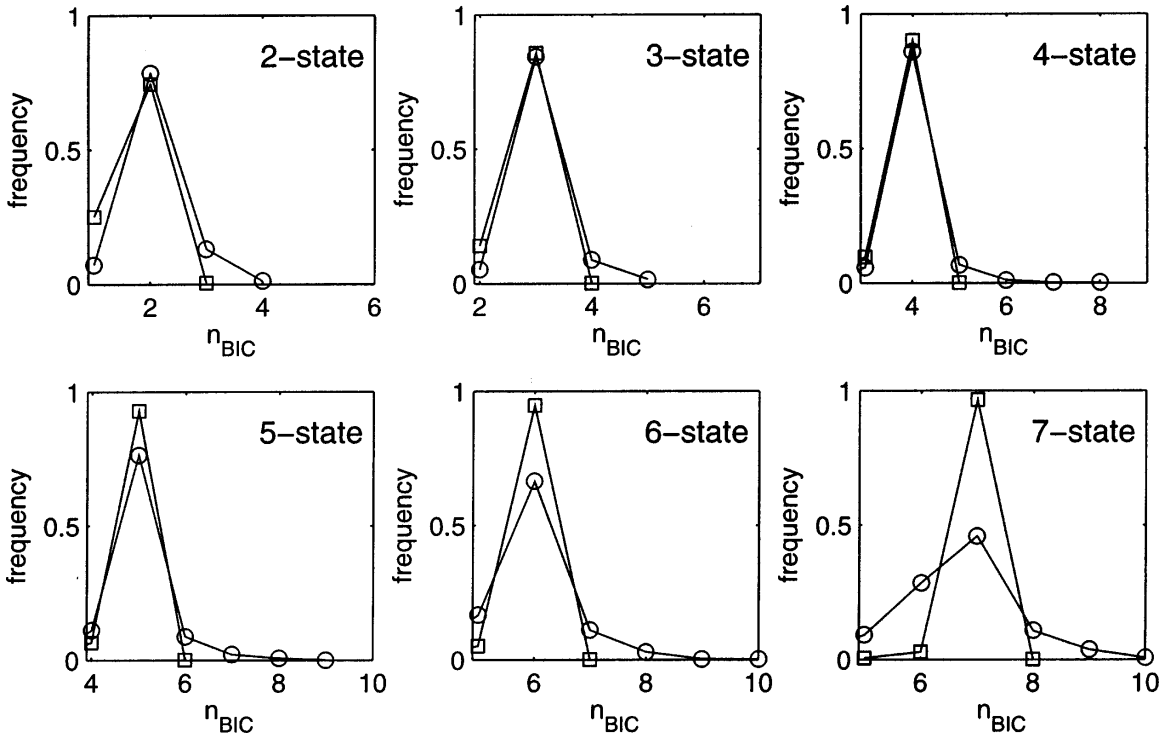


Figure 10.3: Histogram of BIC predictions of numbers of states. Each panel summarized results from simulations with a set number of intensity states. The true number of states is found in the top-right corner of each plot. Predictions of the number of states were made on the basis of trajectories including 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.

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 Fig. 10.3. 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[85] are not completely satisfied 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.

Part III

Miscellaneous Statistical Methods

Knowledge of the free energy surface on which a biological macromolecule resides allows a quantitative understanding of phenomena ranging from folding to catalysis. Its features give important clues to the dynamic structure-function relationship. In addition, accurate experimental characterizations of the free energy surface under physiological conditions provide stringent constraints for tests of theoretical models. These include the identification of conformational species, the determination of their relative population and the heights of the barriers that separate them, and the characterization of their structural flexibility, as indicated by the width of distribution. The distribution of molecular conformations, an experimentally coarse-grained manifestation of the free energy surface, can in principle be directly measured using single-molecule fluorescence spectroscopy [140, 97, 147].

In determining biomolecular conformational distributions from single-molecule measurements, the experimentalist is faced with the statistical uncertainties associated with low-light detection as well as with other measurement errors [73, 88]. This is exacerbated in time-dependent measurements where one relies on only a few photons to determine a molecular parameter. To illustrate the challenges in this area, Fig. 10.4 compares histograms constructed using a commonly adopted approach with the true probability distribution function (PDF) from a simulated single-molecule FRET trajectory. To construct these distributions, one first chooses a time period with which to bin the trajectory, computes the distance value (or FRET efficiency) in each time bin, and chooses an interval in which to bin the distance measurements. As illustrated in Fig. 10.4, both the choice of time bins and the choice of distance bins will affect the shape of the distribution, potentially impacting the interpretation of experiments. While the choice of time bins has been discussed previously, [137, 138] the choice of distance bins represents yet another obstacle towards realizing the full potential of single-molecule spectroscopy, measurement of the distribution of molecular properties. This article seeks to address this issue by developing a comprehensive method for the extraction of probability distribution functions from single-molecule measurements.

When a trajectory is treated using the Maximum Information Method (MIM) each data point has the same statistical significance [137]. This is superior to equal-time binning, where different data points may have wildly different variances. MIM treatment can thus be regarded as equal-information binning along the time trajectory. This is advantageous in probability density (histogram) estimation and is critical to the use of the maximum entropy method (MaxEnt) [71] for removal of the broadening of the histogram that occurs due to photon-counting statistics. The MaxEnt approach is employed because it offers an unprejudiced framework for extraction of the molecular conformational distribution, constrained by available information and known experimental uncertainties. Prior knowledge about the molecular system can be easily included with proper statistical weighing. When little is known about the density distribution, which usually the case at the single-molecule level, the MaxEnt approach allows quantitative recovery of the underlying distribution without assuming any models or shapes for the unknown probability density function. This is consistent with our previous development of information-based, model-free approaches to analysis of fluorescence single-molecule data [137, 138]. To evaluate the accuracy of the deconvolved functions, we have also derived analytical expressions for the covariance matrices of the measured PDF.

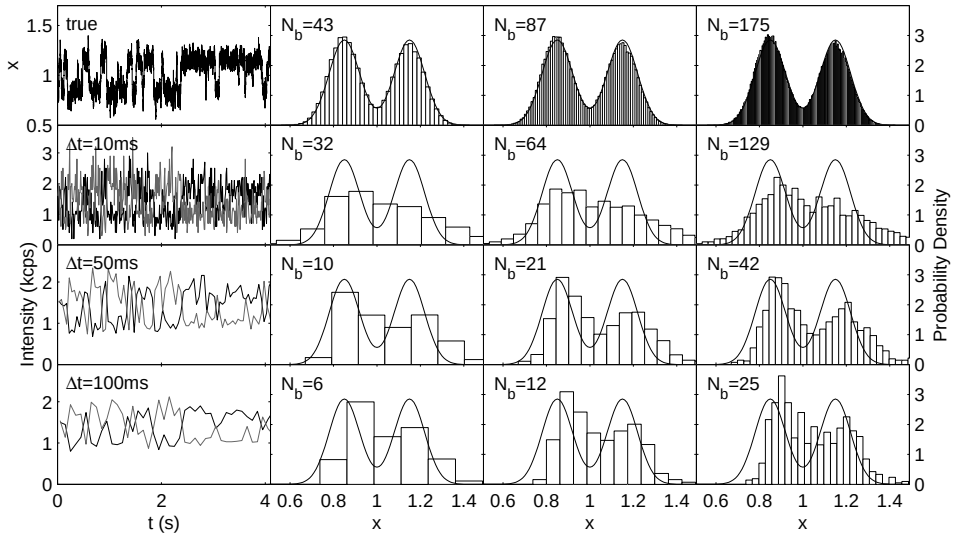


Figure 10.4: Comparison of histograms constructed from a simulated trajectory by constant-time binning with the true probability density (solid lines), illustrating the challenge in determining conformational distributions from single-molecule measurements. Here $x \equiv R/R_0$ is the normalized donor-acceptor distance in a Förster-type resonance energy transfer measurement (cf. Eq. 13.3). The left column shows the true x trajectory, as well as the donor (black) and acceptor (gray) intensity trajectories, binned at 10, 50, and 100 ms. The other three columns compare the underlying probability density (—) with histograms computed from the equal-time binned intensity trajectories, using N_b bins in the x coordinate. The numbers of bins for the set of histograms in the third column are generated according to Scott's formula for the optimal bin width [121]. This formula, like most non-parametric density estimation [122], assumes that there is no error associated with each datum. This assumption is not valid for single-molecule measurements. The columns to the left and right use bin numbers half and twice the optimal value, respectively. Note that different regions of the histogram are broadened differently because of the changes in the variance of the distance estimator as a function of distance. [137]

Chapter 11

Recovering a Probability Density Function from Single Molecule Trajectories

Distributions measured using single-molecule fluorescence methods are commonly visualized by constructing a histogram from a binned time trajectory (averaged over every, *e.g.*, 50 or 100 ms to reduce Poisson counting noise), and have already allowed researchers to uncover many new features in various systems [148, 117, 128, 20, 150, 123], including studies on the dynamics and folding of short peptides – one of the first treatments of the potential of mean force and photon statistics in relation to single molecule measurements – and the discovery of the dynamic equilibrium between closed and open forms of syntaxin I. [92] A quantitative assessment of the underlying probability density function (PDF) is therefore expected to provide further insight for the systems of interest.

When constructed from a fluorescence single-molecule time trajectory, the PDF contains contributions from both the molecular property and from photon-detection statistics. [128, 83, 82] An information-based method such as MIM, in addition to its exact accounting of time resolution and measurement uncertainty, is advantageous for quantitative construction of the molecular distribution. Since the information content in each measurement (be it efficiency or distance) is constant, MIM can also be understood as equal-information binning, in contrast to the commonly used equal-time binning. That the information content is the same for every measurement is an important property that allows one to construct statistically robust distribution functions. This further affords model-free deconvolution to uncover the sought molecular property distribution in an unbiased, objective way. While the ideas contained in the following discussion is general, the development focuses on statistical methods that are applicable to experiments with immobilized single molecules. Such an experimental scheme can in principle provide dynamical information on a time scale covering several decades.

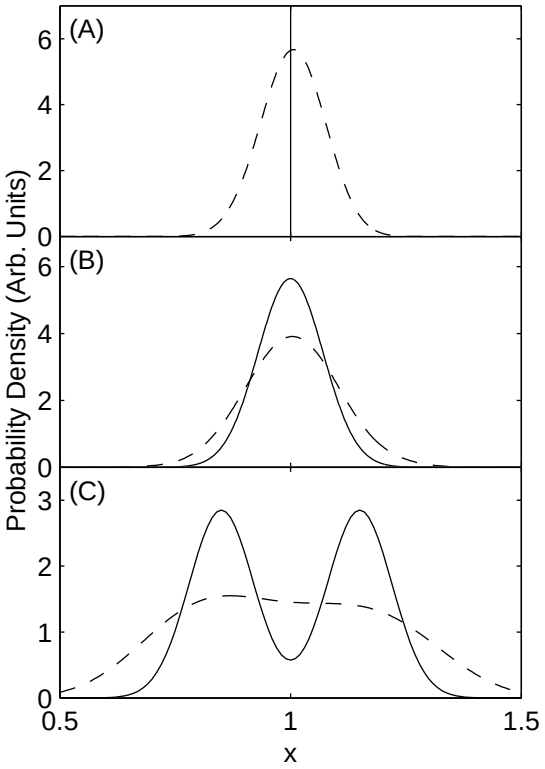


Figure 11.1: Comparison of true (—) and raw (– –) PDFs for trajectories simulated (A) at constant x, (B) on a harmonic potential, and (C) on a bimodal potential.

11.1 Gaussian kernel density estimation

To estimate the distance distribution from a single-molecule time trajectory, one starts by constructing the raw experimental PDF, $\hat{r}(x)$—containing contributions from photon-counting-related measurement uncertainties—from MIM-extracted distances $\hat{x}_i$. The maximum-likelihood estimators $\hat{x}_i$ are asymptotically normal (Gaussian distributed) and are centered around the true but unknown distance, x_i. By virtue of the equal-information binning (cf. Eq. 13.1), each $\hat{x}_i$ has the same variance α^2. [137] This naturally leads to the use of the Gaussian kernel estimator for $r(x)$,

$$\hat{r}(x) = \frac{1}{T} \sum_{i=1}^{N} \Delta t_i \, k\left(x; \alpha^2\right), \tag{11.1}$$

where N is the number of MIM measurements made from the trajectory, and Δt_i and $\hat{x}_i$ are the duration and distance estimate from the i-th measurement.

To illustrate how experimental measurements yield overly broad density functions, raw PDFs calculated from three trajectories simulated under different conditions are compared in Fig. 11.1 with their respective true PDFs (see *Computational validation* for simulation methods). As can be seen, the raw PDF is an entirely inadequate measurement of the underlying PDF. It should be emphasized, however, that due to the statistically uniform nature of MIM measurements, these raw PDFs already represent an improvement over histograms constructed from constant-time binned trajectories. In equal-time measurements, each time bin contributes to the overall histogram with different significance levels, bringing additional bias and skewness to the resulting histogram.

Mathematically, Fig. 11.1 can be understood by considering the raw density as the underlying molecular PDF, $h(x)$, twice convoluted with the Gaussian kernel $k\left(x; \alpha^2\right)$. One convolution is due to the measurement error in $\hat{x}$, normally distributed with a variance α^2 by virtue of maximum likelihood estimation, and one is due to use of the Gaussian kernel estimator in Eq. 11.1, introducing an additional variance α^2. That is,

$$\begin{aligned} \hat{r}\,(x) &= \left[h(x) \otimes k(x; \alpha^2)\right] \otimes k(x; \alpha^2) \\ &= h(x) \otimes k(x; 2\alpha^2), \end{aligned}$$

where $\otimes$ denotes the convolution operation. The task at hand is then to recover the true molecular PDF, $h(x)$, from knowledge of the raw PDF $r(x)$ and the convolution kernel $k(x)$.

11.2 Covariance of the raw density

Due to the limited duration of single-molecule trajectories, the raw PDF will contain errors resulting from the non-zero relaxation time of the distance correlation function, and from lack of suitable sampling of the raw histogram. Errors of the second sort can be assessed by application of Efron's bootstrap method, explained further in Appendix C. The use of the bootstrap, however, requires the data to be independent and identically distributed. In general, a single-molecule time trajectory may exhibit significant time correlation. That is, the discrete distance measurements (coarse-grained in time) made by

the maximum information method are not necessarily independent, though they should be identically distributed.

To calculate the covariance matrix for the raw density, the notation is changed slightly from that of the previous section. Instead of writing the raw density as a weighted sum of Gaussians (cf. Eq. 11.1), it is written as an integral over time. This makes the treatment more general, since it is not constrained to situations with discrete distance measurements. Given knowledge of the estimated trajectory $\hat{x}(t)$, the raw PDF at a particular x obtained from a trajectory $\hat{x}(t)$ of duration T can be written as

$$\hat{r}(x) = \frac{1}{T}\int_0^T k\left(\hat{x}(t) - x, \alpha^2\right) dt.$$

Note that, in the case of discrete measurements, this reduces to Eq. 11.1. This PDF is estimated from a trajectory of limited duration, so it may contain statistical errors due to insufficient sampling of the conformational space, as has been discussed by Zwanzig and Ailawadi. [159] But its ensemble average will be the true raw PDF,

$$\begin{aligned} r(x) &= \left\langle \int_0^T k\left(\hat{x}(t) - x, \alpha^2\right) dt \right\rangle \\ &= \int_0^T \left\langle k\left(\hat{x}(t) - x, \alpha^2\right)\right\rangle dt. \end{aligned}$$

The second equality holds because the ensemble average and the time integral operations commute. Writing the difference between the measured raw PDF and the true raw PDF as $\delta\hat{r}(x) \equiv \hat{r}(x) - r(x)$, the covariance matrix of the raw PDF is

$$\begin{aligned} \sigma_r^2(x_1, x_2) &= \langle \delta\hat{r}(x_1)\, \delta\hat{r}(x_2)\rangle \\ &= \frac{1}{T^2}\int_0^T \int_0^T \left\langle k\left(\hat{x}(t_1) - x_1, \alpha^2\right) k\left(\hat{x}(t_2) - x_2, \alpha^2\right)\right\rangle \\ &\qquad - \left\langle k\left(\hat{x}(t_1) - x_2, \alpha^2\right)\right\rangle \left\langle k\left(\hat{x}(t_2) - x_2, \alpha^2\right)\right\rangle dt_1 dt_2. \end{aligned}$$

The integrand is a time correlation function and, to a very good approximation, will only be a function of $|t_2 - t_1|$. When the conformational space projected on the x-coordinate is appropriately sampled, this correlation should decay on a time scale much shorter than the length of the trajectory, T. This allows simplification to a more convenient form,

$$\begin{aligned} \sigma_r^2(x_1, x_2) = \frac{2}{T}\int_0^T &\left\langle k\left(\hat{x}(0) - x_1, \alpha^2\right) k\left(\hat{x}(t) - x_2, \alpha^2\right)\right\rangle \\ &- \left\langle k\left(\hat{x}(0) - x_1, \alpha^2\right)\right\rangle \left\langle k\left(\hat{x}(t) - x_2, \alpha^2\right)\right\rangle dt. \end{aligned} \quad (11.2)$$

This formula is the integral of a correlation function, and is simple to evaluate. The ensemble average in the integrand may be converted to a time average for calculation of the correlation function. Given multiple trajectories from the same sample, the correlation function should be averaged across trajectories before integration. Once again, errors in this correlation function may be evaluated by the method of Zwanzig and Ailawadi. [159] To find the covariance of a raw PDF that has been averaged over multiple trajectories, T in Eq. 11.2 should be the sum of the durations of all the trajectories.

11.3 Maximum entropy method

To deconvolve the raw PDF, a one-dimensional form of the maximum entropy method can be used. [69, 70, 71] A merit function $\mathcal{M}$ is constructed for a trial molecular PDF $h(x)$,

$$\mathcal{M}[h(x), \lambda] = \chi^2 + \lambda H. \tag{11.3}$$

χ^2 is a measure of the goodness-of-fit between the raw PDF and the convolution of the proposed molecular PDF,

$$\chi^2 = \int_{-\infty}^{\infty} \frac{\left[r(x) - h(x) \otimes k(x; 2\alpha^2)\right]^2}{\sigma_r^2\,(x,x)} dx, \tag{11.4}$$

H is the negative entropy of the proposed molecular histogram,

$$H = \int_{-\infty}^{\infty} h(x) \ln h(x) dx, \tag{11.5}$$

and λ is a Lagrange multiplier, adjusted so that the final χ^2 after optimization is within $1 \pm 1/\sqrt{N}$. An initial guess of $h = \hat{r}\,(x)$ is used and a provisional $h(x)$ is found when $\mathcal{M}[h(x), \lambda]$ is minimized. The minimization is performed numerically using a steepest descent algorithm.

The provisional $h\,(x)$ are used to find the correct λ, and thus the experimentally justified $h\,(x)$. This is the core concept behind the maximum entropy method. Given a set of underlying PDFs, all of which adequately represent the data, the one with the highest entropy is the only one that is justified by the data. While one of the higher-entropy PDFs may be more correct, the data is insufficient to show this. This correct $h\,(x)$ can be found by varying the Lagrange multiplier λ used in the minimization until χ^2 is within the range $1 \pm 1/\sqrt{N}$.

11.4 Covariance of the deconvolved density

Since $r\,(x) = h\,(x) \otimes k\,(x)$, $\delta r\,(x) = \delta h\,(x) \otimes k\,(x)$, and the covariance matrix σ_h^2 of the molecular PDF can be calculated by the two-dimensional deconvolution of the covariance matrix of the raw PDF,

$$\begin{aligned} \langle \delta r\,(x_1)\, \delta r\,(x_2) \rangle &= \langle \delta h\,(x_1)\, \delta h\,(x_2) \rangle \otimes k\,(x_1) \otimes k\,(x_2)\,, \\ \sigma_r\,(x_1, x_2) &= \sigma_h\,(x_1, x_2) \otimes k\,(x_1) \otimes k\,(x_2)\,. \end{aligned}$$

Knowledge of this covariance matrix is important for several reasons. The diagonal term gives the variance of the deconvolved PDF as a function of x, a measure of the overall and point-by-point accuracy. Just as important are the off-diagonal terms, which provide information about the relative accuracy between different regions of the density. For instance, in a trajectory from a bimodal PDF, the time-scale of equilibration between the two high-density regions will be much longer than that within the two regions. This means that the deconvolved densities at values of x within the same potential well should be accurate

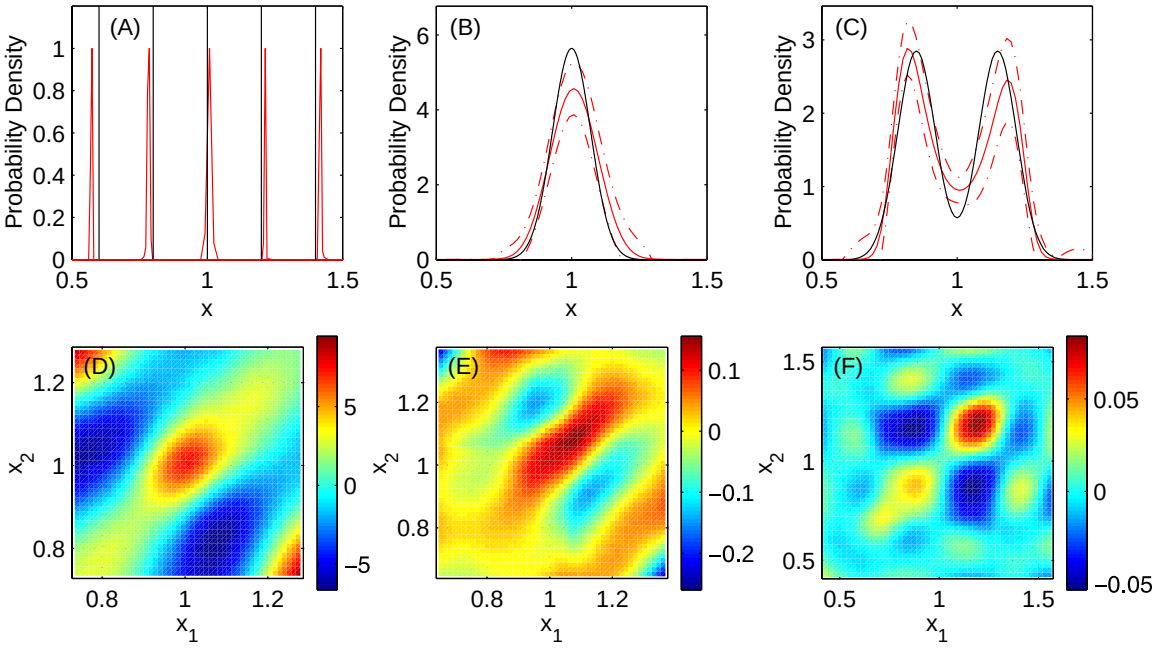


Figure 11.2: Results from deconvolution of test trajectories: (A) True (black) and deconvolved (red) probability densities for constant trajectories at $x = 0.6$, 0.8, 1.0, 1.2, and 1.4. (B) True (black) and deconvolved (red) probability densities for a trajectory simulated on a harmonic potential centered at $x = 1.0$. The 95% confidence interval for the deconvolved density is indicated by dashed lines. (C) True (black) and deconvolved (red) probability densities for a trajectory simulated on a bimodal potential centered at $x = 1.0$. The 95% confidence interval for the deconvolved density is indicated by broken lines. (D) Covariance matrix for the deconvolved density shown in A at $x = 1.0$. (E) Covariance matrix of the deconvolved density shown in B. (F) Covariance matrix of the deconvolved density shown in C.

with respect to one another. That is, they should be positively correlated. This will be reflected by a positive covariance. The densities at values of x in different potential wells, on the other hand, should be negatively correlated, since more time spent in one potential well means less time spent in the other. This will produce a negative covariance. Thus, the covariance between two points in the PDF is primarily determined by the time-scale of equilibration between those two points. Therefore, the deconvolved covariance matrix provides further insights into the dynamics afforded by the experimentally measured PDF.

11.5 Computational validation

To test this procedure, some basic simulations were performed. Three classes of trajectories were produced: constant x, a harmonic potential, and a bimodal potential. The trajectories were produced by simulation of one-dimensional high-friction Langevin

dynamics and subsequent conversion to photon-arrival-time data, as previously described. [137] The trajectories on harmonic potentials were simulated at a temperature of $\beta = 100$, with friction coefficient $\gamma = 1.0$. The bimodal trajectories were simulated at the same temperature, with a friction coefficient of $\gamma = 0.1$. The combined number of photons emitted before bleaching of the dyes was 2×10^5, and the signal-to-background ratio on both channels was 5.0.

As can be seen in Figure 11.2, the deconvolution procedure performs well. The constant-distance trajectories in Fig. 11.2(A) all deconvolve to delta functions. Bias in the location of the peak is small (less than 0.023). This is expected based on the bias analysis of the maximum information method. [137] Trajectories from harmonic potentials produce Gaussian PDFs in Fig. 11.2(B) that match the true PDF from the underlying trajectory. The deconvolved height and standard deviation of the Gaussian profile are 4.57 ± 0.70 and 0.22 ± 0.06, respectively, agreeing with the true values of 5.64 and 0.17. The covariance matrices for these deconvolved PDFs, shown in Fig. 11.2(D) and (E), are similar. Densities that are close to each other are positively correlated, while points farther away are negatively correlated. If the trajectory spends more time in one part of the density function, it spends less time in the other.

The deconvolution of the trajectory from a bimodal potential is more revealing. The raw PDF is too broad and it obscures the true separation of the two potential wells. From the raw PDF (cf. Fig. 11.1), it is not possible to measure the width of the individual peaks or the depth of the barrier that separates them. The deconvolved PDF shown in Fig. 11.2(C), on the other hand, matches well with the underlying PDF. The heights of the peaks and the depth of the well between them are 2.88 ± 0.37, 2.45 ± 0.57, and 1.71 ± 0.51, respectively; all compare well with the true values of 2.85, 2.85, and 2.28. The deconvolved covariance matrix in Fig. 11.2(F) is similarly informative. Two distinct regions of positive covariance can be identified, corresponding to the two wells in the potential. The covariance between points in different wells is negative, indicating the slower time-scale of equilibration between the wells. None of these properties would be apparent from the raw histogram alone.

The differences between the true density and the recovered density can be attributed to the finite length of the trajectories. This root cause manifests itself in two ways. The accuracy of the density at a particular x value depends on how often the trajectory visits that value. Additionally, the accuracy of the covariance estimate depends on the accuracy of the correlation functions, which are themselves strongly dependent on the length of the trajectory.

Chapter 12

Photon-by-photon Calculation of Intensity Correlation

The time correlation function of a photon-by-photon single-molecule intensity trajectory can be directly calculated by representing the time-dependent intensity as a series of Dirac delta functions [151]

$$I_n(t) = \sum_{i=1}^{N} \delta\left(t - \tau_i^{(n)}\right),$$

where $\{\tau_i^{(n)}\}$ is the set of photon arrival times on channel n. The true correlation function is

$$C_{nm}(t) = \langle I_m(t) I_n(0) \rangle - \langle I_m \rangle \langle I_n \rangle,$$

where $\langle \cdots \rangle$ indicates an ensemble average.

If the single-molecule trajectory is long enough, the ensemble average may be converted to a time average,

$$\begin{aligned} C_{nm}(t) &= \frac{1}{(T-t)} \int_0^{T-t} I_n(\tau) \cdot I_m(t+\tau)\, d\tau - \bar{I}_n \bar{I}_m \\ &= \frac{1}{(T-t)} \sum_{i=1}^{N_n} \sum_{j=1}^{N_m} \delta\left(t + \tau_i^{(n)} - \tau_j^{(m)}\right) - \bar{I}_n \bar{I}_m, \end{aligned}$$

where T is the duration of the entire trajectory (cf. region I in Fig. 13.2). The correlation function is just a sum of scaled delta functions which can be computed directly from the photon arrival sequence. The correlation function $C_{mn}(t)$ may be further averaged over a time interval $[t_a, t_b]$ to reduce the stochastic noise,

$$\begin{aligned} \bar{C}_{nm}(t_a, t_b) &= \frac{1}{t_b - t_a} \int_{t_a}^{t_b} C_{nm}(t)\, dt \\ &= \frac{1}{(t_b - t_a)} \sum_{i=1}^{N_n} \sum_{j=1}^{N_m} \frac{\mathbf{1}_{t_a \le \tau_j^{(n)} - \tau_i^{(n)} \le t_b}}{T - \left(\tau_j^{(m)} - \tau_i^{(n)}\right)} - \bar{I}_n \bar{I}_m. \end{aligned}$$

$\mathbf{1}_{expr}$ is the indicator function, equal to 1 when *expr* is true and 0 otherwise.

Since the correlation function is calculated as a time average over a single trajectory, errors may arise due to incomplete sampling of the conformational space. These errors are estimated using the method of Zwanzig and Ailawadi [159]. In the averaging of correlation functions from multiple trajectories, errors may be propagated in the usual manner. This allows calculation of intensity auto-correlation and cross-correlation in FRET trajectories on a photon-by-photon basis and is analogous to most implementations of fluorescence correlation spectroscopy (FCS) [91]. As such, it is straightforward to use this microscope in FCS-type applications.

Part IV

Application of the Statistical Methods to a Model System: Polyproline

As an experimental demonstration of the statistical techniques developed in Parts I–III, we have measured distances and distance distributions in a series of poly(L-proline). Previous studies have shown an excellent correspondence between bulk and single-molecule FRET measurements of distance [118, 84, 111]. Therefore, this part focuses on the discussion of the issues mentioned above. The constituent amino acids of poly(L-proline) are expected to exist primarily in the *trans*- state and should be fairly static on the time scale of single-molecule measurements, with no complex dynamics [56]. They are thus a good test case for our new approach. Indeed, we recovered narrow distance distributions for donor-acceptor distances in the poly-prolines. The means of these distributions are well explained by application of the worm-like chain model with a persistence length of 23 Å[93]. While the short persistence length may have further implications for the structure of proline-rich signaling proteins [74]—which are beyond the scope of this article—it is hoped that the general methodologies described herein will aid in the development of quantitative and predictive understanding of the dynamic behavior of biological macromolecules.

Chapter 13

Methods and Materials

13.1 A server-based single-molecule microscope

Microscope Construction. The design of the microscope is diagrammed in Fig. 13.1. The 532-nm excitation light from a continuous-wave laser—a diode-pumped solid-state laser, or DPSS (Coherent, Compass 315M-100)—is passed through a cleanup filter (Chroma, HQ545/10x) and expanded to a diameter of ∼8 mm to match the back aperture of the microscope objective. To minimize sample exposure to light, a shutter is installed in the beam path and controlled by the control server (see System Software). A polarization element is placed immediately before the entrance of the microscope; it may be a $\lambda/2$ plate, or a $\lambda/4$ plate, a Pockel cell, or any combination thereof. In the experiments reported here, a $\lambda/4$ plate was used to ensure circularly polarized excitation. After entering the microscope, the excitation beam is reflected from a dichroic mirror (Chroma, Z532rdc) into an 60×, infinity-corrected, N.A. 1.4, oil-immersion objective (Olympus, PlanApo). The objective focuses the light to a diffraction-limited spot on the surface of the sample cover slip, which is secured onto a custom-made, temperature-regulated vacuum chuck mounted on top of a nm-resolution piezoelectric stage (Physik Instrumente, P734). The piezoelectric stage is driven by a high-voltage driver (Physik Instrumente, E509.C2 and E503.00) interfaced to the control server computer (see below).

Emitted light from individual molecules immobilized on the surface of the slide is collected by the same objective and passes through the dichroic mirror. It then passes through a tube lens and is separated by another dichroic mirror (Chroma, Q645LP) into donor and acceptor channels. The photons on their respective beam paths are spectrally filtered by bandpass optics (Chroma, HQ600/80m for donor emission and HQ705/130m for acceptor emission) before being focused on a pair of single-photon-counting avalanche photodiodes (APDs, Perkin-Elmer, SPCM-AQR13). Each APD is connected to a photon registration server and outputs a TTL pulse upon detecting a photon. The modular design of this microscope is such that if more parameters such as polarization were to be measured, one simply drops in additional APD/server modules and the software (see below) will take care of coordination.

The filters and dichroic mirrors used were chosen to match the absorption and fluorescence properties of the fluorophores used in these experiments. Since crosstalk can

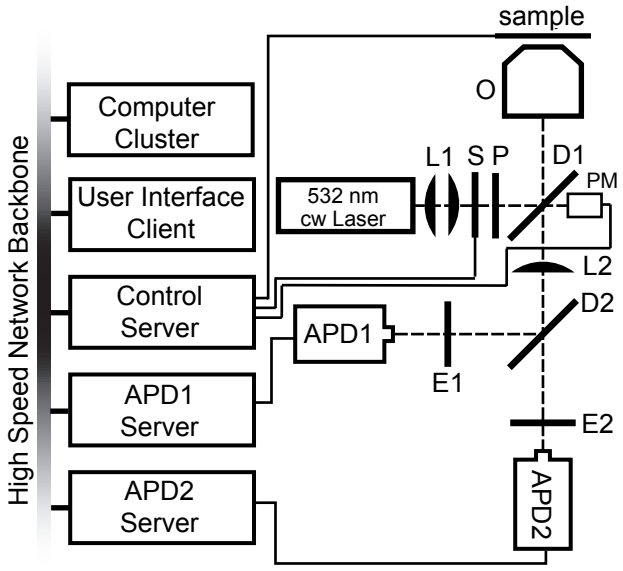


Figure 13.1: Configuration of a cw-excitation photon-by-photon microscope. Illumination is provided by a continuous-wave diode pumped solid-state laser. Telescope lens assembly L1 expands the beam to 8 mm in order to fill the back aperture of the objective O. A shutter S is installed in the beam path to minimize unnecessary illumination. A polarization element P is placed immediately before the entrance of the home-built microscope body. Dichroic mirror D1 reflects the excitation light into the back aperture of an infinity-corrected objective O, which focuses it onto the sample mounted on a piezo stage. Excitation light that leaks through D1 is detected by a photodiode PM to monitor laser excitation power. Fluorescence from the sample is collected and collimated by objective O and passes through dichroic D1 before being focused by tube lens L2 and split into donor and acceptor channels by dichroic mirror D2. The emitted photons finally pass through emission filters E1 and E2 before being recorded by avalanche photodiodes APD1 and APD2. The stage and the APDs are each controlled by a separate computer server, all of which are connected via a high-speed network backbone via the TCP/IP protocol to a client computer which runs the user interface. The modular design of the system is such that if more parameters such as polarization are to be measured, one simply drops in additional APD/server modules.

be fully treated – accounting for the changing intensities on each channel – by including the crosstalk coefficients in the signal-to-background ratio, as previously shown [137], much broader bandpass filters can be used, allowing for more effective collection of photons.

System Software. To coordinate the complicated tasks of real-time data acquisition and analysis, the system software is split into its core functionalities, which are composed of instrument control, photon registration, and user interface. The piezoelectric stage and each APD are controlled by server programs running on separate computers. All servers are connected by a high-speed TCP/IP network to the client computer running the user interface program. The current implementation utilizes a 1 gigabit per second intranet backbone. No data stream latency was observed.

The instrument control server interfaces with the microscope via a multi-function I/O card (National Instruments, PCI-6052E) to perform a variety of control and measurement functions. These functions include setting and measurement of the position of the piezoelectric stage, control of the shutter, measurement of the laser power, and operation of any other physical components necessary to a particular experiment. The photon registration servers measure and record absolute arrival times of photons at the APDs via a timing / counting interface (National Instruments, PCI-6602). The arrival times are measured against an internal 80-MHz clock on the board, providing 12.5-ns resolution for photon arrival times which, with the ∼100-ns dead time on the APDs, is more than sufficient for the CW excitation being used. Note that the 12.5-ns time resolution refers only to the timing on individual photons, not to the time required to make a distance measurement. Theoretical limits on time resolution in the measured distance trajectories have been presented before [137] and the experimental realization of those results is expanded upon below. To ensure that no detected photons are missed, the measured TTL stream is polled via a direct memory access (DMA) channel and buffered before sending out to the client. While it is possible to run them all on the same computer, this generally results in dropped photons at high count rates. Performance is significantly improved when each APD is monitored by a different computer. For high count rate applications, the TTL pulses from the APD are monitored simultaneously by two counters on the same card. The data from each of these is then compared and errors are corrected before the data are sent to the user interface. Tests showed that such a dual counter configuration should be used when the average count rate is higher than ∼20 kilo-counts per second (kcps). This configuration was tested using a pulse generator (Stanford Research, DG535) for constant-rate detection and a light-illuminated APD for exponentially distributed inter-photon duration at average 10–1000 kcps. No missing photons were observed up to an average count rate of 100 kcps for ∼20 s. At greater count rates, impractical for single-molecule fluorescence experiments, the setup misses 1 photon per 3 s at 500 kcps and 7 photons per 0.3 s at 1000 kcps. The standard deviation of chronological time registration was found to be ∼3–40 ns. Both autocorrelation and cross correlation analyses were performed on these test trajectories. No correlations were found in the entire photon detection and data registration process.

The client user interface, which may be run on yet another computer, controls the operations of all of the data and control servers, as well as providing real-time data analysis capabilities by networking with a computer cluster that offers parallel computation. The entire software suite was coded in C++ and the client runs on either the Windows or the

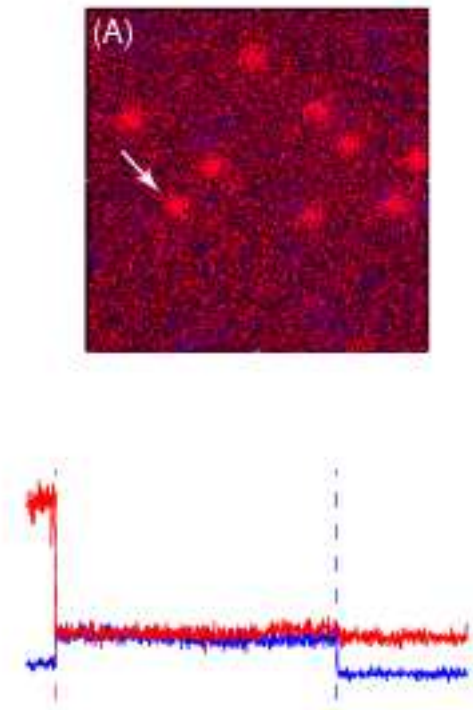


Figure 13.2: (A) Raster-scanned image of doubly labeled $P_{12}CG_3K$(biotin), where the donor and acceptor emission are represented in blue and red false color, respectively. The image was binned at 40 nm and filtered with a 5-pixel by 5-pixel Gaussian averager. (B) A representative intensity trajectory acquired from the spot indicated by the arrow in (A). The segmentation of different regions for data analysis is shown by the vertical dashed lines.

GNU/Linux operating systems.

Imaging. To locate molecules for time-resolved observation and on-line analysis, a fluorescence image of the surface that contains immobilized single-molecules must be obtained. The stage is raster scanned across a predefined area, generally 10 μm$\times$10 μm, and photon arrival times are recorded on all channels, along with the position of the stage as a function of time. The reference times for position measurements and photon arrival times on each channel are all synchronized by a trigger pulse generated by the shutter upon opening. With this data, each photon can be assigned a specific origin on the sample cover slip. For viewing, the photon origins are spatially binned into pixels and displayed on screen. These images are generally acquired at the lowest possible excitation power to guard against irreversible photochemical reactions, or photo bleaching. The data in this part were collected at an excitation power of 780 nW (or 350 W/cm^2 assuming a diffraction-limited focal disk). Inspection of the two channel image allows selection of doubly labeled molecules suitable for recording single-molecule trajectories. Molecules labeled only with the acceptor will not

be visible, whereas molecules labeled only with the donor or doubly labeled molecules with a bleached acceptor will be visible only on the donor channel. Correctly labeled molecules will be visible on both channels.

Time Trajectories. Once an appropriately labeled molecule is found, the stage is moved (with shutter closed) so that the selected molecule is at the focus of the objective. The counters are then armed and the shutter is opened. As in the imaging algorithm, the opened shutter generates a pulse that simultaneously triggers the counters, synchronizing their zero-times. A typical intensity trajectory is shown in Fig. 13.2, where the characteristic bleaching pattern of a single molecule FRET trajectory can be observed. The acceptor usually bleaches first, causing the intensity on the acceptor channel to drop to the level of the background plus crosstalk from the donor channel, while the intensity on the donor channel increases to its value in the absence of the acceptor. When the donor fluorophore bleaches, the intensities on both channels drop to their respective background levels. The high background level on the acceptor channel is caused by plastic coverwells (Molecular Probes, C18139) that were used to prevent sample evaporation. Subsequent experiments have found that the use of plastic spacers (Molecular Probes, P18178) combined with quartz covers reduce the background level to less than 200 cps. The maximum intensity on the donor channel (Region II) is lower than that on the acceptor channel because, while our filter sets were optimized to include the tail of the acceptor emission, the tail region of the donor's emission spectrum overlaps considerably with the acceptor's emission spectrum. Photons are thus much more efficiently collected from the acceptor than they are from the donor. Data acquired with this microscope may be analyzed using any of the various powerful statistical methods available [138, 137, 50, 103, 42, 151, 5, 116, 3, 81, 142, 143, 41, 55, 54].

13.2 Sample preparation and characterization

A series of peptides with the sequence P_nCG_3K(biotin) (n = 8, 12, 15, 18, 24) was synthesized using the Fmoc solid-phase synthesis technique. The C-terminal lysine was prefunctionalized with biotin on the amine (Nova Biochem, 04-12-1237). The poly(L-proline) peptides are expected to be predominantly in the *trans*- isomer (forming a poly-proline-II helix) under experimental conditions [64, 124]. The peptides were labeled with Alexa Fluor 647 C_2-maleimide (Invitrogen/Molecular Probes, 20347) on the cysteine and Alexa Fluor 555 carboxylic acid C_5-succinimidyl ester (Invitrogen/Molecular Probes, 20009) on the N-terminal proline (cf. Fig. 13.3). The free dye was removed by the addition of 0.2 mg/mL streptavidin (Invitrogen/Molecular Probes, S888) and subsequent centrifugal filtration. Streptavidin-bound proline-dye conjugates were retained by the filter, while free dyes passed through. Unfortunately, the chemical structures of these dyes are proprietary and not available. These two fluorophores form a FRET pair with an R_0 of 51 Å, calculated from the absorption and fluorescence spectra. The ensemble-averaged steady-state anisotropies (measured on a SPEX Fluorolog) are given in Table 13.1. Large anisotropies are an indication that the orientation of the excited optical dipole does not have time to randomize before it relaxes back to the ground state. The fluorescence lifetime of Alexa 555 is 0.27 ns [90], much shorter than the timescale of rotation of the proline-streptavidin complex ($\sim$ 10 ns). Thus, in this case, even though the acceptor's anisotropy is low, κ^2 is

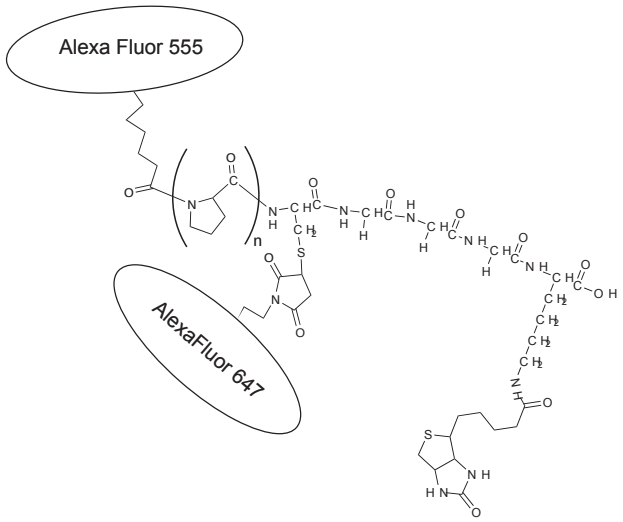


Figure 13.3: Chemical structure of doubly labeled P_nCG_3K(biotin), where n = 8, 12, 15, 18, 24.

n	donor	acceptor
8	0.205 (9)	-0.000 (7)
12	0.244 (8)	-0.003 (6)
15	0.235 (12)	0.016 (11)
18	0.223 (11)	0.023 (14)
24	0.193 (9)	0.040 (20)

Table 13.1: Steady-state fluorescence anisotropy of doubly labeled poly-prolines, excited at 532nm.

not in the dynamically averaged regime for single photon emissions. However, the donor-acceptor distance measurement is made on a photon by photon basis over a much longer timescale, ~ 1 ms — far longer than the timescales for rotation of the dyes. Averaged over this ~ 1 ms timescale (which typically contains 15-25 photons), the value of κ^2 approaches the 2/3 limit. This is demonstrated in detail in a later chapter (cf. Eq. 15.1, Fig. 15.1).

The labeled peptides were immobilized via biotin-streptavidin chemistry [61]. This immobilization scheme has been shown to exhibit minimal interaction with tethered molecules [104]. Briefly, quartz cover slips (Technical Glass Products, $1\times1\times0.17$ mm^3) were first cleaned by sequential sonication in 1M KOH, absolute ethanol, 1M KOH, and ethanol. They were then dried and silanized with (3-aminopropyl)-trimethoxysilane (APS) by soaking for two minutes in a 2% solution of APS in acetone followed by 30 minutes at 110°C. The silanized cover slips were functionalized with PEG-SPA and PEG-biotin by incubation for 3 hours in a water solution of 10% PEG-SPA, 0.1% PEG-biotin, and 0.01M $NaHCO_3$ at pH 8.2. Finally, the streptavidin-bound fluorescently labeled peptide was incubated for five

minutes on the active side of the cover slip at a concentration of ∼10 pM. The sample cover slip was then secured on the microscope for observation. No deoxygenation agents were used in the present study. More than 60 valid single-molecule trajectories were acquired for each n.

13.3 Calibration of the maximum information method for the experimental system

The maximum information method (MIM) allows one to quantitatively follow single-molecule FRET efficiency and distance dynamics with the highest time resolution allowed by the information content in an experimental data set [137]. It has, for example, allowed identification of two coexisting conformations of the cdAE1 protein [108]. The time resolution (Δt) for each maximum-information measurement is determined by the expected measurement error $\alpha \equiv \delta x/x$. In a two-channel FRET setting, it is given by

$$\Delta t = \frac{1}{\alpha^2}\left(\frac{I_d^\beta}{\zeta_d\left(x\right)}\left[\frac{\partial \zeta_d}{\partial x}\right]^2 + \frac{I_a^\beta}{\zeta_a(x)}\left[\frac{\partial \zeta_a}{\partial x}\right]^2\right)^{-1}. \tag{13.1}$$

As before, $x = R/R_0$ is the normalized distance, R is the donor-acceptor distance, and R_0 is the Förster radius—a function of the orientation factor κ^2 that accounts for the relative orientation of the donor and acceptor transition dipoles. I_d^β and I_a^β are the maximum intensities on each channel (I_d^β, donor intensity in the absence of the acceptor, and I_a^β, acceptor intensity at $R \ll R_0$), $\zeta\left(x\right)$ is the distance dependent intensity scaling function with $\zeta_d(x) = x^6/(1+x^6)$ and $\zeta_a(x) = 1/(1+x^6)$. Within the time period Δt allowed by α, the donor-acceptor energy transfer efficiency and corresponding distance are computed as

$$\hat{E} = \frac{I_d^\beta n_a - I_a^\beta n_d \beta_a^{-1}}{I_d^\beta n_a\left(1-\beta_d^{-1}\right) + I_a^\beta n_d\left(1-\beta_a^{-1}\right)} \tag{13.2}$$

and

$$\hat{x} = \left(\frac{\beta_a}{\beta_d}\cdot\frac{I_d^\beta n_a - I_a^\beta n_d \beta_d}{I_a^\beta n_d - I_d^\beta n_a \beta_a}\right)^{1/6}. \tag{13.3}$$

In the above equations, n_d and n_a are the number of photons detected within the chosen time interval on the donor and the acceptor channels respectively and β_d and β_a are the signal to background ratios. This method is robust against intermittency and transient variations in the dyes' quantum efficiency [137], The orientation factor $\kappa^2 = 2/3$ was used for calculating the donor-acceptor distances [34]. Ensemble-averaged steady-state anisotropy measurements of doubly labeled poly(L-proline) conjugated to streptavidin indicate that the fluorescent probes have already experienced depolarization within the ∼35 ns protein rotation time. The linkers used for tethering fluorescent probes to the protein surface are expected to exhibit segmental dynamics on the ultrafast to ns time scales [21, 65]. Together with the absence of correlation in the single-molecule time trajectories, these considerations lead to the conclusion that 2/3 is a good approximation for κ^2. This assumption was made

in several recent studies, showing that accurate distance information can be obtained from immobilized DNA molecules [111] as well as from diffusing single molecules such as DNA [84] and poly-proline [118, 55] using FRET. These works all point to the importance of carefully considering contributions from background, cross talk, and other instrumentation factors. Note that in addition to these parameters—time-independent cross talk has been shown to be a form of background [137]—the correction factor for detector efficiency for the donor and acceptor channels, $(\Phi_a\eta_a)/(\Phi_d\eta_d)$, has been explicitly included in the derivation of Eqs. 13.2 and 13.3. This allows further correction of potential molecule-to-molecule variations in the absorption cross section or emission spectrum as a result of heterogeneity in the microscopic environment.

Calibration. Before analysis of the photon arrival time data using the MIM algorithm, several calibration values must be determined. These include the background levels on each channel, the maximum intensities on each channel (the background levels and maximum intensities may vary from molecule to molecule), and the crosstalk coefficients (χ_d and χ_a, constant for a given experimental setup). χ_d is the fraction of donor fluorescence that will be observed on the acceptor channel. Likewise, χ_a is the fraction of acceptor fluorescence that will be observed on the donor channel. They may thus be calculated directly from the fluorescence spectra $F_d(\nu)$ and $F_a(\nu)$ of the donor and the acceptor, the transmission curves $T_d(\nu)$ and $T_d(\nu)$ of the emission filters that define the donor and acceptor channels, and the response curve $R_A(\nu)$ of the APD itself by

$$\chi_d = \frac{\int_0^\infty F_d(\nu)\, T_a(\nu)\, R_A(\nu)\, d\nu}{\int_0^\infty F_d(\nu)\, T_d(\nu)\, R_A(\nu)\, d\nu}$$
$$\chi_a = \frac{\int_0^\infty F_a(\nu)\, T_d(\nu)\, R_A(\nu)\, d\nu}{\int_0^\infty F_a(\nu)\, T_a(\nu)\, R_A(\nu)\, d\nu}.$$

To calculate the other required parameters, each single-molecule time trajectory is divided into three regions, I (FRET), II (donor-only), and III (background) (cf. Fig. 13.2). The intensity changes between regions I and II, due to acceptor photo-bleaching, and between II and III, due to donor photo-bleaching, are abrupt. Quantitative segmentation of the time trajectory was accomplished by the intensity change point detection algorithm detailed above. Depending on the relative intensities, this method allows one to determine the intensity change point to within a few photons. Regions I, II, and III, as determined by the change point algorithm, were then used to determine the calibration parameters. In region I, the period from the beginning of the trajectory to the time the acceptor photo-bleaches, the observed intensities on the donor and acceptor channel can be written as

$$\bar{I}_d^{(\mathrm{I})} = I_d^0\zeta_d(\bar{x}) + \chi_a I_a^0\zeta_a(\bar{x}) + B_d,$$
$$\bar{I}_a^{(\mathrm{I})} = I_a^0\zeta_a(\bar{x}) + \chi_d I_d^0\zeta_d(\bar{x}) + B_a,$$

where $\bar{x}$ is the (unknown) average distance in region I and $\bar{I}_d$ and $\bar{I}_a$ are the average intensities on the donor and acceptor channels. These are computed by applying the maximum-likelihood estimator $\bar{I}^{(\mathrm{I})} = N_\mathrm{I}/T_\mathrm{I}$, where N_I is the total number of photons in the region and T_I is the total time duration of the region.

In region II, the period from the time that the acceptor photo-bleaches to the time that the donor photo-bleaches, the effective donor-acceptor distance is $x \to \infty$, and the observed intensities will be

$$\begin{aligned} \bar{I}_d^{(\mathrm{II})} &= I_d^0 + B_d, \\ \bar{I}_a^{(\mathrm{II})} &= \chi_d I_d^0 + B_a. \end{aligned}$$

In region III, the period from the time the donor photo-bleaches and until the end of the trajectory scan, the intensities are just the background counts:

$$\begin{aligned} \bar{I}_d^{(\mathrm{III})} &= B_d, \\ \bar{I}_a^{(\mathrm{III})} &= B_a. \end{aligned}$$

Solving these equations, one obtains expressions for the desired calibration values.

$$\begin{aligned} I_d^0 &= \bar{I}_d^{(\mathrm{II})} - \bar{I}_d^{(\mathrm{III})}, \\ I_a^0 &= \frac{I_d^0 + \chi_d I_d^0 P}{P + \chi_a}, \end{aligned}$$

with $P = (\bar{I}_d^{(\mathrm{II})} - \bar{I}_d^{(\mathrm{I})})/(\bar{I}_a^{(\mathrm{I})} - \bar{I}_a^{(\mathrm{II})})$. Once the calibration parameters have been determined for an individual molecule, the photon arrival times from region I can be subjected to the MIM algorithm, generating the desired energy-transfer efficiency or distance trajectory. The uncertainties in the determination of these parameters are propagated to assist in the assessment of variations in molecule-to-molecule measurements.

Chapter 14

Static Poly(l-Proline) End-to-End Distance on Single-Molecule Experiment Time Scales

Representative intensity trajectories for all peptide lengths are shown in Fig. 14.1 (first column), along with the reconstructed efficiency (second column) and distance trajectories (third column) from MIM analysis with a relative error of $\alpha = 0.1$ (equivalent to a distance uncertainty of $\Delta R = 5.1$ Å). The bleaching times for both the donor (∼160 s) and the acceptor (∼10 s) appear roughly exponentially distributed (Histograms are presented in the Supplementary Material) and are summarized in Table 14.1. The average time resolution (Δt) is ∼26–70 ms.

The fluorescence intensities appear constant over time, indicative of constant energy transfer efficiency on the time scales accessible to fluorescence single-molecule experiments. Similarly, the MIM-determined FRET efficiencies and distances appear to fluctuate randomly about their respective mean values. To examine if the extrinsic probes may transiently interact with the peptide in a non-specific way [72, 52, 36, 22], correlation analyses on the fluorescence intensity, energy transfer efficiency, and distance were carried out on all the trajectories. The lack of correlation on the experimental time scales indicate that single poly(L-proline) molecules interrogated in this study can be considered as exhibiting

n	$\bar{\tau}_a(s)$	$\sigma_{\tau_a}(s)$	$\bar{\tau}_d(s)$	$\sigma_{\tau_d}(s)$	$\overline{\Delta t}$(ms)
8	6.0	6.9	118	154	70
12	7.8	8.9	161	176	70
15	11	14	192	168	26
18	11	16	135	135	30
24	16	19	194	192	46

Table 14.1: Mean ($\bar{\tau}$) and standard deviation (σ_τ of donor (d) and acceptor (a) bleaching times for P_nCG_3K(biotin).

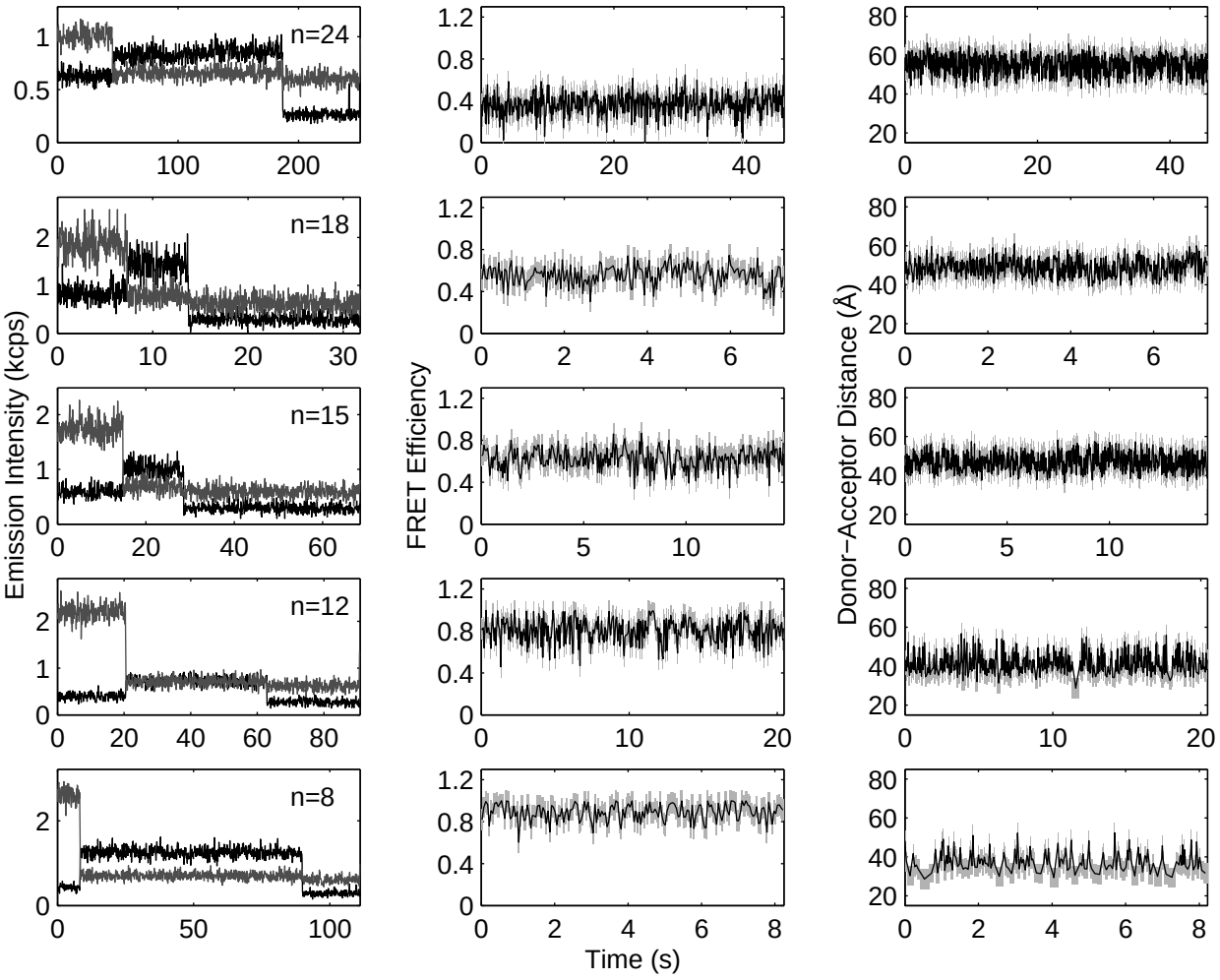


Figure 14.1: Intensity (left), FRET efficiency (center), and calculated donor-acceptor distance (right) as a function of time. Intensity trajectories are binned at 100 ms on both the donor (black) and acceptor (gray) channels. Distance and efficiency trajectories trajectories are calculated using the maximum information method and assume $R_0 = 51$ Å. Dimensions of the gray boxes on the distance and efficiency trajectories indicate the time resolution (horizontal dimension) and expected 95% confidence interval in energy transfer efficiency or donor-acceptor distance (vertical dimension).

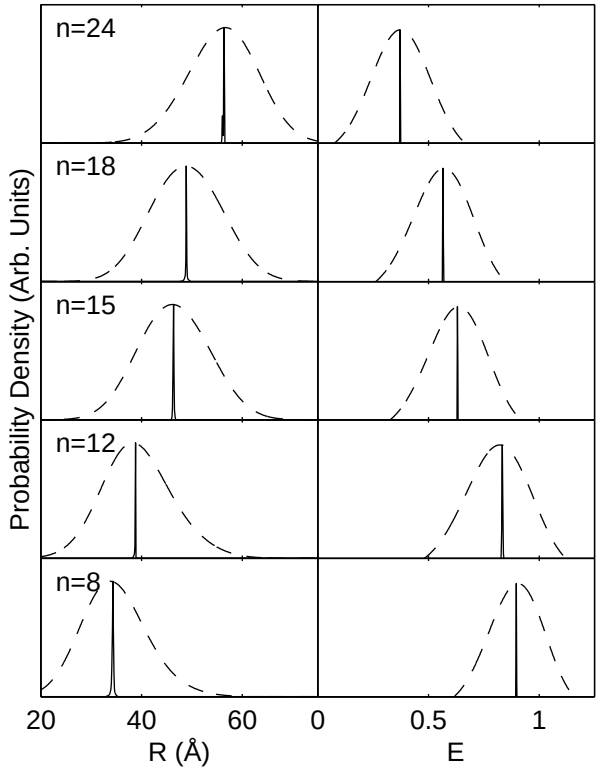


Figure 14.2: Raw (- -) and deconvolved (—) distribution functions from single P_nCG_3K(biotin) trajectories shown in Fig. 14.1.

static, mean end-to-end distances on time scales from ms to tens of seconds (see Supplementary material for correlation results), in contrast to studies where the measurement time scale is comparable to that of molecular motions [72, 52]. Higher time resolution data, or more extended poly(L-proline) molecules may allow direct observation of conformational dynamics.

For molecules such as poly(L-proline) that presumably exhibit constant energy transfer efficiencies, one expects a sharp distribution peaking at the mean value. The raw distribution functions $\hat{r}_n(E)$ $(n = 8, 12, 15, 18, 24)$ constructed using the Gaussian kernel estimator (cf. Eq. 11.1), however, appear very broad (cf. Fig. 14.2). This is not surprising, as they are broadened both by photon-counting noise and by the density estimation procedure. To recover the underlying efficiency distribution of individual poly(L-proline) molecules, the Maximum Entropy deconvolution procedure was applied to the raw distribution functions $\hat{r}_n(E)$. As shown in Fig. 14.2, the Maximum Entropy deconvolution drastically reduces the distribution to sharply-peaked $\hat{h}_n(E)$, as one would have expected from poly(L-proline) molecules with time-invariant energy transfer efficiency.

Chapter 15

Towards Quantitative FRET Measurement

15.1 Sufficient sampling of relative donor-acceptor orientation using MIM

The use of $\kappa^2 = 2/3$ in R_0 implies that orientational correlations between the donor and acceptor dyes disappear on a time-scale shorter than the inter-photon timing, and that the number of photons used in distance calculations is sufficient to ensure that the distribution of κ^2 is close to normal. For the former, one examines the intensity auto- and cross-correlation functions. They show no significant correlation at short time-scales, in support of this randomization assumption.

With the operating assumption that the relative donor-acceptor orientation randomizes on a time scale much faster than inter-photon timing, each detected photon can be considered as an instantaneous sampling of the probability density function for κ^2 [29].

$$p(\kappa^2) = \frac{2}{\sqrt{3\kappa^2}} \left[\ln\left(2 + \sqrt{3}\right) - g(\kappa^2) \right], \tag{15.1}$$

where $g(\kappa^2) = 0$ when $0 < \kappa^2 < 1$ and $g(\kappa^2) = \ln\left(\sqrt{\kappa^2 - 1} + \sqrt{\kappa^2}\right)$ when $1 < \kappa^2 < 4$. The excitation-emission cycling within a single molecule then allows repeated sampling of different relative orientations. In using a set of photons for a distance measurement, the effective κ^2 for the measurement will be the mean, $\langle \kappa^2 \rangle$. As shown by the numerical study presented in Fig. 15.1, as few as ~ 10 photons are required before the central limit theorem takes effect and the means approach the $\langle \kappa^2 \rangle \rightarrow 2/3$ limit. This implies that if the spectra of the dyes and the refractive index of the medium between the dyes do not change appreciably over the course of the experiment, the measured distance is linearly related to the actual molecular distance in any one trajectory.

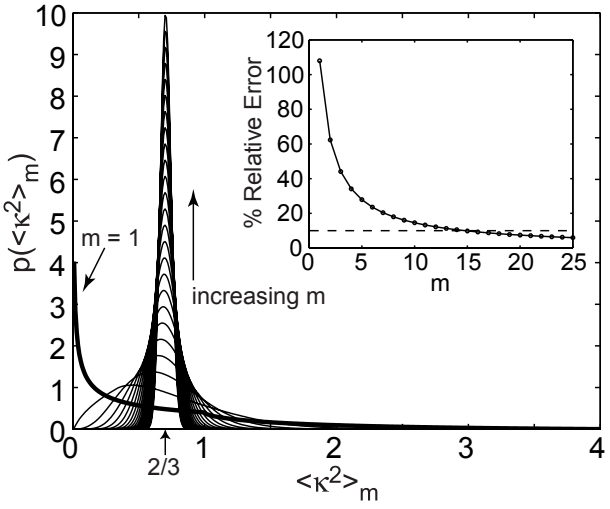


Figure 15.1: Probability density $p(\langle\kappa^2\rangle_m)$ as a function of the m-measurement mean value $\langle\kappa^2\rangle_m$. The bold line highlights the $m = 1$ function. As m increases, the mean value rapidly peaks at the ensemble-averaged value, $\langle\kappa^2\rangle_{m\to\infty} = 2/3$. The inset shows the relative standard deviation of $\langle\kappa^2\rangle_m$ as a function of m, which quickly decreases to less than 10% (at $m \simeq 15$).

15.2 Molecule-to-molecule variations are dominated by parameter calibration uncertainty.

The underlying distribution of FRET efficiency and distance within individual molecules (relative distribution) can be reliably recovered with combined use of MIM and MaxEnt deconvolution. This permits one to begin discussing potential complications related to variations between molecules. The results are summarized in Fig. 15.2.

For this data set, in general, a broad molecule-to-molecule variation is observed in the measured absolute distances $\hat{h}_n(x)$ and energy transfer efficiencies $\hat{h}_n(E)$. The width of the distance distribution for a given oligopeptide exceeds what would have been expected from statistical errors in the MIM analysis of individual trajectories. Control experiments using linearly polarized excitation light at 0°, 45°, and 90° at the same molecule resulted in trajectories of constant intensity within measurement uncertainties after correcting for depolarization effects in the optical components. This observation rules out the scenario in which either the donor or the acceptor probe is locked in a fixed orientation during observation period.

Instead, it was found that the spread was dominated by variation in the observed calibration values (I_d^β, I_a^β, β_d, and β_a in Eqs. 13.3 and 13.2). These calibration related uncertainties may result from variations in locating individual molecules from the single-molecule image, or from variations in the immediate chemical environment of the molecule under investigation. This is visualized in Fig. 15.2 by comparing the distributions of $\hat{h}_n(x)$

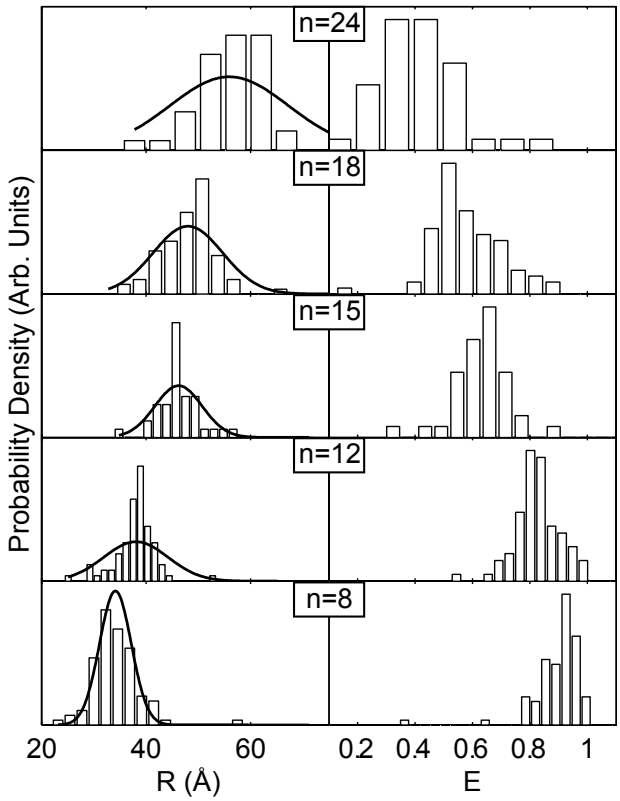


Figure 15.2: Distributions of donor-acceptor distances and FRET efficiency of P_nCG_3K(biotin), $(n = 8, 12, 15, 18, 24)$. The solid line is a Gaussian distribution with a variance that is the mean of expected variance of individual molecules by propagating uncertainties in parameters calibration. This indicates that the molecule-to-molecule distributions are dominated by uncertainties in parameter calibration.

from all molecules (bars) with a Gaussian distribution (thick solid lines) having a variance of $\sigma^2(\hat{x}) = \frac{1}{M}\Sigma_{j=1}^{M}\sigma_{\hat{x}}^2(j)$, where M is the total number of molecules of a given poly(L-proline) length, and $\sigma_{\hat{x}}^2(j)$ is the expected variance of the distance measure for the j-th molecule by propagating errors in parameters calibration. More accurate measurements such as those using multispectral methods [90] will be needed in order to address issues such as the shape of the molecule-to-molecule distribution. Indeed, it will be interesting to examine the possibility that individual polyproline molecules exist in different conformation and do not interconvert on the time scale of observation. One likely physical origin is that the number of *cis*- residues contained in individual polyproline molecules may vary from molecule to molecule, resulting in such a broad end-to-end distance distribution. Work along this direction is underway. Here, we will focus on the trend of the mean end-to-end distance exhibited by the series of polyprolines, discussed below.

Chapter 16

Worm-Like Chain Model for Poly(l-Proline) Molecules with a Short Persistence Length

The experimentally determined mean donor-acceptor distances may be compared with those predicted by three different models for polymer chains. In all of these comparisons, a unit length increment of 3.12 Å from C_α to C_α will be used for calculating the contour length, l_c [27]. In order of decreasing rigidity, these models are (A) A rod-like poly(L-proline) which exhibits an effective persistence length $l_p \to l_c$. This model appears to be implicitly assumed in the original paper for the use of Förster-type resonance energy transfer as a bulk-level spectroscopic ruler [127]. A concurrently proposed theoretical model is also consistent with this rod-like picture for short poly(L-proline) chains [115], and is therefore included in this category. (B) A less rigid model with the widely-used persistence length of $l_p = 220$ Å for all-*trans*-poly(L-proline) [18, 19, 45]. (C) A flexible model with a $l_p = 23$ Å persistence length, derived from osmometric experiments on high molecular weight poly(L-proline) [93].

These models will be discussed in the framework of a statistical description of stiff-chain polymers, the worm-like chain (WLC) model. The expected end-to-end distance, $\langle R \rangle$, of the WLC model is calculated using the mean-field expression for its probability density [13, 58],

$$p(r;u) = \frac{4\pi\mathcal{N}r^2}{(1-r^2)^{9/2}} \exp\left(-\frac{3u}{4}\frac{1}{(1-r^2)}\right),$$

where $r = R/l_c$ and $u = l_c/l_p$. The normalization constant $\mathcal{N}$ is given by

$$\mathcal{N} = \frac{4s^{3/2}e^s}{\pi^{3/2}(4 + 12s^{-1} + 15s^{-2})},$$

with $s = 3u/4$. Thus, the expectation value of R is,

$$\frac{\langle R \rangle}{l_c} = \frac{4\sqrt{3u}(5+u) - 2e^{3u/4}\sqrt{\pi}(-10+3u)\mathrm{erfc}[\sqrt{3u}/2]}{\sqrt{\pi}[20 + 3u(4+u)]}, \tag{16.1}$$

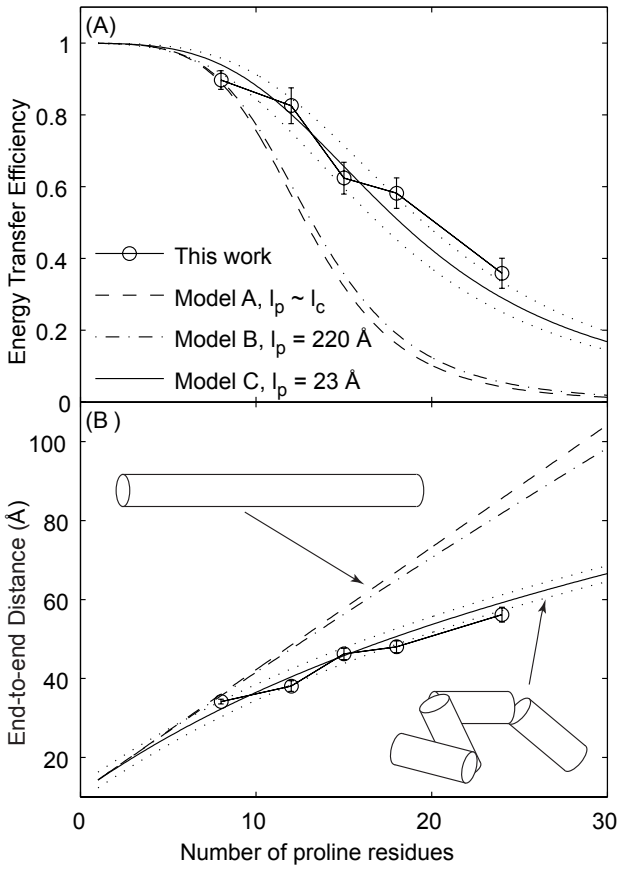


Figure 16.1: Comparison of experimental results with various models for poly(L-proline). As a reference, ± 2 Å ranges for the model are also displayed. Experimental error bars represent 95% confidence interval of the experimental mean.

where $\mathrm{erfc}[z] = \frac{2}{\pi}\int_z^\infty e^{-a^2} da$ is the complementary error function.

With the Förster radius $R_0 = 51$ Å determined from the spectral overlap of the donor and acceptor probes, the only remaining parameter to be determined is the distance between the center of the emitting dipole to the C_α to which the donor or the acceptor is tethered. Unfortunately, no chemical structure is available for the Alexa Fluor 555 and 647 dyes. Nevertheless, the structures for the coupling moieties, maleimide and succinimidyl ester (cf. Fig. 13.3), are known and can be used to estimate a lower bound for the linker distance. For this purpose, one counts 9 chemical bonds from either the N- or the C-terminus C_α for the linkers. The linkers are also described within the framework of the WLC model, using a 6.5-Å persistence length for polymethylene as an approximation and a C-C contour increment of 1.26 Å[149, 63]. Using Eq. 16.1, a lower bound of 12.2 Å for the joint linker distance was obtained. Constrained by this lower bound constraint, the WLC model with the $l_p = 23$ Å persistence length (model C) appears to describe the experimental data well, with a fitted linker distance of 11.23 Å, as summarized in Fig. 16.1.

It is evident that poly(L-proline) exhibits considerable flexibility even for the relatively short chains studied here. These results are consistent with the recent studies from diffusing single molecules [118] and from NMR experiments [68]: both found shorter-than-expected end-to-end distance if compared with a rigid poly-proline model. While poly(L-proline) is believed to exist predominantly in the *trans-* form in room-temperature aqueous solutions [64, 124], theoretical considerations indicate that the inclusion of 5% *cis-* residues in an otherwise *trans-* polyproline is sufficient to reduce the apparent persistence length significantly [130], Therefore, it is very likely that a small number of proline residues exist in the *cis-* form for the short chains studied here, giving rise to the observed flexibility.

Part V

Conclusion

A detailed and quantitative study of reactive dynamics in biomolecular systems requires both an method to turn experimental data into information about biomolecular conformation and a reliable estimate of the errors involved in that transformation. This work has taken general methods from the statistical literature, such as the Fisher Information and Maximum Likelihood Estimators, and applied them to single molecule data.

Part I used maximum likelihood estimators and the Fisher information to analyze photon by photon measurements of single molecule trajectories. The analysis achieved the best possible resolution given the constraint of making a model-free measurement of the experimental system.

The accuracy of a single photon counting experiment is determined by Poisson statistics. For example, if one is trying to measure the distance between two fluorophores in an experimental system, the distance information that one seeks is carried by each detected photon and is acquired at a constant rate in time. The rate of information acquisition will vary depending on the experimental configuration. Any measurement will be limited in precision by the amount of information obtained, as specified by the Cramér-Rao-Fréchet bound. By analyzing data in a way that achieves this limit, one can assure that the maximum information is extracted from the data.

The algorithm presented in Figure 5.1 achieves this goal, yet allows great flexibility in determining the relative values of the temporal and spatial resolutions. This concept is: 1), generally applicable to a variety of systems; 2), independent of kinetic models; 3), easy to implement in practical experiments; 4), efficient, since it extracts information photon by photon; 5), quantitative; and 6), most importantly, applicable to reactive systems.

Experimentally, it is helpful to remember that the algorithm is based on the detection of individual photons. The maximum information method relies upon the Poissonian noise that is inherent in photon counting applications. Arbitrary subtraction of background from the measured signal distorts the essential Poisson statistics. To increase experimental time resolution, all that must be done is to increase the excitation power, and thus the average detected intensity. Conversely, one can choose the intensity based on the desired time resolution. This allows one to take into account other experimental limitations, such as fluorophore photobleaching and triplet blinking.

Another advantage of this method is that it analytically propagates all potential distortions. For instance, in FRET measurements, cross-talk is now properly seen as merely a source of background noise. If cross-talk between the donor and acceptor channels is ignored, one would naturally excite the donor at the maximum in its absorption spectrum to give the highest signal/background ratio in a single molecule experiment. But the acceptor in a FRET pair frequently absorbs at that wavelength, producing cross talk. Since we now recognize cross talk as just another contribution to the background—i.e., its *only* effect is to decrease the ratio of signal/ background—it is no longer necessary to ensure that the acceptor is perfectly transparent at the excitation wavelength. With our information analysis, the excitation wavelength can be adjusted intelligently and the maximal signal/background ratio can be achieved.

All of these issues arise from the central idea of analysis grounded in statistics and information theory. In any experiment the fundamentally limited parameter is information. Since the amount of information does not increase, it is to the advantage of the

experimentalist to be as flexible as possible in choosing where to allocate that information. With our maximum information analysis the experimentalist is given optimal control over the information.

Part II continued this theme and shifted the context from continuous changes in intensity to sudden changes. 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.

The changepoint detection method detailed 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 arbitrariness in deciding whether a transition occurs, and 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 data quality is not unnecessarily reduced. It is therefore possible to achieve very high time resolution using our method. 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 non-stationary cases. This makes the changepoint method complementary to correlation function approaches [135, 151, 4], 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. Our study shows that the correct emission states can be *quantitatively* recovered from a single molecule trajectory.

Part III extended the photon-by-photon analysis of Parts I–II to two more scenarios: Probability Density Functions for molecular properties and intensity correlation functions. The calculation of accurate Probability Density Functions would not have been possible without the consistent, known errors calculated from the maximum information method. And the photon-by-photon calculation of intensity correlation functions follows directly from our approach in the rest of this work.

These ideas are perfectly general. Our new approach provides a framework for photon by photon analysis of change points that can be extended to the treatment of any signal. Once the statistical distributions underlying the data are known, the tests discussed above—and many other statistical tests—become directly applicable. We hope that this new method, which dramatically increases the amount of information that can be extracted from a single molecule trajectory, will allow new and unexpected phenomena to be observed and studied.

Finally, Part IV applied these techniques to polyproline as a model system.

While spectroscopy at the single-molecule level in principle allows the direct measurement of molecular property distributions, a quantitative determination of these distributions remain challenging, especially in time-dependent experiments. Uncertainties associated with low-light detection broaden and sometimes skew the experimentally obtained distribution. To address this issue, a deconvolution procedure has been developed using the distance dependent Förster-resonance energy transfer (FRET) as an illustrative example.

An uniformly broadened PDF is first prepared using the previously developed maximum-information approach. This amounts to equal-information binning and ensures that every point in the underlying histogram is broadened by the same amount. Straightforward deconvolution, attempting to make the best fit possible between the experimental data and the re-convoluted PDF, produces an over fit that is not supported by experimental data. The resulting deconvolved PDF is too rough and its features are too sharp. This leads naturally to the use of a maximum entropy-based method. It has two necessary components: statistical uniformity of the underlying data, already provided by the maximum information method, and accurate knowledge of the variance in the raw experimental histogram. The analytical expressions for the variance derived in this work provide the second requirement. It should be emphasized that the calculation of this variance takes into account the timescale of dynamics in the system under observation. Furthermore, the calculation of the full covariance matrix of the deconvolved PDF allows accurate assessment of the relative heights, widths, and importance of each observed mode. Prior assumptions about the functional form of the underlying probability density of the molecular parameter are no longer necessary for its accurate calculation.

For each single poly(L-proline) molecule studied here, sharply peaked distance and FRET efficiency distributions were observed, suggesting a time-independent end-to-end distance on the time scale of fluorescence single-molecule spectroscopy. This, in turn, allows discussion of molecule-to-molecule variations in the measured distance (FRET efficiency) on more quantitative terms. It was found that these variations were dominated by uncertainties in parameter calibration. The systematic study of a series of poly(L-proline) allows one to assess models of differing rigidity. It was found that a worm-like chain model with the $l_p = 23$ Å persistence length (derived from high molecular weight osmometry studies [93]) was in very good agreement with the present single-molecule results. While an all-*trans*-poly-proline chain is expected to exhibit a persistence length much longer than oligopeptides of other composition, it has been suggested that a small percentage of *cis*- residues would be sufficient to allow some flexibility in the otherwise rigid chain [130]. Indeed, the presence of *cis*- residues cannot be ruled out in room-temperature solutions. Therefore, an emerging picture for short poly(L-proline) chains is that they are composed of short *trans*- repeats interspersed with the occasional *cis*- residue. Longer time trajectories are expected to allow more detailed examination of this model, and to provide insights into the nature of the molecule-to-molecule variations.

The approach presented here for recovering the underlying molecular property distribution is general and is expected to be applicable to other experimental observables. With methods like this, an understanding of conformational features, as well as the dynamics within, may begin to be developed and placed within a quantitative, predictive theory. As

an example for future applications, this approach will allow quantitative comparison of the manner by which molecular property distributions may change as a result of changes in the underlying molecule, and to identify subtle yet functionally important molecular conformations.

Bibliography

[1] K. Adachi, R. Yasuda, H. Noji, H. Itoh, Y. Harada, M. Yoshida, and K. Kinosita. Stepping rotation of F-1-ATPase visualized through angle- resolved single-fluorophore imaging. *Proc. Natl. Acad. Sci. USA*, 97(13):7243–7247, 2000.

[2] N. Agmon. Conformational cycle of a single working enzyme. *J. Phys. Chem. B*, 104(32):7830–7834, 2000.

[3] M. Andrec, R. M. Levy, and D. S. Talaga. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden markov models. *J. Phys. Chem. A*, 107(38):7454–7464, 2003.

[4] V. Barsegov, V. Chernyak, and S. Mukamel. Multitime correlation functions for single molecule kinetics with fluctuating bottlenecks. *J. Chem. Phys.*, 116(10):4240–4251, 2002.

[5] V. Barsegov and S. Mukamel. Probing single molecule kinetics by photon arrival trajectories. *J. Chem. Phys.*, 116:9802–9810, 2002.

[6] A. P. Bartko and R. M. Dickson. Three-dimensional orientations of polymer-bound single molecules. *J. Phys. Chem. B*, 103(16):3053–3056, 1999.

[7] Andrew P. Bartko, Kewei Xu, and Robert M. Dickson. Three dimensional single molecule rotational diffusion in glassy state polymer films. *Phys. Rev. Lett.*, 89:026101–1–026101–4, 2002.

[8] W. Becker, H. Hickl, C. Zander, K. H. Drexhage, M. Sauer, S. Siebert, and J. Wolfrum. Time-resolved detection and identification of single analyte molecules in microcapillaries by time-correlated single-photon counting (TCSPC). *Rev. Sci. Instrum.*, 70(3):1835–1841, 1999.

[9] A. M. Berezhkovskii, M. Boguna, and G. H. Weiss. Evaluation of rate constants for conformational transitions using single-molecule fluorescence spectroscopy. *Chem. Phys, Lett.*, 336(3–4):321–324, 2001.

[10] A. M. Berezhkovskii, A. Szabo, and G. H. Weiss. Theory of single-molecule fluorescence spectroscopy of two- state systems. *J. Chem. Phys.*, 110(18):9145–9150, 1999.

[11] A. M. Berezhkovskii, A. Szabo, and G. H. Weiss. Theory of the fluorescence of single molecules undergoing multistate conformational dynamics. *J. Phys. Chem. B*, 104(16):3776–3780, 2000.

[12] Claude Berney and Gaudenz Danuser. FRET or no FRET: A quantitative comparison. *Biophys. J.*, 84(6):3992–4010, 2003.

[13] J. K. Bhattacharjee, D. Thirumalai, and J. D. Bryngelson. Distribution function of the end-to-end distance of semiflexible polymers. *arXiv:cond-mat/9709345*, 1997.

[14] V. Biju, M. Micic, Dehong Hu, and H. P. Lu. Intermittent single-molecule interfacial electron transfer dynamics. *J. Am. Chem. Soc.*, 126:9374–9381, 2004.

[15] Martin Böhmer, Francesco Pampaloni, Michael Wahl, Hans-Jürgen Rahn, Rainer Erdmann, and Jörg Enderlein. Time-resolved confocal scanning device for ultrasensitive fluorescence detection. *Rev. Sci. Instrum.*, 72(11):4145–4152, 2001.

[16] Gregory Bokinsky, David Rueda, Vinod K. Misra, Maria M. Rhodes, Andrew Gordus, Hazen Babcock, Nils G. Walter, and Xiaowei Zhuang. Single-molecule transition-state analysis of RNA folding. *Proc. Natl. Acad. Sci. USA*, 100:9302–9307, 2003.

[17] H. Boudjellaba, B. MacGibbon, and P. Sawyer. On exact inference for change in a poisson sequence. *Commun. Statist.—Theory Meth.*, 30:407–434, 2001.

[18] David A. Brant and Paul J. Flory. The configuration of random polypeptide chains. I. Experimental results. *J. Am. Chem. Soc.*, 87:2788–2791, 1965.

[19] David A. Brant and Paul J. Flory. The configuration of random polypeptide chains. II. Theory. *J. Am. Chem. Soc.*, 87:2791–2800, 1965.

[20] S. Brasselet, E. J. G. Peterman, A. Miyawaki, and W. E. Moerner. Single-molecule fluorescence resonant energy transfer in calcium concentration dependent cameleon. *J. Phys. Chem. B*, 104(15):3676–3682, 2000.

[21] Joanne Budzien, Colleen Raphael, Mark D. Ediger, and Juan J. de Pablo. Segmental dynamics in a blend of alkanes: Nuclear magnetic resonance experiments and molecular dynamics simulation. *J. Chem. Phys.*, 116:8209–8217, 2002.

[22] V. Buschmann, K. D. Weston, and M. Sauer. Spectroscopic study and evaluation of red-absorbing fluorescent dyes. *Bioconjugate Chem.*, 14:195–204, 2003.

[23] J. S. Cao. Event-averaged measurements of single-molecule kinetics. *Chem. Phys, Lett.*, 327(1-2):38–44, 2000.

[24] Jie Chen and A. K. Gupta. On change point detection and estimation. *Commun. Statist.-Simula.*, 30:665–697, 2001.

[25] Yu Chen, Dehong Hu, Erich R. Vorpagel, and H. Peter Lu. Probing single-molecule t4 lysozyme conformational dynamics by intramolecular fluorescence energy transfer. *J. Phys. Chem. B*, 107(31):7947–7956, 2003.

[26] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, Inc., New York, 1991.

[27] P. M. Cowan and S. McGavin. Structure of poly-l-proline. *Nature*, 176:501–503, 1955.

[28] Harald Cramér. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, 1946.

[29] R. E. Dale, J. Eisinger, and W. E. Blumberg. The orientational freedom of molecular probes the orientation factor in intramolecular energy transfer. *Biophys. J.*, 26:161–194, 1979.

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood for incomplete data via the em algorithm. *J. R. Stat. Soc. B*, 39:1–38, 1977.

[31] Laura A. Deschenes and David A. Vanden Bout. Single molecule photobleaching: increasing photon yield and survival time through supression of two-step photolysis. *Chem. Phys, Lett.*, 365(5–6):387–395, 2002.

[32] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. E. Moerner. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388:355–358, July 1997.

[33] M. Diez, B. Zimmermann, M. Börsch, M. König, E. Schweinberger, S. Steigmiller, R. Reuter, S. Felekyan, V. Kudryavtsev, C.A.M. Seidel, and P. Gräber. Proton-powered subunit rotation in single membrane-bound f_0f_1-atp synthase. *Nat. Struct. Mol. Biol.*, 11:135–141, 2004.

[34] Cristobal G. dos Remedios and Pierre D. J. Moens. Fluorescence resonance energy transfer spectroscopy is a reliable "ruler" for measuring structural changes in proteins dispelling the problem of the unknown orientation factor. *J. Struct. Biol.*, 115:175–185, 1995.

[35] L. Edman, Z. Foldes-Papp, S. Wennmalm, and R. Rigler. The fluctuating enzyme: a single molecule approach. *Chem. Phys.*, 247(1):11–22, 1999.

[36] L. Edman, Ü. Mets, and R. Rigler. Conformation transitions monitored for single molecules in solution. *Proc. Natl. Acad. Sci. USA*, 93:6710–6715, 1996.

[37] Bradley Efron. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7:1–26, 1979.

[38] Bradley Efron and Gail Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37:36–48, 1983.

[39] C. Eggeling, J. R. Fries, L. Brand, R. Gnter, and C. A. M. Seidel. Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proc. Natl. Acad. Sci. USA*, 95:1556–1561, 1998.

[40] C. Eggeling, J. Widengren, R. Rigler, and C.A.M. Seidel. Photobleaching of fluorescent dyes under conditions used for single-molecule detection: Evidence of two-step pholysis. *Anal. Chem.*, 70(13):2651–2659, 1998.

[41] J. Enderlein, P. M. Goodwin, A. van Orden, W. P. Ambrose, R. Erdmann, and R. A. Keller. A maximum likelihood estimator to distinguish single molecules by their fluorescence decays. *Chem. Phys, Lett.*, 270:464–470, 1997.

[42] Jörg Enderlein and Markus Sauer. Optimal algorithm for single-molecule identification with time-correlated single-photon counting. *J. Phys. Chem. A*, 105:48–53, 2001.

[43] Chris Farley. Algorithms for model-based gaussian hierarchical clustering. *SIAM J. Sci. Comput.*, 20:270–281, 1998.

[44] R. A. Fisher. Theory of statistical estimation. *Proc. Cam. Phil. Soc.*, 22:700–725, 1925.

[45] Paul J. Flory. *Statistical Mechanics of Chain Molecules*. John Wiley & Sons, New York, 1969.

[46] Th. Förster. Experimentelle und theoretische untersuchung des zwischenmolekularen bergangs von elektronenanregungsenergie. *Z. Naturforsch.*, 4a:321–327, 1949.

[47] Chris Fraley and Adrian E. Raftery. How many clusters? which clustering mothod? answers via model-based cluster analysis. *Comput. J.*, 41:578–588, 1998.

[48] Chris Fraley and Adrian E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.*, 97:611–631, 2002.

[49] M. Fréchet. Sur la convergence "en probabilité". *Metron*, 8(4):3, 1930.

[50] Joachirn R. Fries, Leif Brand, Christian Eggeling, Malte Köllner, and Claus A. M. Seidel. Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *J. Phys. Chem. A*, 102:6601–6613, 1998.

[51] E. Geva and J. L. Skinner. Theory of single-molecule optical line-shape distributions in low-temperature glasses. *J. Phys. Chem. B*, 101(44):8920–8932, 1997.

[52] E. Geva and J. L. Skinner. Two-state dynamics of single biomolecules in solution. *Chem. Phys, Lett.*, 288(2–4):225–229, 1998.

[53] Edit Gombay and Lajos Horvath. On the rate of approximations for maximum likelihood tests in change-point models. *J. Multivariate Anal.*, 56:120–152, 1996.

[54] I. Gopich and A. Szabo. Fluorophore-quencher distance correlation functions from single-molecule photon arrival trajectories. *J. Phys. Chem. B*, 109:6845–6848, 2005.

[55] Irina Gopich and Attila Szabo. Theory of photon statistics in single-molecule Förster resonance energy transfer. *J. Chem. Phys.*, 122:014707, 2005.

[56] F. Gornick, L. Mandelkern, A. F. Diorio, and D. E. Roberts. Evidence for a cooperative intramolecular transition in poly-L-proline. *J. Am. Chem. Soc.*, 93:1769–1777, 1971.

[57] H.B. Gray and J. R. Winkler. Electron transfer in proteins. *Ann. Rev. Biochem.*, 65:537–561, 1996.

[58] Alexander Grosberg, editor. *Theoretical and Mathematical Models in Polymer Research.* Academic Press, New Yrok, 1998.

[59] T. Ha, Th. Enderle, D. F. Ogletree, D. S. Schemla, P. R. Selvin, and S. Weiss. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. USA*, 93:6264–6268, 1996.

[60] T. Ha, J. Glass, Th. Enderle, D. S. Chemla, and S. Weiss. Hindered rotational diffusion and rotational jumps of single molecules. *Phys. Rev. Lett.*, 80(10):2093–2096, 1998.

[61] T. Ha, I. Rasnik, W. Cheng, H. P. Babcock, G. H. Gauss, T. M. Lohman, and S. Chu. Intiation and re-initiation of dna unwinding by the escherischia coli rep helicase. *Nature*, 419:638–641, 2002.

[62] Taekjip Ha. Single-molecule fluorescence resonance energy transfer. *Methods*, 25(1):78–86, 2001.

[63] L. Harnau, R. G. Winkler, and P. Reineker. On the dynamics of polymer melts: Contribution of rouse and bending modes. *Europhys. Lett.*, 45:488–494, 1999.

[64] William F. Harrington and Michael Sela. Studies on the structure of poly-L-proline in solution. *Biochim. Biophys. Acta*, 27:24–41, 1958.

[65] Yiyong He, T. R. Lutz, M. D. Ediger, Chakravarthy Ayyagari, Rmitry Bedrov, and Grant D. Smith. NMR experiments and molecular dynamics simulations of the segmental dynamics of polystyrene. *Macromolecules*, 37:5032–5039, 2004.

[66] Robin Henderson. A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika*, 77:835–43, 1990.

[67] Christian G. Hübner, Renn Alois, Renge Indrek, and Wild Urs P. Direct observation of the triplet lifetime quenching of single dye molecules by molecular oxygen. *J. Chem. Phys.*, 115:9619–9622, 2001.

[68] Jaison Jacob, Brian Baker, Robert G. Bryant, and David S. Cafiso. Distance estimates from paramagnetic enhancements of nuclear relaxation in linear and flexible model peptides. *Biophys. J.*, 77:1086–1092, 1999.

[69] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

[70] E. T. Jaynes. Information theory and statistical mechanics 2. *Phys. Rev.*, 108:171–190, 1957.

[71] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proc. IEEE*, 70(9):939–952, 1982.

[72] Yi-wei Jia, Alexander Sytnik, Liangguan Li, Serquei Vladimirov, Barry S. Cooperman, and Robin M. Hochstrasser. Nonexponential kinetics of a single t-RNA$^{\mathrm{Phe}}$ molecule under physiological conditions. *Proc. Natl. Acad. Sci. USA*, 94:7932–7936, 1997.

[73] Y. Jung, E. Barkai, and R.J. Silbey. Current status of single-molecule spectroscopy: Theoretical aspects. *J. Chem. Phys.*, 117(24):10980–10995, 2002.

[74] B. K. Kay, M. P. Williamson, and M. Sudol. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB*, 14(2):231–241, 2000.

[75] Malte Köllner and Jürgen Wolfrum. How many photons are necessary for fluorescence-lifetime measurement? *Chem. Phys. Lett.*, 200(1-2):199–204, 1992.

[76] Todd D. Krauss, Stephen O'Brian, and Louis E. Brus. Charge and photoionization properties of single semiconductor nanocrystals. *J. Phys. Chem. B*, 105:1725–1733, 2001.

[77] R. Kuhnemuth and C. A. M. Seidel. Principles of single molecule multiparameter fluorescence spectroscopy. *Single Mol.*, 2(4):251–254, 2001.

[78] Florian Kulzer and Michel Orrit. Single-molecule optics. *Ann. Rev. Phys. Chem.*, 55:585–611, 2004.

[79] M. Kuno, D.P. Fromm, H.F. Hamann, A. Gallgher, and D. Nesbitt. "on" / "off" fluorescence intermittency of single semiconductor quantum dots. *J. Chem. Phys.*, 115:1028–1040, 2001.

[80] A. D. Lanterman. Schwarz, wallace, and rissanen: Intertwining themes in theories of model selection. *International Statistical Review*, 69(2):185–212, 2001.

[81] Ted A. Laurence, Achillefs N. Kapanidis, Xiangxu Kong, Daniel S. Chemla, and Shimon Weiss. Photon arrival-time interval distribution (paid): A novel tool for analyzing molecular interactions. *J. Phys. Chem. B*, 108:3051–3067, 2004.

[82] M. Lee, J. Kim, J. Tang, and R. M. Hochstrasser. Fluorescence quenching and lifetime distributions of single molecules on glass surfaces. *Chem. Phys. Lett.*, 359:412–419, 2002.

[83] M. Lee, J. Y. Tang, and R. M. Hochstrasser. Fluorescence lifetime distribution of single molecules undergoing förster energy transfer. *Chem. Phys. Lett.*, 344:501–508, 2001.

[84] Nam Ki Lee, Achillefs N. Kapanidis, You Wang, Xavier Michalet, Jayanta Mukhopadhay, Richard H. Ebright, and Shimon Weiss. Accurate fret measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.*, 88:2939–2953, 2005.

[85] Brian G. Leroux. Consistent estimation of a mixing distribution. *Ann. Stat.*, 20:1350–1360, 1992.

[86] Li-Qiang Li and Lloyd M. Davis. Single photon avalanche diode for single molecule detecion. *Rev. Sci. Instrum.*, 64(6):1524–1529, 1993.

[87] M. Lippitz, C. G. Hübner, Th. Christ, H. Eichner, P. Bordat, A. Herrmann, K. Müllen, and Th. Basché. Coherent electronic coupling versus localization in individual molecular dimers. *Phys. Rev. Lett.*, 92(10):103001–1–103001–4, 2004.

[88] Markus Lippitz, Florian Kulzer, and Michel Orrit. Statistical evaluation of single nano-object fluorescence. *ChemPhysChem*, 6:770–789, 2005.

[89] H. Peter Lu, L.-Y. Xun, and X. Sunney Xie. Single-molecule enzymatic dynamics. *Science*, 282(5395):1877–1882, 1998.

[90] A. Khai Luong, Claudiu C. Gradinaru, David W. Chandler, and Carl C. Hayden. Simultaneous time- and frequency-resolved fluorescence microscopy of single molecules. *J. Phys. Chem. B*, 109(33):15691–15698, 2005.

[91] Doublas Magde, Elliot Elson, and W. W. Webb. Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.*, 29:705–708, 1972.

[92] M. Margittai, J. Widengren, E. Schweinberger, G. F. Schröder, S. Felekyan, E. Haustein, M König, D Fasshauer, H Grubmüller, R. Jahn, and C. A. M. Seidel. Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin-1. *Proc. Natl. Acad. Sci. U.S.A.*, 100(26):15516–15521, 2003.

[93] Wayne L. Mattice and Leo Mandelkern. Conformational properties of poly-L-proline form II in dilute solution. *J. Am. Chem. Soc.*, 93:1769–1777, 1971.

[94] E. Mei, S. Vinogradow, and R. M. Hochstrasser. Direct observation of triplet state emission of single molecules: single molecule phosphorescence quenching of metalloporphyrin and organometallic complexes by molecular oxygen and their quenching rate distributions. *J. Am. Chem. Soc.*, 125:13199–13204, 2003.

[95] W. E. Moerner and L. Kador. Optical-detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.*, 62(21):2535–2538, 1989.

[96] W.E. Moerner and D.P. Fromm. Methods of single-molecule fluorescence spectroscopoy and microscopy. *Rev. Sci. Instrum.*, 74(8):3597–3619, 2003.

[97] W.E. Moerner and M. Orrit. Illuminating single molecules in condensed matter. *Science*, 283:1670–1676, 1999.

[98] C.C. Moser, J.M.Keske, K. Warncke, R.S. Farid, and P.L. Dutton. Nature of biological electron transfer. *Nature*, 355(6363):796–802, 1992.

[99] R. G. Neuhauser, K. T. Shimizu, W. K. Woo, S. A. Empedocles, and M. G. Bawendi. Correlation between fluorescence intermittency and spectral diffusion in single semiconductor quantum dots. *Phys. Rev. Lett.*, 85:3301–3304, 2000.

[100] S. M. Nie and R. N. Zare. Optical detection of single molecules. *Ann. Rev. Biophys. Biomol. Struct.*, 26:567–596, 1997.

[101] M. Nirmal, B. O. Dabbousi, M. G. Bawendi, J. J. Macklin, J. K. Trautman, T. D. Harris, and L. E. Brus. Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature*, 383:802–1106, 1996.

[102] Marc Noé. The calculation of distributions of two-sided kolmogorov-smirnov type statistics. *Ann. Math. Statist.*, 43:58–64, 1972.

[103] Eugene Novikov, Johan Hofkens, Mircea Cotlet, Michael Maus, Frans C. De Schryver, and Noël Boens. A new analysis method of single molecule fluorescence using series of photon arrival times: theory and experiment. *Spectrochim. Acta*, 57(11):2109–2133, 2001.

[104] B. Okumus, T. J. Wilson, D. M. J. Lilley, and T. Ha. Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys. J.*, 87:2798–2806, 2004.

[105] J. N. Onuchic, J. Wang, and P. G. Wolynes. Analyzing single molecule trajectories on complex energy landscapes using replica correlation functions. *Chem. Phys.*, 247(1):175–184, 1999.

[106] M. Orrit and J. Bernard. Single pentacene molecules detected by fluorescence excitation in a para-terphenyl crystal. *Phys. Rev. Lett.*, 65(21):2716–2719, 1990.

[107] Art B. Owen. Nonparametric likelihood confidence bands for a distribution function. *J. Am. Statist. Assoc.*, 90:516–521, 1995.

[108] Prithwish Pal, John F. Lesoine, M. Andreas Lieb, Lukas Novotny, and Philip A. Knauf. A novel immobilization method for single protein spfret studies. *Biophys. J.*, page To Appear., 2005.

[109] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, New York, 1988.

[110] C. R. Rao. Sufficient statistics and minimum variance estimates. *Proc. Cambridge Phil. Soc.*, 45(4):213–218, 1949.

[111] Chandran R. Sabanayagam, John S. Eid, and Amit Meller. Using fluorescence resonance energy transfer to measure distances along individual dna molecules: Corrections due to nonideal transfer. *J. Chem. Phys.*, 122:061103, 2005.

[112] M. Sauer, K.H. Drexhage, U. Lieberwirth, Muller. R., S. Nord, and C. Zander. Dynamics of the electron transfer reaction between an oxazine dye and DNA oligonucleotides monitored on the single-molecule level. *Chem. Phys, Lett.*, 284(3-4):153–163, 1998.

[113] Greg K. Schenter, H. Peter Lu, and X. Sunney Xie. Statistical analyses and theoretical models of single-molecule enzymatic dynamics. *J. Phys. Chem. A*, 103(49):10477–10488, 1999.

[114] Mark J. Schervish. *Theory of Statistics*. Springer, 2 edition, 1997.

[115] Paul R. Schimmel and Paul J. Flroy. Conformational energy and configurational statistics of poly-l-proline. *Proc. Natl. Acad. Sci. U.S.A.*, 58:52–59, 1967.

[116] G. F. Schröder and H. Grubmüller. Maximum likelihood trajectories from single molecule fluorescence resonance energy transfer experiments. *J. Chem. Phys.*, 119:7830–7834, 2003.

[117] B. Schuler, E. A. Lipman, and W. A. Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419(6908):743–747, 2002.

[118] Benjamin Schuler, Everett A. Lipman, Peter J. Steinbach, Michael Kumke, and William A. Eaton. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. U.S.A.*, 102:2754–2759, 2005.

[119] Gideon Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, 1978.

[120] A. J. Scott and M. J. Symons. Clustering methods based on likelihood ratio criteria. *Biometrics*, 27:387–397, 1971.

[121] D. W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.

[122] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, New York, 1998.

[123] B. D. Slaughter, J. R. Unruh, M. W. Allen, R. J. B. Urbauer, and C. K. Johnson. Conformational substates of calmodulin revealed by single-pair fluorescence resonance energy transfer: Influence of solution conditions and oxidative modification. *Biochem.*, 44(10):3694–3707, 2005.

[124] Izchak Z. Steinberg, William F. Harrington, Arieh Berger, Michael Sela, and Ephraim Katchalski. The configurational changes of poly-L-proline in solution. *J. Am. Chem. Soc.*, 82:5263–5279, 1960.

[125] B. C. Stevens and T. Ha. Discrete and heterogeneous rotational dynamics of single membrane probe dyes in gel phase supported lipid bilayer. *J. Chem. Phys.*, 120:3030–3039, 2004.

[126] Lubert Stryer. Fluorescence energy transfer as a spectroscopic ruler. *Ann. Rev. Biochem.*, 47:819–846, 1978.

[127] Lubert Stryer and Richard P. Haugland. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci. U.S.A.*, 58:719–726, 1967.

[128] D. S. Talaga, W. L. Lau, H. Roder, J. Y. Tang, Y. W. Jia, W. F. DeGrado, and R. M. Hochstrasser. Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 97(24):13021–13026, 2000. NOV 21 PROC NAT ACAD SCI USA.

[129] Elliot Tan, Timothy J. Wilson, Michelle K. Nahas, Robert M. Clegg, David M. J. Lilley, and Taekjip Ha. A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Natl. Acad. Sci. USA*, 100(16):9308–9313, 2003.

[130] Seiji Tanaka and Harold A. Scheraga. Calculation of the characteristic ration of randomly coiled poly(L-proline). *Macromolecules*, 8:623–631, 1975.

[131] P. Tinnefeld, D. P. Herten, and M. Sauer. Photophysical dynamics of single molecules studied by spectrally-resolved fluorescence lifetime imaging microscopy (sflim). *J. Phys. Chem. A*, 105:7989–8003, 2001.

[132] D.A. van den Bout, W.T. Yip, D.K. Fu, T.M. Swager, and P.F. Barbara. Discrete intensity jumps and intramolecular electronic energy transfer in the spectroscopy of single conjugated polymer molecules. *Science*, 277:1074–1077, 1997.

[133] V. G. Voinov and M. S. Nikulin. *Unbiased Estimators and Their Applications*, volume I. Kluwer Academic Publishers, Boston, 1993.

[134] L. Yu. Vostrikova. Detection of a "disorder" in a wiener process. *Theory Probab. Appl.*, 26:356–362, 1981.

[135] J. Wang and P. Wolynes. Intermittency of single moleclue reaction dynamics in fluctuating environments. *Phys. Rev. Lett.*, 74(21):4317–4320, May 1995.

[136] Joe H. Ward, Jr. Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.*, 58:234–244, 1963.

[137] L. P. Watkins and H. Yang. Information bounds and optimal analysis of dynamic single molecule measurements. *Biophys. J.*, 86:4015–4029, 2004.

[138] L. P. Watkins and H. Yang. Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B*, 109:617–628, 2005.

[139] Lucas P. Watkins and Haw Yang. A continuous-wave photon-by-photon single-molecule spectrometer. *manuscript in preparation*, 2003.

[140] Shimon Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283:1676–1683, 1999.

[141] K. D. Weston and L. S. Goldner. Orientation imaging and reorientation dynamics of single dye molecules. *J. Phys. Chem. B*, 105:3453–3462, 2001.

[142] J. B. Witkoskie and J. S. Cao. Single molecule kinetics. I. Theoretical analysis of indicators. *J. Chem. Phys.*, 121:6361–6372, 2004.

[143] J. B. Witkoskie and J. S. Cao. Single molecule kinetics. II. Numerical bayesian approach. *J. Chem. Phys.*, 121:6373–6379, 2004.

[144] K. J. Worsley. The power of likelihood ratio and cumulative sum tests for a change in a bionomial probability. *Biometrika*, 70:455–464, 1983.

[145] K. J. Worsley. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73:91–104, 1986.

[146] K. J. Worsley. Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics*, 44:259–263, 1988.

[147] X. Sunney Xie and Jay K. Trautman. Optical studies of single molecules at room temperature. *Ann. Rev. Phys. Chem.*, 49:441–480, 1998.

[148] Z. Xie, N. Srividya, T. R. Sosnick, T. Pan, and N. F. Scherer. Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, 101(2):534–539, 2004.

[149] Hiromi Yamakawa. Stiff-chain macromolecules. *Ann. Rev. Phys. Chem.*, 35:23–47, 1984.

[150] Haw Yang, Guobin Luo, Pallop Karnchanaphanurach, Tai-Man Louie, Luying Xun, and X. Sunney Xie. Single-molecule protein dynamics on multiple time scales probed by electron transfer. *Science*, 302:262–266, October 2003.

[151] Haw Yang and X. Sunney Xie. Probing single molecule dynamics photon by photon. *J. Chem. Phys.*, 117:10965–10979, 2002.

[152] Haw Yang and X. Sunney Xie. Statistical approaches for probing single molecule dynamics photon by photon. *Chem. Phys.*, 284:423–437, 2002.

[153] S. L. Yang and J. S. Cao. Two-event echos in single-molecule kinetics: A signature of conformational fluctuations. *J. Phys. Chem. B*, 105(28):6536–6549, 2001.

[154] R. Yasuda, H. Noji, K. Kinosita, and M. Yoshida. F-1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degrees steps. *Cell*, 93(7):1117–1124, 1998.

[155] Ryohei Yasuda, Tomoko Masaike, Kengo Adachi, Hiroyuki Noji, Hiroyasu Itoh, and Kazuhiko Kinosita Jr. The ATP-waiting conformation of rotating F_1-ATPase revealed by single-pair fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA*, 100(16):9314–9318, 2003.

[156] Ahmet Yildiz, Joseph N. Forkey, Sean A. McKinney, Taekjip Ha, Yale E. Goldman, and Paul R. Selvin. Myosin v walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. *Science*, 300:2061–2065, 2003.

[157] X. W. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. J. Ha, D. Herschlag, and S. Chu. A single-molecule study of RNA catalysis and folding. *Science*, 288(5473):2048–2051, 2000.

[158] X. W. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu. Correlating structural dynamics and function in single ribozyme molecules. *Science*, 296(5572):1473–1476, 2002.

[159] Robert Zwanzig and Narinder K. Ailawadi. Statistical error due to finite time averaging in computer experiments. *Phys. Rev.*, 182:280–283, 1969.

Appendix A

FRET information

Maximum likelihood estimators are not guaranteed to be unbiased. In this section we calculate the bias in the estimators we have given and consider the effects it might have on the results obtained by the maximum information algorithm discussed in the body of this article. In general, the bias b_n in an estimator F_n of some parameter x is

$$b_n = \langle F_n \rangle - x. \tag{A.1}$$

Here n is the number of observations in the data set and $\langle \cdots \rangle$ indicates an average over all possible n-point datasets, weighted by the probability density of the observation of that dataset.

In the cases considered in this article, the probability density to observe a particular data set is Poisson. In the context of dynamic measurement, however, we are concerned more with the time in photon acquisition rather than the number of photons in a certain observation interval as indicated by the Fisher information. The bias in our estimators will therefore by given by

$$b\left(T\right) = \sum_{n_1 \ldots n_m = 0}^{\infty} F\left(n_1 \ldots n_m\right) \prod_{i=1}^{m} \frac{\left[I(x)T\right]^{n_i}}{n_i!} e^{-I(x)T} - x, \tag{A.2}$$

where, as before, m denotes the number of channels.

The estimators of FRET efficiency given in Eqs. 3.12 and 3.14 can be analytically shown to be unbiased. The sums for the estimators of distance in FRET and ET (given in Eqs. 3.13, 3.15, and 4.4) was evaluated numerically. A photon trajectory was generated at constant x, the photons were binned into time intervals T, and the appropriate estimator was applied. When negative, infinite, or imaginary distances were generated, the data point was discarded, just as in the maximum information algorithm in which T is increased until the set uncertainty level α is reached.

These calculations were performed at a variety of constant x-values. The results, plotted as a function of the information per bin, can be seen in Fig. A.1. Information per bin is the most natural coordinate for the bias plot in the context of the maximum information algorithm. As an example, if a trajectory is analyzed at $\alpha = 0.07$ (as in Fig. A.1, A and B), the information per bin is 204.

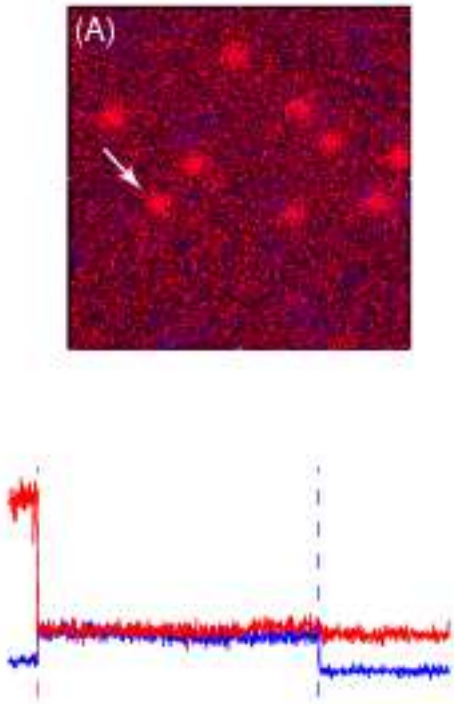


Figure A.1: Bias in estimators of x based on simulations of (A) one-channel FRET and (B) two-channel FRET at distances of 0.8, 1.0, and 1.2 and electron transfer at distances of (C) x= 9, 12, and (D) x 6 (- - -). Wedges are placed to indicate the values of the information (J) which satisfy the cutoff values of a 0.1 and a 0.07, as discussed in the main text.

There are strong fluctuations in the bias at extreme values of x in single-channel FRET and electron transfer. These arise as a consequence of the discrete nature of photons. Because photons are quantized, the possible values of the estimator at a given bin time are quantized. Since the estimator is highly nonlinear, the possible values of the estimator are also highly oscillatory as a function of the bin time. This oscillatory effect is most pronounced for the extreme values of the estimator. At extreme values of x, the average that is calculated to determine the bias is heavily influenced by the extreme value of the estimator. This produces the oscillations in bias. As the bin time increases, the number of possible values of the estimator also increases, and the oscillations damp out. For two-channel FRET, the estimator is not a function of bin size, so there are no oscillations.

For both one- and two-channel FRET, bias is smallest when $x \sim 1.0$. In all the curves the bias approaches zero as the information increases, confirming the asymptotic unbiasedness of these estimators. For ET at $x = 6$, the estimator is quite biased. At larger distances for ET, and at all distances for FRET measurements, the bias is at least an order-of-magnitude smaller than the standard deviation ($J^{1/2}$).

Appendix B

Computation of Critical Regions for Changepoint Detection

Under the $H_\circ$ 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. 8.2 picks up a change point is $\Pr\left(\mathcal{L}_k^\circ \geq \tau_{1-\alpha}^\circ(N); k = 1, \ldots\right.$ α, 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 $V_1^N \leq V_2^N \leq \cdots \leq V_{N-1}^N$ 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 [145]

$$\Pr\left(\mathcal{L}_k^\circ \leq \tau_{1-\alpha}^\circ; k = 1, \ldots, N-1 | T\right) = \Pr\left(a_k \leq V_k \leq b_k; k = 1, \ldots, N-1\right). \tag{B.1}$$

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 [102], but if a_k and b_k can be found it can be easily modified to calculate such probabilities as Eq. B.1. Owen has described an efficient way to implement Noé's algorithm [107]. In our implementation, a_k and b_k were found by numerically solving $\mathcal{L}_k^\circ = \tau_{1-\alpha}^\circ(N)$ (cf. Eq. 8.1) 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 [109]. 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. B.1 can be computed using Noé's algorithm. The critical region $\tau_{1-\alpha}^\circ(N)$ then can be found by numerically solving Eq. B.1 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 $Z_N^{1/2}$,

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

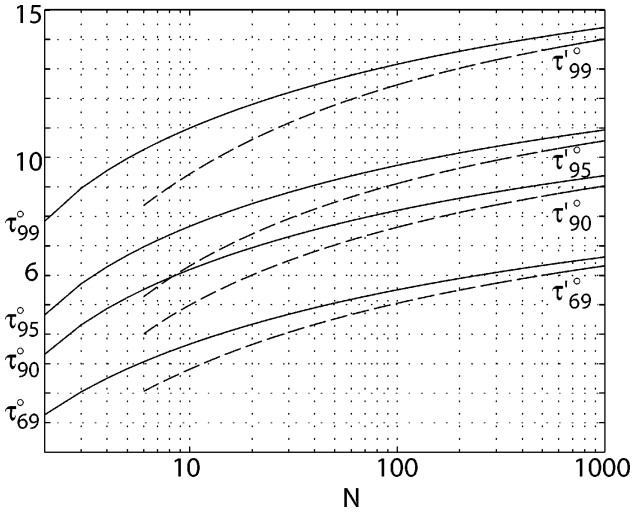


Figure B.1: Numerical values of critical regions $\tau^{\circ}_{1-\alpha}$ (—) as functions of N for Eq. 8.2. Critical regions $\tau'^{\circ}_{1-\beta}$ (−−) for conservative confidence intervals analogous to Eq. 8.6 are also plotted.

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

$$\Pr\left\{\sup_{0\le y\le Y} U(y) > \sqrt{\tau^{\infty}_{1-\alpha}}\right\} = \alpha, \tag{B.2}$$

where $0 \le y \le Y = \ln[(1-h)(1-l)/hl]$, $h(N) = l(N) = (\ln N)^{3/2}/N$, and the supremum, $\sup_{0\le y\le Y} U(y)$, is the least upper bound of $U(y)$. The analytical expression of the distribution of $U(y)$ in Eq. B.2 has been obtained by Vostrikova to give [134],

$$\alpha = \frac{(\tau^{\infty}_{1-\alpha})^{d/2} e^{-\tau^{\infty}_{1-\alpha}/2}}{2^{d/2}\Gamma(d/2)} \left[Y - \frac{d}{\tau^{\infty}_{1-\alpha}} Y + \frac{4}{\tau^{\infty}_{1-\alpha}} + O\left(\frac{1}{(\tau^{\infty}_{1-\alpha})^2}\right)\right], \tag{B.3}$$

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 $\tau^{\infty}_{1-\alpha}$ can be found by numerically solving Eq. B.3 and serves as the upper bound for finding $\tau^{\circ}_{1-\alpha}(N)$. We have computed the critical regions of Eq. 8.2 for $N = 1 \ldots 1000$ at $\alpha = 0.31$ (note that this does not correspond to one standard deviation but is included here only as a reference), $\alpha = 0.1$, $\alpha = 0.05$, and $\alpha = 0.01$. The results are plotted in Fig. B.1 as a function of N, a portion of which is listed in Table 8.1.

Analogous procedures can be used to find the critical regions $\tau_{1-\alpha}$ for Eq. 8.3 by numerically solving Eq. 8.4. The results are plotted in Fig. B.2, and partially listed in Table 8.1.

We next turn to computation of the critical regions $\tau'_{1-\beta}$ (or $\tau'^{\circ}_{1-\beta}$) for β-level confidence intervals once the intensity change point is found. The conservative confidence

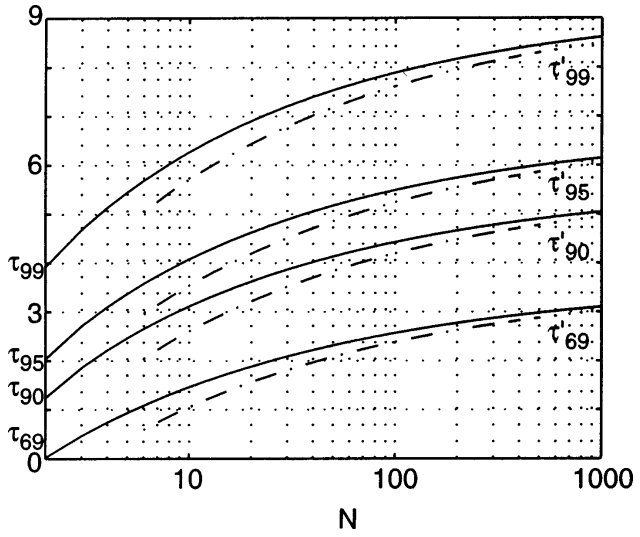


Figure B.2: Numerical values of critical regions $\tau_{1-\alpha}$ (—) and $\tau'_{1-\beta}$ (−−) as functions of N for Eqs. (8.3) and (8.6), respectively.

interval in Eq. 8.6 at various N and β can be found by solving

$$\begin{aligned} & \Pr\left(Z_N - \mathcal{L}_k \leq \tau'_{1-\beta}\right) = 1-\beta \\ = \; & \Pr\left(a_k \frac{N}{N-1} \leq V_k \leq b_k \frac{N}{N-1}; k = 1 \ldots N-2\right), \end{aligned}$$

where the scaling factor $N/(N-1)$ results from Worsley's approximation to conditionality of T and T_k in Eq. 8.5. 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 Figs. B.1 and B.2.

Finally, a note on the application of these results. Although the standardized and weighted likelihood ratio test in Eq. 8.3 offers a better performance than the classic test in Eq. 8.2, deviation of type-I error rate from ideality (cf. Fig. 8.1) 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 Supplementary Material.

Appendix C

Efron's bootstrap

An alternative method for the determination of errors in the raw histogram is the bootstrap method by Efron [37, 38]. Given the raw histogram, a set of auxiliary histograms is constructed by re-sampling each original data point from the raw histogram. The standard deviation $\sigma_B(x)$ of this set of auxiliary histograms has been shown, subject to certain assumptions, to be an good estimator of the error in the original histogram.

The assumption required by the bootstrap method is that all data points in the original histogram are independent and identically distributed. In single-molecule time trajectories, though, this assumption may not always be justified. If slow dynamics are being manifested in the trajectories being studied, the data points in the original histogram will not be independent, though they should be identically distributed. This means that, if the original number of data points is used to resample the raw histogram, the standard deviation determined will be significantly lower than is justified.

This oversight can be remedied by estimation of the dominant time-scales of the trajectories under consideration. In the spirit of Zwanzig's use of correlation times for calculation of errors, the number of uncorrelated distance estimates can be estimated by dividing the total duration of the trajectory by the $1/e$ time of the correlation function. The bootstrapped error calculated from this number of independent points is generally comparable with the error calculated by the analytical method described in the main text and may be more expedient in situations where the analytical approach cannot be applied, or where fast calculations are required.