# A distribution-based method to resolve single-molecule Forster resonance energy transfer observations

4 AUTHORS, INCLUDING:

E. Shane Price
William Jewell College

**13** PUBLICATIONS **143** CITATIONS

# A distribution-based method to resolve single-molecule Förster resonance energy transfer observations

Mihailo Backović,[1,a)] E. Shane Price,[2,b)] Carey K. Johnson,[2,c)] and John P. Ralston[1,d)]

[1]*Department of Physics & Astronomy. University of Kansas, Lawrence, Kansas 66045, USA*
[2]*Department of Chemistry. University of Kansas, Lawrence, Kansas 66045, USA*

We introduce a new approach to analyze single-molecule Förster resonance energy transfer (FRET) data. The method recognizes that FRET efficiencies assumed by traditional ensemble methods are unobservable for single molecules. We propose instead a method to predict *distributions* of FRET parameters obtained directly from the data. Distributions of FRET rates, *given* the data, are precisely defined using Bayesian methods and increase the information derived from the data. Benchmark comparisons find that the response time of the new method outperforms traditional methods of averaging. Our approach makes no assumption about the number or distribution of underlying FRET states. The new method also yields information about joint parameter distributions going beyond the standard framework of FRET analysis. For example, the running distribution of FRET means contains more information than any conceivable single measure of FRET efficiency. The method is tested against simulated data and then applied to a pilot-study sample of calmodulin molecules immobilized in lipid vesicles, revealing evidence for multiple dynamical states. © *2011 American Institute of Physics*. [doi:10.1063/1.3568946]

## I. INTRODUCTION

Experimental studies of protein dynamics moved from ensemble-based averages to single-molecule measurements for good reasons.[1] Single molecule observations probe dynamical features that are washed out in ensembles. Technical challenges of getting single-molecule data are significant, but they are not always the limiting factor. The challenges of making the most effective use of single-molecule data are a new frontier where much remains to be explored.[2]

FRET, standing for Förster resonance energy transfer,[3] is perhaps the most important tool for exploring single-molecule dynamics.[4–8] FRET probes protein configurations via time variations of fluorescent response. However, shot-noise limitations can degrade useful timing resolution by orders of magnitude compared to the intrinsic timing resolution of instruments. Analysis of single-molecule FRET has therefore used increasingly sophisticated statistical approaches to predict the FRET states underlying single-molecule measurements.[9–16] Analysis of dynamics is even harder. In this paper we explore a new method designed to make the most of the information available in FRET data while optimizing the possible time resolution.

## A. The FRET efficiency

The main tool of traditional FRET analysis is the efficiency $E_n$, defined by

$$E_n = n_a/(n_a + n_b), \tag{1}$$

where $n_a$ and $n_b$ are the number of detected photons in the acceptor or donor channels. The instantaneous estimate of $E_n(t)$ at time $t$ is quite meaningless, given that $n_a(t)$ and $n_b(t)$ are either 1 or 0 instant by instant. For this reason the values of $n_a$ and $n_b$ are commonly "binned" to make more stable estimators. For justification, it is common to appeal to maximum likelihood estimators.[17] Although data binning has a certain appeal, we will show that it can degrade performance without adding reliability.

In general, the concept of "intrinsic values" of $n_a(t)$, $n_b(t)$, and $E_n(t)$ is flawed, because *these quantities are not physically observable* for single molecules. To be specific, let $\mu$ be the mean for a sample with (for example) Poisson-distributed photon counts. We cannot possibly measure $n$ photons in a time $t$ and know $\mu$. At most we can learn something about the probability of $\mu$ via its probability distribution $P(\mu; t)$. Rather than the estimation of a single parameter at each time step, our goal is therefore to calculate a *probability distribution $P(\mu; t)$ evolving in time*.

Applying this notion to single-molecule FRET experiments, and considering what is observable, at best one could observe a *distribution of FRET efficiencies*, and this distribution may depend on time. At best, a single-molecule FRET experiment can only determine a *probability distribution* for the FRET efficiency of the single molecule system. Our goal is to implement a method that develops $P(E, t)$, *given the data*. This approach does not merely estimate a parameter value ($E$),

a)Electronic mail: mihailo@ku.edu.
b)Electronic mail: pricees@william.jewell.edu.
c)Electronic mail: ckjohnson@ku.edu.
d)Electronic mail: ralston@ku.edu.

but a time-dependent function, the probability distribution of *E*. This distribution provides a new way to explore what can be known from the data.

## B. Bayesian updating

The distribution-based method we describe uses the data in real time to update the distribution being estimated. This concept is Bayesian and implies active, real-time adaptation of the updating rules. A conceptual review of the Bayesian updating method is given in Ref. 18. In this paper we extend the method and test it against simulated data, comparing its performance to a simple maximum likelihood method, the running average. We also apply the method for the first time to the experimental FRET data on calmodulin, an important calcium-signaling protein.[19]

Our method converts running data of photon numbers $n_a(t)$, $n_b(t)$ into a time-dependent joint distribution $P(\mu_a, \mu_b; t)$. Starting from an assumed "prior" distribution (comprising what we already know about the distribution before data collection), the distribution is updated at each point based on the data. It is important that the output is a probability distribution rather than merely a parameter. Integration over the joint means distribution produces a running efficiency distribution $P(E, t)$. As previously emphasized, no particular FRET efficiency can possibly be observable, whereas the efficiency distribution is well defined from the data. Just as a wave function in quantum mechanics contains much more information than an expected value, the distribution $P(E, t)$ contains more information than any single efficiency parameter. The distribution can respond quickly to short term changes. It also provides useful global information about dynamical states of the system.

In this paper we show that the method enhances the timing resolution of the data, because much more information from the data is used from every run. For example, a 1000 point run treated by the time binning method might be reduced to 10 numbers, each representing an independent average over 100 photons. In our method a distribution having 20 $\mu$-bins (say) is produced at each time step. The 1000 points of the data are transformed into 20 000 pieces of information. There is no paradox in generating this much information. Suppose each time step is represented as (red, green) in the set (0, 0), (1, 0), or (0, 1). There are three possible values and $3^{1000} = 1.43 \times 10^{477}$ different patterns in 1000 points. Using all possible information about the actual state of the molecule, photon by photon can produce better short-term predictions. The task of converting the information received into a running distribution turns out to be straightforward and statistically robust and is described in Sec. II.

It is interesting that the time evolution predicted by Bayesian updating is not a strict "Markov" process. An important potential feature of single-molecule data is timing patterns or correlations in the data that come from fluctuations in single-molecule conformations.[10, 13, 16, 20] This is because hidden degrees of freedom coupled to fluorescence observables may affect the time evolution that is observed, leading to time correlations.[10, 20] Such correlations and fluctuations

are exactly the topic single-molecule experiments wish to explore. With Markov assumptions the statistical rules of evolution are independent of the history of the state. While attractive, the Markov assumption may be inappropriate to single molecule studies because it ignores information in the history of fluctuations. Allowing a role for fluctuations opens the door to improve single-molecule timing observations traditionally washed out by ensemble methods.

Several recent reports also highlight the utility of Bayesian analysis for single-molecule data. Cao and Witkoskie introduced a Bayesian Markov chain indicator to compare the probability of different kinetic models in accounting for single-molecule trajectories.[16] Landes *et al.* used Bayesian inference to identify periods of photoblinking by comparing the posterior probability of photoblinking and not-photoblinking, given single-molecule data, permitting photoblinking periods to be excluded from further analysis.[21] Ensign and Pande introduce a Bayes factor to evaluate the evidence for the existence of change points in single-molecule trajectories.[22] While the above methods were applied to single-channel data, a recent paper by Wiggins *et al.* described the application of a Bayesian approach to single-molecule FRET data.[23] Given a general model (e.g., hidden Markov), the method they describe predicts the number of underlying states as well as model parameter values. This approach may be very useful if the general nature of the underlying model is known. Cumulatively, the treatments described above, as well as the approach we report in this paper, illustrate the capability of Bayesian methods to add information beyond that available from maximum likelihood estimations. The approach we present here uses Bayesian methods in an iterative fashion to predict the evolving FRET probability distribution. Our method does not assume a number of underlying FRET states as a part of the model. Instead, using our method, the number of states or the distribution of FRET efficiencies appear as a result of data analysis. As desired, parameters such as the FRET efficiency transition times can be estimated from the probability distribution.

Section II describes the distribution-based method and Sec. III compares with a benchmark standard. We show that the procedure is not inherently sensitive to fine procedural details. Section IV describes the experimental method generating the data on calmodulin. The new method is applied to the calmodulin data in Sec. V. Despite using sample sizes that are rather small, the procedure develops distribution-based conclusions that are quite robust.

## II. THE EFFICIENCY DISTRIBUTION

The fundamental quantity of interest is the efficiency distribution developed from the distribution of rate parameters $\mu_a, \mu_b$. The efficiency itself is defined by

$$E_\mu = \frac{\mu_a}{(\mu_a + \mu_b)}. \tag{2}$$

For comparison, the integer-based efficiency distribution is defined by $E_n$ in Eq. (1). One cannot use $E_n$ and $E_\mu$ interchangeably. There is little doubt that $E_n$ creates a bias from

integer ratios, and $E_\mu$ is physically correct. Indeed the original Förster analysis was based on the underlying rate parameters, which corresponds to using $E_\mu$. In Subsection II A we show how our method produces the distribution $P(\mu)$, and henceforth we use symbol $E$ for "efficiency" consistently meaning $E_\mu$.

## A. The distribution of the mean

We use symbol $P(A \mid B)$ to be the normalized conditional probability (a distribution) of $A$, given $B$. We let $P(A)$ be the corresponding marginal distribution, implying integration over the variables not written. By definition the joint probability $P(A, B)$ is given by

$$P(A, B) = P(A \mid B)P(B). \quad (3)$$

Models commonly specify $P(n \mid \mu)$, namely, the distribution of $n$, given some mean parameter $\mu$. However, experiments seldom seek the probability of $n$ photons, given a mean. Most experiments seek $P(\mu \mid n)$, which is *the probability that the mean is $\mu$ given that n photons have been seen,* and subject to prior conditions. The concepts of "probability of the mean" and "probability of the rate" are made precise using Eq. (3) and Bayes' theorem:

$$P(\mu \mid n)P(n) = P(\mu, n) = P(n \mid \mu)P(\mu);$$

$$P(\mu \mid n) = \frac{P(n \mid \mu)P(\mu)}{P(n)}. \quad (4)$$

The denominator in Eq. (4) amounts to an overall normalization factor. The definition of $P(\mu)$ must be self-consistent:

$$P(\mu) = \sum_n P(\mu, n) = \sum_n P(\mu \mid n)P(n). $$

As shown in the Appendix, an appropriate model is a local-plus-broad distribution:

$$P(\mu \mid n) = (1 - \alpha)P_{\text{local}}(n \mid \mu)P(\mu) + \alpha/N_{\mu-\text{bins}}. \quad (5)$$

The local distribution $P_{\text{local}}(n \mid \mu)$ can be chosen by the user, while we use the Poisson distribution $P_{\text{local}}(n \mid \mu) \rightarrow \mu^n e^{-\mu}/n!$ for much of our data analysis. Symbol $\alpha$ is a parameter that is small compared to 1 and which is derived from the probability the system makes by a significant transition involving broad distributions $P_{\text{broad}}(\mu), P_{\text{broad}}(n)$. For all practical purposes, $P_{\text{broad}}(\mu), P_{\text{broad}}(n)$ can be approximated with a flat distribution over the entire range of $n, \mu$. The parameter $\alpha$ sets the sensitivity to the new information in the data.

We will show that updating is rather insensitive to the details of $P_{\text{broad}}$ and the constant $\alpha$, which we call the "annealing parameter" by a thermal physics analogy. For one thing, rather small parameters $\alpha \sim 1\%$ suffice to get relatively sensitive response to transitions. We report on calculations using a random positive distribution $P_{\text{broad}} \sim P_{\text{ran}}(\mu)$ and a flat distribution.

Equation (5) gives us a prescription for an iterative algorithm capable of "learning" from data. With each incoming data point $t$, the distribution will follow the iteration

$$P(\mu \mid n(t)) = (1 - \alpha)P_{\text{local}}(n(t) \mid \mu)P(\mu \mid n(t-1))$$

$$+ \alpha/N_{\mu-\text{bins}}. \quad (6)$$

We turn to obtaining $P(\mu, t)$ from the data using Bayesian updating.

### 1. Single-channel updating

The self-consistent determination of the distribution $P(\mu, t)$ is done by using the data itself as it arrives at each moment $t$:

- Begin at time step $t = 0$ with a "prior" or "seed" distribution $P(\mu) = P_0(\mu, 0)$. $P(\mu)$ is numerically defined on discrete steps $\mu_1 \mu_p$, and normalized by $\sum_j^p P(\mu_j) = 1$. Some reasonable tuning of the $\mu$ range and discretization can be expected: indeed, it is part of improving upon a maximum likelihood method, as described below. Over-populating the number of bins in $\mu$ has no effect other than increasing the computational load. Unless data sets are very small (a few points), the details of the seed distribution matter very little, and will be overcome by the data after a few steps. We will document this.
- At time step 1, get the data $n(1)$ and calculate $P_1(\mu \mid n(1))$ using Eq. (6). This produces the probability of $\mu$, *given that $n(1)$ was observed and given* the previous history.
- At time step 2, get the data $n(2)$ and use $P_1(\mu \mid n(1))$ as the updated prior for step 2.
- Iterate, producing $P(\mu, t) = P_t(\mu \mid n(t))$. After a few steps the procedure "learns" the data distribution, and the seed distribution becomes irrelevant.

A detailed example calculation, showing how to implement a numerical version of the updating algorithm can be found in Appendix B.

### 2. General scope of the description

Assuming a Poisson process, the time-dependent number distribution $P(n, t)$ is given by integrating over the distribution of $\mu$:

$$P(n, t) = \int d\mu \, \frac{\mu^n e^{-\mu}}{n!} P(\mu, t). \quad (7)$$

It is interesting that the formula suggests $P(n, t)$ should *not* generally be distributed by the classic Poisson rule, save the exceptional circumstance $P(\mu) \rightarrow \delta(\mu - \bar{\mu})$. This condition is unlikely to apply except under very narrow (and probably uninteresting) conditions. In general, we expect a distribution of rates $\mu$ resulting from heterogeneous and fluctuating interactions of the fluorophore with its surroundings.

One might ask what general class of functions has the representation of Eq. (7), and, in particular, if some data set might be incompatible. A little algebra resolves this. Let $\mathcal{P}(\mu) = e^{-\mu}P(\mu)$ and $\mathcal{P}(n) = n!P(n)$, suppressing label $t$.

Equation (7) becomes

$$\mathcal{P}(n) = \int_0^\infty d\mu \, \mu^n \, \mathcal{P}(\mu), \tag{8}$$

which happens to be a Mellin transform. It is invertible with

$$\mathcal{P}(\mu) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} dn \, \mu^{-n-1} \, \mathcal{P}(n). \tag{9}$$

The integration of Eq. (9) in the complex $n$-plane runs over a strip where Eq. (8) converges. When $\mu$ and $n$ are reduced to discrete variables the corresponding problem in linear algebra can always be solved. This shows that Eq. (7) is perfectly general, and always "exact," *given* the information about the distributions at hand. This fact is reassuring but only a side issue compared to the task of *estimating* the distribution of $\mu$ from data, which is our main topic.

### 3. Cumulative distributions

As in previous approaches, one of the main goals of our approach is the reduction of a data set to summaries capable of quantifying the type and number of states observed in the data set. For this we use the time-cumulative distribution of $\mu$, defined by

$$\bar{P}(\mu) = \frac{1}{\tau} \sum_t^\tau P(\mu, t). \tag{10}$$

The time-cumulative distribution is very robust when evaluated with many time steps $\tau \gg 1$. Each bin in $\mu$ tends to be stabilized by the central limit theorem.

### B. Two channel distribution

The two-channel process is a direct generalization of the single-channel case. Let $n_a$, $n_b$ be photon counts in channels $a, b$, with (idealized) Poisson means $\mu_a$, $\mu_b$. Define the joint probability of two rates:

$$P(\mu_a, \mu_b; t) = P_t(\mu_a, \mu_b | n_a(t), n_b(t)).$$

The subscript on $P_t$ is redundant but emphasizes that the shape of the function has been updated and evaluated at each time slice. At each instant $t$, the updating rule is

$$
\begin{aligned}
P(\mu_a, \mu_b | n_a(t), n_b(t)) &= (1 - \alpha) P(n_a(t), n_b(t) \,|\, \mu_a, \mu_b) \\
&\quad \times P(\mu_a, \mu_b \,|\, n_1(t-1), n_b(t-1)) \\
&\quad + \alpha / N_{\mu\text{-bins}}^2.
\end{aligned}
\tag{11}
$$

As in the single-channel case, the value of $P(\mu_a, \mu_b)$ at each step comes from the previous step. The distribution of FRET efficiencies is then

$$P(E, t) = \int d\mu_a d\mu_b \, P(\mu_a, \mu_b; t) \delta\left(E - \frac{\mu_a}{\mu_a + \mu_b}\right). \tag{12}$$

We turn to models developing the updating rule from the data.

### 1. The updating model

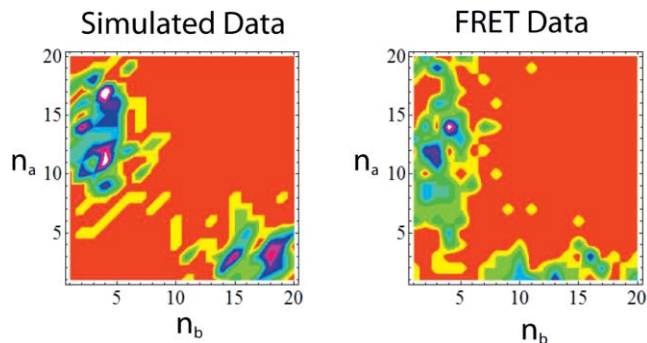FRET data taken at ordinary data rates and with good timing resolution has donor and acceptor channels that are



FIG. 1. Motivation for the *L*-zone distribution. Right panel: contour plot of a typical FRET data set. Left panel: contour plot of an equal number of points from the *L*-zone distribution described in the text. Parameters $\beta = 0.1$ and $\gamma = 0.2$ were used for the figure.

mutually exclusive. A molecule under FRET excitation either emits a "red" or a "green" photon, but not both. However, typical data, when binned over typical time intervals, does not show perfectly exclusive correlations. Simulations also need to take into account a finite probability that the emitted photon in either channel might not be observed.

We consult the time-cumulative $(n_a, n_b)$ distribution of a typical data set to proceed. Figure 1 shows the distribution for a subset selected as clean, as judged by strong signals and clear demarkation of the point of bleaching. Observe that the distribution of this data set tends to lie in an L-shaped zone. This does not imply that all data have the same distribution because probability may have moved or evolved within the L-shaped zone in any number of ways to give the same time-cumulative result. There is no reason to expect that the L shape is general. Other shapes may characterize other data, and the experimenter is free to adapt an updating model appropriate for the data.

Continuing, any *ansatz* capable of describing the probability inside the time-cumulative zone is suitable for updating purposes. Requirements are not particularly demanding. In general, an updating model should be able to accommodate the following:

- The ability to represent anticorrelation between acceptor and donor channels without imposing it automatically.
- A description of exclusive conditions where $n_a$, $n_b$ are as sharp as shot-noise permits.
- Capability of representing low numbers of photons in either channel, or both.
- Sensible upper limits on the number of photons in each channel; infinite numbers should be suppressed.

### 2. "L-Zone" distributions

A simple model satisfying these requirements is

$$
\begin{aligned}
P(n_a, n_b \,|\, \mu_a, \mu_b) = A \big( &e^{-\beta(n_a - \mu_a)^2 - \gamma(n_b - \mu_b)} \\
&+ e^{-\beta(n_b - \mu_b)^2 - \gamma(n_a - \mu_a)} \big).
\end{aligned}
\tag{13}
$$

The parameters $\gamma$, $\beta$ are estimated from the data. Parameter $\gamma$ cuts off the regions of $n \to \infty$. The L-zone distribution meets

each of the requirements enumerated above. The model can represent anticorrelation between acceptor and donor channels through the two probability zones represented by the Gaussians in $n_a$ and $n_b$. The "sharpness" parameter $\beta$ is chosen based on the data (second requirement), including low numbers in one or both channels (third requirement). We note that the L-zone distribution does not assume either high or low-FRET states because $\mu_a$, and $\mu_b$ are repeatedly updated and assume values driven by the data. Finally, the decay parameters restrict the upper limits for photon numbers in each channel (fourth requirement). So long as $\gamma$ is well-chosen it has no effects other than transferring wasted probability at infinity back into the experimental region. The width parameter $\beta$ is best obtained by fits to marginal distributions. One might introduce more than one $\beta$, $\gamma$ parameters at the cost of complication. The gist of Eq. (13) allows a doubly infinite number of locations of Gaussian bumps.

The Gaussian L-zone model was validated with Monte Carlo simulations. A summary of the simulations can be seen in Fig. 1. For the simulation, 3000 random pairs of numbers $(n_a, n_b)$ were drawn from the L-zone distribution with means $\mu_1 = 15$, $\mu_2 = 1$. The L-zone model is not unique. Any other localized functions can replace the Gaussians, and the model serves as a convenient starting point to illustrate a large and versatile class. The L-zone distribution also does not *a priori* assume a particular number of underlying FRET states. As a result, the distribution of FRET states appears as a product of the data-driven analysis, not as an input requirement. The need for assumptions about the number of FRET states are a known handicap of many Markov chain models. In this respect, our method using the L-zone distribution, or any other suitable distribution, has an advantage.

## III. BENCHMARK COMPARISONS

In this section we compare our procedure with the model "data" generated by simulations. Simulated data is used to characterize nominal states, establish the fluctuation of the detection statistic, and quantify response times. We find that the center of the data-derived efficiency distribution faithfully tracks the value of the efficiency put into the simulation. Under wide conditions we can deduce transitions in the data's efficiency values with an enhanced timing resolution compared to a simple running average.

Before reporting on those results we briefly compare standard benchmarks.

### A. Maximum likelihood

The appeal of maximum likelihood lies in its apparent objectivity. We review the simple case of estimating a *value* of a Poisson mean $\mu$ from a data set $n_1, n_2, \ldots, n_N$.

The likelihood $L_n$ of the data is the probability that uncorrelated data was seen given $\mu$:

$$L_n = P(n_1, n_2, \ldots n_N) = P(n_1 \,|\, \mu)P(n_2 \,|\, \mu)\ldots P(n_N \,|\, \mu).$$
$$(14)$$

The log-likelihood is the sum of the logs:

$$\log(L_n) = \sum_j^N \log\left(\frac{\mu^{n_j} e^{-\mu}}{n_j!}\right)$$

$$= \sum_j^N (n_j \log(\mu) - \log(n_j!) - N\mu.$$

The maximum likelihood estimate for $\mu$ comes from taking the derivative and setting it to zero:

$$\frac{\partial \log(L_n)}{\partial \mu} = \sum_j^N \frac{n_j}{\mu} - N \to 0;$$

$$\mu \to \frac{\sum_j^N n_j}{N} = \langle x \rangle_{\text{data}}. \quad (15)$$

Not surprisingly, the estimator found for the mean parameter is the mean of the data.

It is surprisingly common for the running average to be poorly suited to the experimental data analysis. Maximum likelihood has optimized *the probability $L_n$ of the data, given $\mu$*. But the point of the experiment is to optimize *the probability of $\mu$, given the data*. The calculations are not the same. To calculate the likelihood of $\mu$, denoted $L_\mu$ (note the subscript) after one event $n_1$, use the identities of Eq. (4) ("Bayes Theorem"), from which

$$P(\mu \,|\, n_1) = \frac{P(n_1 \,|\, \mu)P(\mu)}{P(n)}.$$

The right hand side contains $P(\mu)$, the prior information about the parameter $\mu$. With a certain distribution $P(\mu)$, the likelihood of $\mu$, given uncorrelated data $n_1, n_2 \ldots$ is

$$L_\mu = P(n_1 \,|\, \mu)P(n_2 \,|\, \mu)\ldots P(n_N \,|\, \mu)P(\mu). \quad (16)$$

Compare Eq. (16) to the usual likelihood function [Eq. (14)] which lacks a factor of $P(\mu)$. In its typical implementation, omitting $P(\mu)$ from the calculation is equivalent to asserting $P(\mu) \sim constant$ over the interval $0 < \mu < \infty$. It would be highly unusual for a flat distribution of mean parameters in the range $0 < \mu < \infty$ to be a realistic description of any data set. Indeed, the total probability spread over any finite region accessible to an experiment is *zero*.

Compare a calculation where $\mu < \mu_*$ is known to occur. Optimizing $\mu$ subject to $\mu < \mu_*$ will always produce a more accurate estimate than including $\mu \to \infty$. This is the weakness of many maximum likelihood estimators cited in the Introduction: No prepackaged statistic, uninformed about the actual data, could be expected to serve all purposes. By exploiting features learned from the data it is always possible to improve on standard maximum likelihood estimates. More sophisticated applications of maximum likelihood are possible.[13] In this paper, however, we choose to explore the performance of the Bayesian updating method, and we compare our Bayesian updating method to the maximum likelihood in its naive running average form.

Define a running average of a quantity $\xi(t)$ over smoothing time $\Delta T$ by

$$\bar{\xi}(t,\ \Delta T) = \frac{1}{\Delta T} \int_{t=t-\Delta T}^{t} dt\,\xi((t).$$

For discrete data we use the corresponding sum. The formula defines a backward average; other time offsets can also be used. The running average smoothes the data at a price of reducing response to changes over time.

As shown above, the running mean number $\bar{n}(t)$ is a maximum likelihood estimate for $\mu(t)$ that assumes a flat prior distribution over $0 < \mu < \infty$. The central limit theorem implies that under broad conditions the distribution of $\bar{n}$ will be a Gaussian centered at the true mean. The distribution width from averaging $N$ points is predicted to scale like $1/\sqrt{N}$. Apply this to estimate a mean using the running average:

$$\bar{\mu}_{\text{runav}} \sim \mu_0 \pm (\Delta\mu)_{\text{runav}};$$

$$(\Delta\mu)^2_{\text{runav}} \sim \frac{\sigma_\mu^2}{\bar{\mu}\,\Delta T}. \tag{17}$$

With idealized conditions of a steady system the experimenter can run for an unlimited time and reduce the error $\Delta\mu_{\text{runav}}$ to zero.

Yet we are interested in extracting dynamics from the fluctuations, not integrating them away. The uncertainty from making running averages tends to increase the inherent uncertainty that the underlying distribution $\sigma$ represents. We express this with an uncertainty relation

$$\bar{\mu}\,\Delta T\,\Delta\mu^2_{\text{runav}} \geq \sigma_\mu^2.$$

For any given value of $\sigma_\mu$ the condition for the smoothing time $\Delta T$ *not to seriously degrade* resolution is

$$\mu\,\Delta T \gg 1;$$

$$\Delta T \gg \frac{1}{\mu}. \tag{18}$$

This is a very modest requirement. It implies the possibility of averaging on the time scale of the single photon waiting time.

Under the impression it is statistically inevitable, experienced observers will impose $\mu\Delta T \gtrsim 100 - 1000$ to suppress $\sqrt{n}$ fluctuations. There are consequences from the "ordinary" uncertainty principle. Consider an instantaneous transition between two channels at time $t = 0$. Model this with short-term exponentials, starting in state 1 at $t = 0$:

$$n_a(t) \sim \bar{n}_a e^{-\gamma t};$$

$$n_b(t) \sim \bar{n}_b(1 - e^{-\gamma t}).$$

The Fourier resolution is

$$n_a(\omega) \sim \bar{n}_a \frac{1}{\omega + i\gamma}$$

$$n_b(\omega) \sim \bar{n}_b\left(\delta(\omega) - \frac{1}{\omega + i\gamma}\right).$$

The characteristic width in frequency is $\Delta\omega \sim \gamma$. The frequency–time uncertainty principle says $\Delta\omega\Delta T \gtrsim 1$. It is important that this operates independent of sampling limitations. Smearing data over an averaging time $\Delta T$ automatically eliminates the ability to resolve rates $\gamma \gtrsim 1/\Delta T$.

## B. Analysis of a single state

### 1. Dependence of the Bayesian updater on the priors

In this section, we study first the dependence of Bayesian updating on the initial priors. We find that Bayesian updating rapidly becomes insensitive to the choice of the prior. To illustrate the point, we first turn to analytic work. Let a data set consist of a set of $p$ numbers $(n_1, n_2 \ldots n_p)$. Let $P(\mu|n_0)$ be the initial prior and $P_{\text{broad}}(\mu)$ be a distribution obtained from random numbers, normalized to one, which appears in the term $\alpha P_{\text{broad}}(\mu)$ of the Bayesian updating algorithm. In general, after $m$ iterations the sequence looks like

$$P(\mu|n_m) = P(\mu|n_0)(1-\alpha)^m \prod_{k=1}^{m} P(n_k|\mu)$$

$$+ \alpha(1-\alpha)^{m-1} P_{\text{broad}}(\mu) \prod_{k=2}^{m} P(n_k|\mu)$$

$$+ \alpha(1-\alpha)^{m-2} P_{\text{broad}}(\mu) \prod_{k=3}^{m} P(n_k|\mu)$$

$$+ \cdots + \alpha P_{\text{broad}}(\mu). \tag{19}$$

Notice that the recursion relation for the Bayesian updater with annealing is highly nonlinear and does not have a simple closed form solution. The reason is that the number of terms increases with the number of iterations, so that after $m$ iterations the expression contains $m$ terms.

Only the first term in Eq. (19) contains the information about the initial prior, and it is suppressed by a product of $m$ probabilities and $m$ powers of $1 - \alpha$. It is also the only term in the updater containing the entire data history. Other terms involving more recent data are typically the dominant ones. As a result, even a delta function prior will eventually be "forgotten." To illustrate, we perform the following simulation. We draw 40 random numbers from a Poisson distribution with a mean $\mu_* = 3$ and analyze the data with the Bayesian updater. We repeat the procedure with different initial priors presented in Fig. 2. Figure 3 shows the result for $\alpha = 0.01$.

In contrast, the likelihood function [Eq. (14)] always contains the entire data history, rendering the method insensitive to abrupt changes in the underlying distribution $P(n_k|\mu)$, as documented below for the running average. Also, the extreme
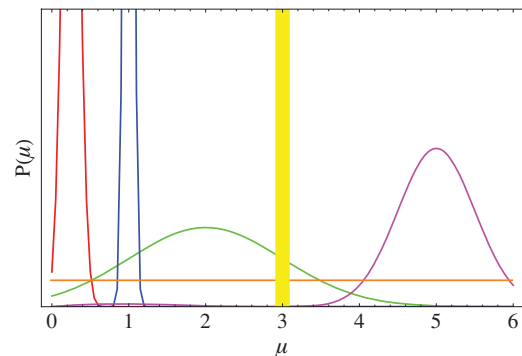


FIG. 2. Different priors used to test the Bayesian updater. The thick, yellow vertical bar represents the mean of the true underlying distribution $P(n|\mu)$.
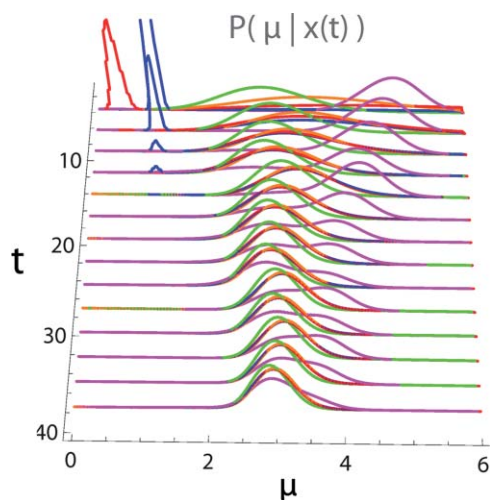
FIG. 3. Evolution of the probability distribution $P(\mu|x(t))$ for different priors given in Fig. 2. $\alpha = 0.01$ for the purpose of the graphic.

case of a delta function prior will, in principle, not allow the likelihood function to evolve past it.

The ability of the Bayesian updater to "forget" the prior is weakly dependent on the choice of the annealing parameter $\alpha$. Figure 4 shows slight differences between three different values of $\alpha = 0.01, 0.05, 0.1$. Different curves in the figure correspond to different priors. Even the most unfavorable prior becomes irrelevant in less than 10 to 20 time steps, depending on the choice of $\alpha$.

### 2. Fluctuations of a single state

For our *distribution-based method*, we compute $\sigma_\mu(t)$ and $\langle\mu\rangle(t)$ defined by

$$\langle\mu\rangle(t) = \int_0^\infty d\mu\, \mu\, P(\mu,\,t);$$

$$\sigma_\mu^2(t) = \langle\mu^2\rangle(t) - \langle\mu\rangle^2(t).$$

We compare the performance of the running average $\bar{n}(t)$, for which $\sigma_{\bar{n}}(t)$ is the standard deviation of $\bar{n}$ evaluated on the same smoothing time.

Each approach has a free parameter that balances stability against responsiveness. The running average depends on the smoothing time $\Delta T$ (the number of points averaged), developing smaller fluctuations but slower response for
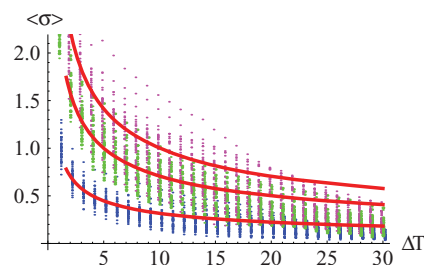


FIG. 5. Simulation results: Standard deviation of the running average mean $\bar{n}(t)$ as a function of smoothing time $\Delta t$. Simulation parameter $\mu_{1*} = 10, 5, 1$ (top to bottom). $\langle\sigma\rangle$ is the standard deviation of $\bar{n}(t)$ averaged over 50-point simulated trajectories. The vertical spread shows results for 50 runs. Curves are $\sqrt{\mu}_{1*}/\Delta T$. This study is done using the first half of simulation runs similar to one in Fig. 7, containing no transitions. Data points are slightly shifted horizontally for visual clarity.

longer $\Delta T$. Figure 5 shows the running average fluctuation $\langle\sigma\rangle$ as a function of $\Delta T$. This study used a data set of 50 numbers drawn from a Poisson distribution with constant means of $\mu_{*1} = 1, 5, 10$. Curves are the analytic prediction $\sigma_{\bar{n}} = \sqrt{\mu_*/\Delta T}$, which fits very well.

In the distribution-based method, $P(\mu, t)$ and its responsiveness depend on the annealing parameter $\alpha$. The distribution of $\sigma_\mu(t)$ over 50 data points is shown as a function of $\alpha$ in Fig. 6. These studies also use $\mu_{*1} = 1, 5, 10$. For small $\mu_* \lesssim 1$ the typical fluctuations tend to be comparable to those of the running average method (Fig. 5). However the fluctuations of the distribution-based method remain small and fixed even when the underlying distribution has $\mu_* \gg 1$. This appears to come from a "memory effect" retaining information about the history, and the larger number of degrees of freedom used in constructing $P(\mu)$.

### C. Analysis of multiple states

#### 1. Response time for a single transition

We simulated processes with transitions between two states as follows. Fifty random numbers were drawn from a Poisson distribution with mean $\mu_{*1}$. This sequence was joined to 100 or more numbers from a Poisson distribution with mean $\mu_{*2}$. The joint set made 100-point runs (allowing extra points for smoothing delays) that have an instantaneous transition at $t = 50$. We made 50 runs of every particular study. A few simulation runs suffice to draw conclusions, because
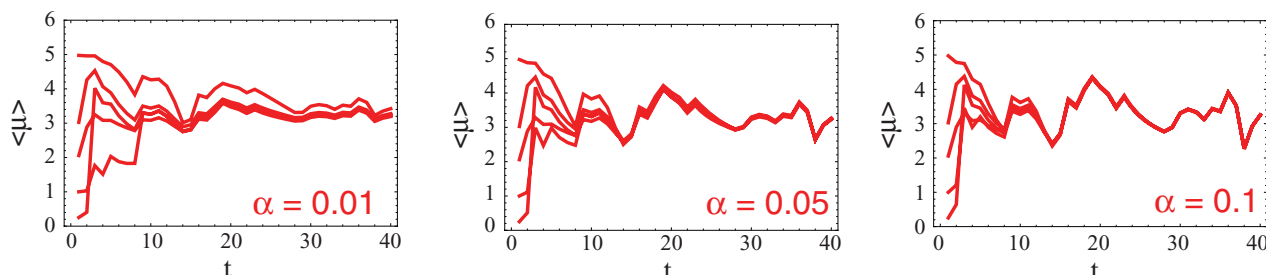


FIG. 4. Dependence of the Bayesian updater on the initial priors. Curves show the means from the Bayesian updater after each data point is analyzed. Different curves correspond to different initial priors $P(\mu|x_0)$ shown in Fig. 2. In all cases, the updater shows weak dependence on the initial prior. Annealing parameter (left to right) $\alpha = 0.01, 0.05, 0.1$.
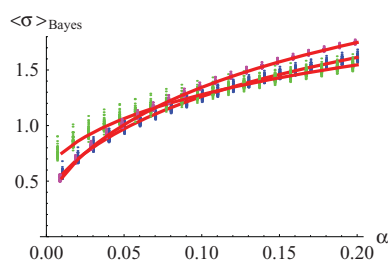
FIG. 6.  Simulation results for the distribution-based method. The distribution of $\langle \sigma(t) \rangle$ is shown as a function of the annealing parameter $\alpha$. Simulation parameters $\mu_{1*} = 1, 5, 10$. The vertical spread represents 50 runs. $\langle \sigma \rangle$ was averaged over the 50 data points. $\mu = 10$ (magenta online), 5 (green online), 1 (blue online). Curves (red online) show the average over the 50 runs.



FIG. 8.  Simulation results: Response times $t_{trans}$ using the running average method as a function of the smoothing parameter $\Delta T$. Points are the integer-valued response times; red (online) curve is the average of 50 runs.

every study is repeated many times with nearby parameter choices that probe the statistical fluctuations.

Figure 7 shows the response of both the Bayesian updater and the 10-point running average to a transition. A state is considered to be "detected" at time $t$ if $\langle \mu(t) \rangle$ lies within $1 \sigma_\mu(t)$ of the mean of the distribution generating it. A "transition" is signaled when the mean shifts to detect a different state. The "response time" $t_{trans}$ is the number of time steps between the actual transition and its detection. The fluctuation $\sigma_{\bar{n}}(t)$ has a trivial tendency to jump in the region of a significant transition. The figure illustrates the faster response of the Bayesian updater. Since a transition is signaled by $\bar{n}(t)$ lying within 1 unit of $\sigma_{\bar{n}}(t)$, the procedure exploiting the jump-related $\sigma_{\bar{n}}$ produces the most rapid detection signal we were able to construct. The statistical significance of any signal under large fluctuation criteria is naturally reduced, so that the procedure cannot be called "conservative." Instead the study establishes *minimum* timing resolutions possible with the running average.

The distribution over the first 50 time points serves as a negative control to show that the Bayesian updater does not show jumps or multiple states when none are present in the
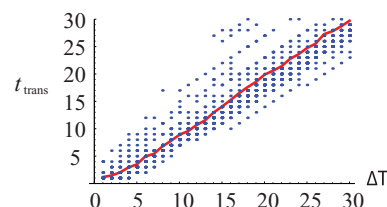
underlying simulated mean. The point is also made by the portion of the trajectory after the transition. The distribution predicted by the Bayesian updater, in fact, has a smaller standard deviation over these regions than the running average, as discussed above.

Running average response times are shown in Fig. 8. The simulation used full runs with $\mu_{*1} = 1$ and $\mu_{*2} = 5$. The average $t_{trans} = \Delta T$ scales accurately, with significant fluctuations about the mean. These simulations conform well to the analysis of the "uncertainty principle" controlling the running average method. In approaching those limits the running average performance is quite satisfactory for data generated by fixed and unique values of $\mu_*$. That is, the running average can be tuned into an excellent estimator *provided that knowing the mean is entirely equivalent to knowing the underlying $P(\mu, t)$ distribution.*

Our distribution-based method automatically retains information about the shape of $P(\mu, t)$. Response times are shown in Fig. 9. Response times are remarkably fast—a few time steps—and also quite insensitive to the value of the annealing parameter $\alpha$ over the entire range shown. Here is an example of superiority of the distribution-based method. To achieve the same average timing response by a running average as $t_{trans} \sim 2$ shown for $\alpha \sim 0.05$ requires a smoothing time $\Delta T \sim 1$ (Fig. 8). This is the most rapid and earliest
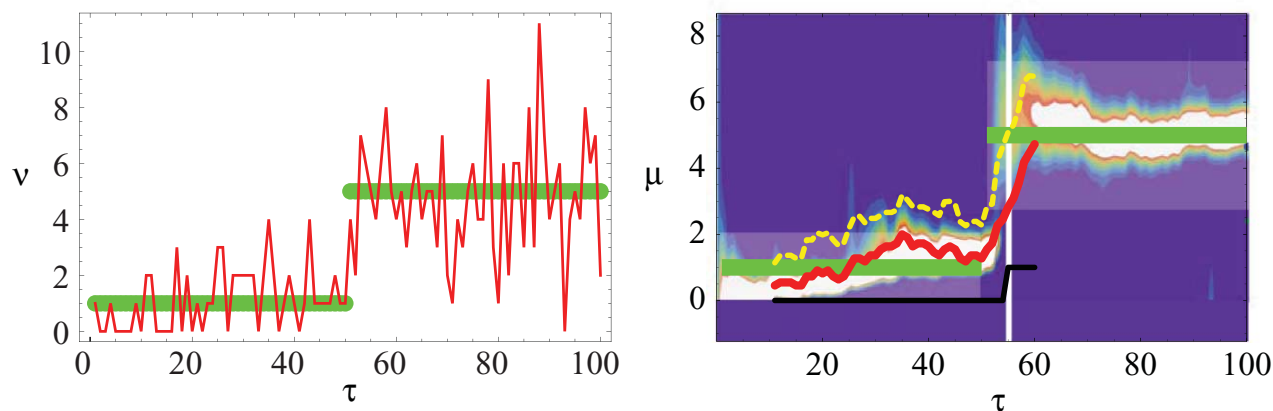


FIG. 7.  Left Panel: Raw data for a simulated two state transition. The green horizontal bars represent the states with $\mu_{1,2} = 1, 5$. Right panel: Typical response to a transition of the Bayesian distribution $P(\mu, t)$ (shown in color scale from blue to white) compared to the 10-point running average $\bar{n}(t) \pm \sigma_{\bar{n}}(t)$ (thick curve, red online) plus and minus the standard deviation (yellow online) represented by the dashed line. The running average is "backward-looking" and therefore increases at time points >50. The fluctuation $\sigma_{\bar{n}}$ is practically guaranteed to increase in the transition region, which makes for a particularly rapid detection scheme. "Hits" where $|\bar{n}(t) - \mu_{*2}| \leq \sigma_{\bar{n}}$ are shown as stepped curve (thin black line) of 1 (detection) or 0 (no detection). The thick horizontal lines (green online) are the means of the Poisson distribution, $\mu_{*1} = 1$ before and $\mu_{*2} = 5$ after the transition. Shaded regions represent the widths of the Poisson distributions, i.e., $\sqrt{\mu_*}$. The vertical line (white online) shows the time where a transition is detected in the running average, i.e., where the running average is $1 \sigma$ from the mean.
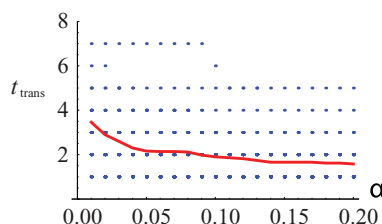
FIG. 9. Simulation results: Response times $t_{trans}$ from our distribution-based method as a function of annealing parameter $\alpha$. Dots are the integer-valued response times; red (online) curve is the average of 50 runs.

response possible. Yet it is certain to come with a fluctuation $\sigma_{\bar{n}} \gtrsim \sqrt{\mu}$ (Fig. 5). For $\mu \gtrsim 5$ the growing fluctuations of the running average method cannot compete with the higher stability of the distribution-based method.

### 2. Efficiency statistics of multiple state transitions

In this section we use simulations to compare the time response of distributions from the Bayesian updating method to the response of the running average. Figures 10 and 11 show typical time histories of multiple state transitions analyzed with the Bayesian updater and running average methods. The raw simulated data used in both figures is shown in Fig. 12.
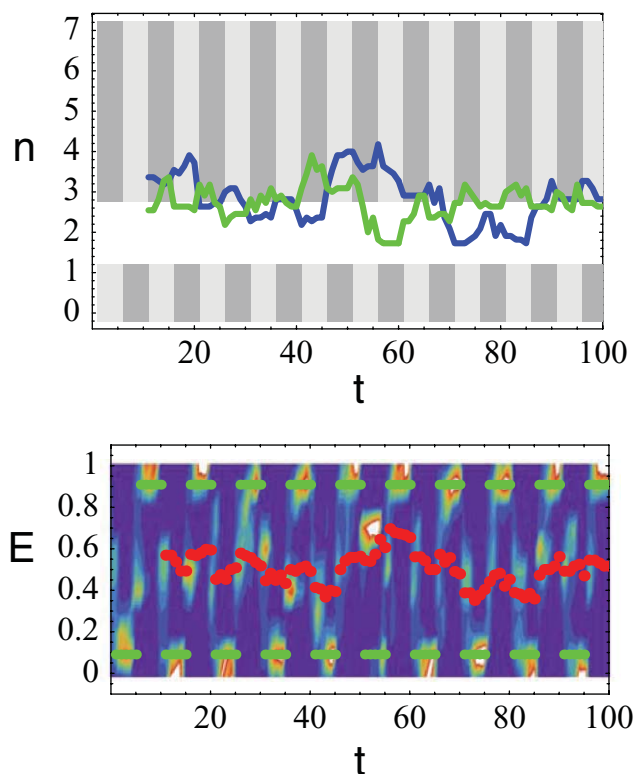




FIG. 10. Typical time histories of multiple state transitions analyzed with the Bayesian updater and the running average. Top panel shows the running average for $\bar{n}_a(t)$, $\bar{n}_b(t)$ smoothed using $\Delta T = 10$. Gray bars indicate the distribution parameters $\mu_*(t) \pm \sqrt{\mu_*(t)}$ oscillating on $t = 5$ units. Bottom panel shows the predicted Bayesian updater probability distributions in contour plot. Dots (red online) show the integer-based efficiency $E_n(t)$ obtained from the running average. Actual efficiencies are dashed horizontal bars (green online). $\alpha = 0.01$, $\beta = 2$, $\gamma = 0.2$ were used for the purpose of the graphic.
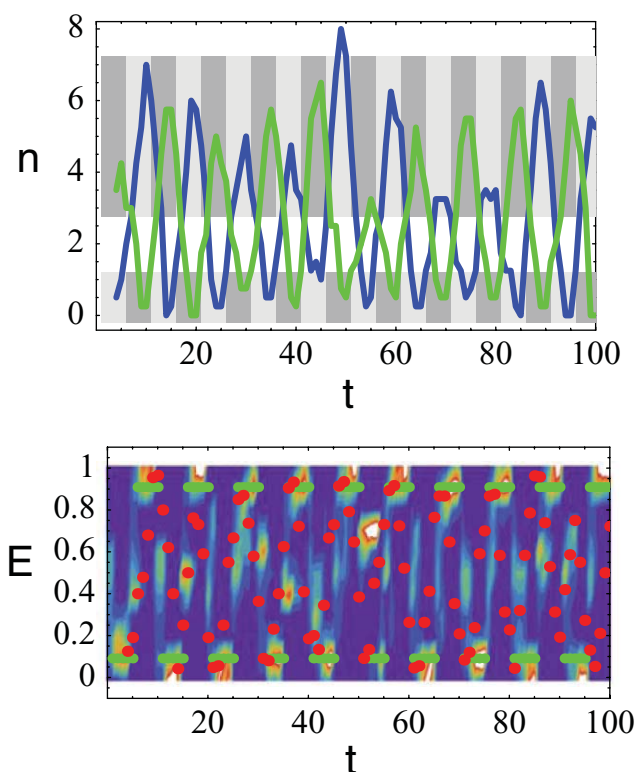




FIG. 11. Time history of multiple state transitions, as in Fig. 10, smoothed using $\Delta T = 3$.

The simulations were done by switching between $\mu_1 = 0.5$ and $\mu_2 = 5$ periodically on a time scale of 5 units. Response of the running average on smoothing times $\Delta T = 10$ and $\Delta T = 3$ are shown. The top panels of each figure show the smoothed $\bar{n}_a(t)$, $\bar{n}_b(t)$ oscillating in each channel. Gray bars indicate the parameters $\mu_*(t) \pm \sqrt{\mu_*(t)}$. The bottom panels show the distribution predicted by the Bayesian updater as contour plots. The distribution can be compared to the integer-based efficiency $E_n(t)$ computed from the running average, along with the actual efficiencies of the underlying distribution shown as horizontal dashed bars. The bottom panels show the distribution predicted by the Bayesian updater compared to the integer-based efficiency $E_n(t)$ running average, along with the actual efficiencies of the underlying distribution.

Several features are clear: (a) Resolving the short time variations of $\bar{n}_j(t)$ with small $\Delta T$ produces wild fluctuations in the efficiency $E_n(t)$ reported by the running average. (b) Smoothing out the short time variations of $\bar{n}_j(t)$ tames the
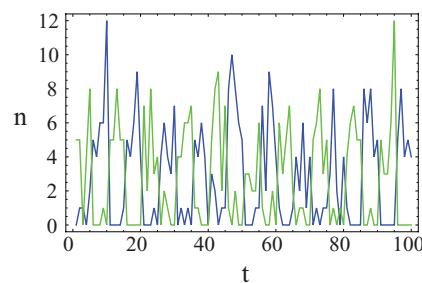


FIG. 12. Raw simulation data for the studies in Figs. 10 and 11.

efficiency fluctuations of the running average while producing a false compromise efficiency that does not represent the underlying distribution. Indeed, it is impossible for the running average method to produce anything other than a compromise when the data has any sort of nontrivial distribution. (c) It is difficult to extract much information from the running average for highly fluctuating $E_n(t)$. (d) The Bayesian updater is much more successful in tracking the underlying FRET efficiencies, showing distribution "peaks" at or near the correct underlying FRET efficiencies of 0.09 and 0.9. Although it also predicts some probability at intermediate FRET values due to the high interchange frequency, the peaks in the predicted distribution are near the correct values. The response time appears even faster than predicted in Fig. 9. The response time of the Bayes method depends on the data history and is probably shorter here because the transition occurs after a shorter time period at the previous efficiency value. Thus, under interchange rates high enough that the running average cannot resolve any population in the underlying high or low-FRET states, the distribution does detect the presence of these populations.

## IV. EXPERIMENTAL METHOD

Here we describe the experimental procedures. CaM T34C/T110C was expressed, purified, labeled, and separated following methods described previously.[24, 25] Purified CaM T34C/T110C was labeled simultaneously with Alexa Fluor 488 maleimide (AF488) as donor (D) and Texas Red maleimide (TR) as acceptor (A). The double-labeled construct (CaM-DA) was separated by reverse phase high-performance liquid chromatography. Mass spectrometry verified the expected dye labeling.

Calmodulin was encapsulated inside lipid vesicles by techniques described in Refs. 26 and 27. Solutions of phosphatidylcholine (PC) and 16:0 biotinyl cap phosphatidylethanolamine (biotin-PE) (Avanti Polar Lipids, Alabaster, AL) were prepared (10 mg/ml PC in chloroform and 1 mg/ml biotin-PE in chloroform). A 100:1 mixture of PC to biotin-PE was prepared and the chloroform was evaporated under nitrogen. CaM-DA was dialyzed into the buffer of choice at a final concentration of $\sim$25 nM. The desiccated lipid was then hydrated with the CaM solution and subjected to ten freeze/thaw cycles between a liquid nitrogen bath and a 30 °C water bath. Large unilamellar vesicles were formed by extrusion through a 100 nm pore membrane. The unencapsulated protein was separated from the vesicle entrapped protein with a Sepharose 4B column (GE Healthcare, Piscataway, NJ) equilibrated with the same buffer used to prepare the vesicles. The vesicles were used within 24 h of formation.

The high $Ca^{2+}$ buffer consisted of 10 mM HEPES, 100 mM KCl, 1 mM $MgCl_2$, and 0.1 mM $CaCl_2$, adjusted to pH 7.4 and filtered through a 0.2 $\mu$m syringe filter. The low $Ca^{2+}$ buffer was made up of 10 mM HEPES, 100 mM KCl, 1 mM $MgCl_2$, and 3 mM EGTA, adjusted to pH 7.4 and filtered.

Vesicles were immobilized in flow cells constructed from microscope cover slips. Two cover slips used as a spacer were sandwiched between cleaned top and bottom cover slips with a gap to form a channel with approximate dimensions 0.3 mm

$\times$ 4 mm $\times$ 22 mm. A lipid solution (20 mg/ml) was incubated in the flow cell for 30 min to form a lipid bilayer. The excess lipid was washed away with the buffer solution, and the flow cell was further incubated with a solution of 0.2 mg/ml streptavidin for 10 min, followed by another washing with the buffer. The final step was incubation with the extruded vesicles at an appropriate dilution for 5 min. Nonspecifically bound vesicles were thoroughly washed away with buffer and the finished flow cell was then placed on the microscope scanning stage. A well-made flow cell contained a $\sim$ 20 $\mu$l volume for at least 30 min. The buffer was replenished in the flow cell regularly to keep the sample hydrated during scanning.

Signals were collected by an inverted fluorescence microscope (Nikon TE2000) with a 60$\times$ objective lens (UPLSAPO 60XW Olympus, Center Valley, PA). The 488-nm line of an Ar ion laser (JDS Uniphase, Milpitas, CA) was directed through a Z488/10$\times$ excitation filter to the microscope dichroic filter (500DCXR). Red and green signals were separated with a 565DCLP FRET dichroic. Green emission was detected through an HQ535/50M green emission filter, and red emission through an HQ620/75M red emission filter. All filters were obtained from Chroma (Rockingham, VT). Signals in red and green channels were detected by avalanche photodiodes (SPCM AQR-14, Perkin Elmer, Vaudreuil, Quebec). Immobilized molecules were located by raster-scanning a 15 $\times$15 $\mu m^2$ region with a piezo-electric scanning stage (Mad City Labs, Madison, WI) with a laser power of 1 $\mu$W at 488 nm while detecting the signal from the red channel. Once an image was generated, the laser was shuttered and molecules were identified based on the intensity of the spots over the background. The scanning stage was then moved one molecule at a time to position each in the observation volume. The laser power was adjusted to 5 $\mu$W, the data collection software started, and the shutter opened until after the molecule under observation had undergone photobleaching. Counts from each channel were collected in the time-tagged mode and stored in computer memory. Trajectories were viewed in a LABVIEW program. For each single-molecule trajectory the start time (laser on) and the stop time (photobleach event) were ascertained by eye and recorded for further analysis. Multiple tracks were analyzed independently.

## V. RESULTS—IMMOBILIZED CALMODULIN EFFICIENCY DISTRIBUTIONS

### A. Time-dependent distributions of calmodulin

We analyzed 26 trajectories of immobilized calmodulin FRET data. Each trajectory comprises $n_a$, $n_b$ values sampled at 300 time steps. Data analyzed was raw: no background has been subtracted. Time step "0" was chosen by eye on the basis of the earliest significant rise in rates after applying full laser power. For the purpose of this analysis, we chose $\beta = 0.01$ and $\gamma = 0.01$, which was approximately fit from the marginal distributions of all samples. An annealing parameter $\alpha = 0.05$ was used in updating. Consistency checks showed little sensitivity to small variations of $\alpha$ just as suggested by the simulations. Each trajectory was treated independently. Raw data can be seen in Fig. 13.
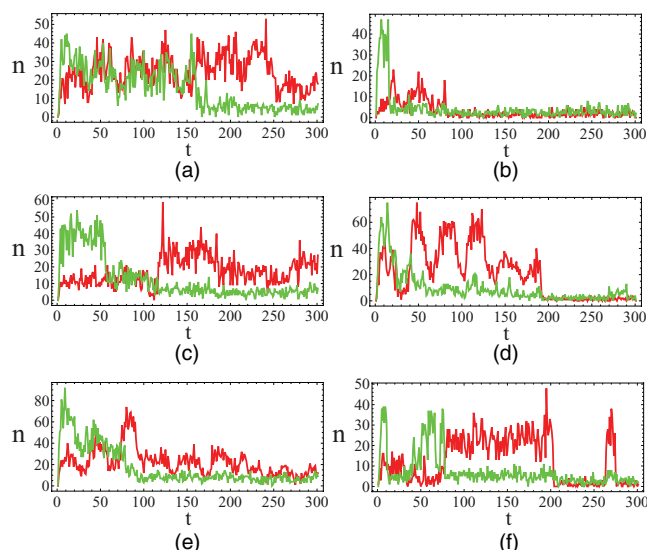
FIG. 13. Raw data used for analysis. Red lines are photon counts in the donor channel, green in the acceptor. Data is binned in $50\mu s$ bins.

Running probabilities $P(\mu_a, \mu_b \mid n_a, n_b, t)$ were developed over the range $0 \le \mu_1, \mu_2 \le 50$ in steps of 0.5. Thus, each time step is represented by a $100 \times 100$ distribution of $10^4$ values. The running efficiency distribution $P(E, t)$

was collected by binning the $E$-distribution on intervals of $\Delta E = 0.01$. We checked that the small effects of discretization of $\mu$ produced negligible bias with this binning.

Time histories of $P(E, t)$ for typical data samples are shown in Fig. 14. Each time history graphically displays considerably more information than it is possible to get from the running average. Meanwhile, the running averages $\bar{E}_n(t)$ (red curve online, using $\Delta T = 10$ units) also show a high degree of coincidence with the centers of the running distribution. Visual inspection of the plots suggests conformational transitions among multiple states.

## B. Time-cumulative distributions of calmodulin

The time-cumulative efficiency distributions developed by the analysis above were calculated with Eq. (10). Results corresponding to the six samples in Fig. 13 are shown in Fig. 15. Set by set there is a statistically significant signal that CaM FRET data consists of multiple states.

We collected the time-cumulative distributions of all 26 data sets for $0 < t < 300$ into one grand-cumulative distribution (Fig. 16). The grand cumulative distribution has a higher statistical significance than each particular cumulative distribution. It also tends to wash out features. We believe it would be possible to resolve features even better by more
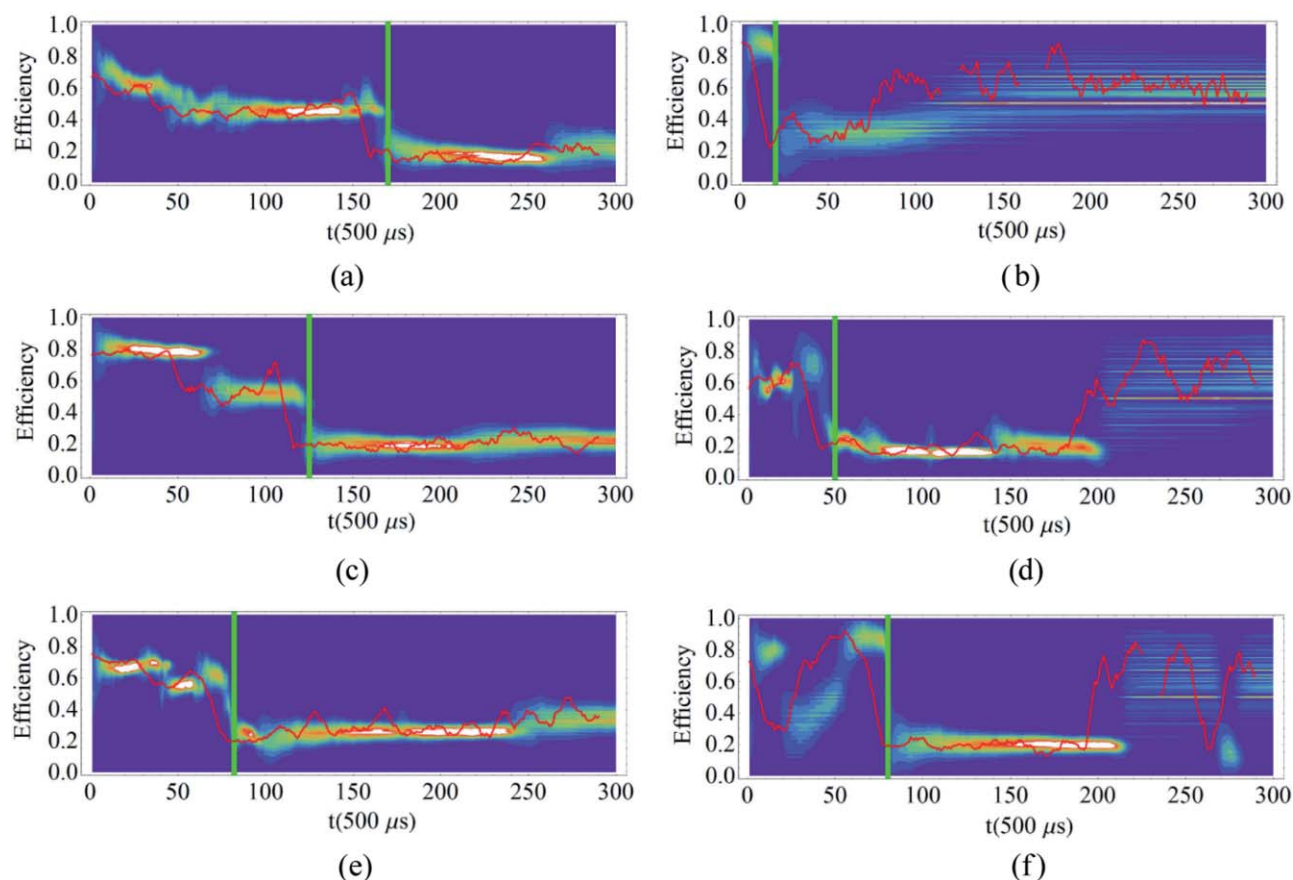


FIG. 14. Samples of running efficiency distribution $P(E, t)$ as a function of time for $\alpha = 0.05$. Red curve shows the running average efficiency $\bar{E}_n(t)$ averaged over a bin of size 10 time steps. The running average here is forward looking, which shifts the averaged trajectory to the left. The vertical green line shows the approximate number of time steps that were used to calculate the cumulative efficiency distributions of Fig. 15. Notice the wide band at the end of several trajectories characteristic of donor photobleaching. Each time step is $500\mu s$.
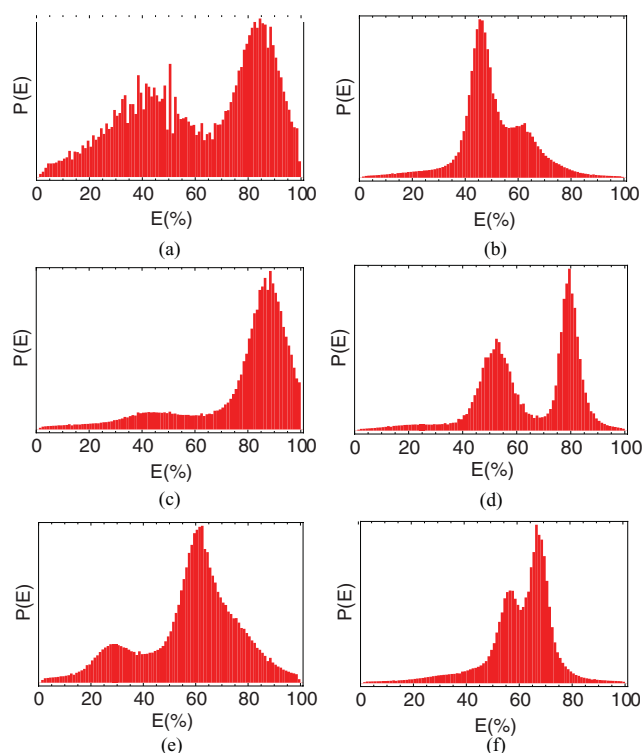
FIG. 15. Non-normalized time-cumulative efficiency distributions $\bar{P}(E)$ derived from the running distributions of Fig. 14. The data have not been corrected for background or cross-talk of donor emission into the red channel.

sophisticated processing, either by the Bayesian distribution approach or by other methods described in the literature.[8–16]

The grand cumulative distribution also supports the observation of multiple states in CaM, which has previously been observed in samples approaching $10^5$ molecules.[28–30] The significance of multiple states of CaM has been discussed elsewhere.[25–27]

## C. The joint $\mu_a$, $\mu_b$ distribution

Bleaching of dyes is a common feature of FRET data and can be treated by many methods. Visual inspection of the time histories, together with the overall rates, indicated that the late-time data $t \gtrsim 150$ units is significantly contaminated by bleaching.
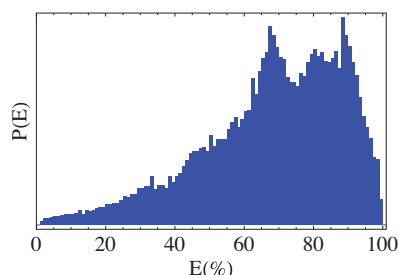


FIG. 16. Non-normalized sum of cumulative distributions of the CaM data over the range $0 < t < 300$ time steps and all the data files. Each time step is $500\mu s$. Parts of the data which were photobleached are taken out.

By correlating this observation with time-histories we were able to make very precise criteria to remove the data from bleaching. Figure 17 summarizes the process using the time-integrated $P(\mu_a, \mu_b)$. The left panel shows the distribution including a significant peak at $\mu_a \to 0$, $\mu_b \to 0$ as well as a few other peaks at high $\mu_a$ and low $\mu_b$ present due to photobleaching. The right panel is the distribution after throwing out the data recorded after the point where $(\mu_a + \mu_b) \to 0$. Careful visual inspection of the raw data coincided with the analytic signature of bleaching to within a few time steps. The numerous peaks seen in $P(\mu_a, \mu_b)$ after bleached areas are removed appear to probe fluorescent states and the inherent dynamics of the molecule. A similar analysis may be useful for identifying different states or environments of the fluorophores. Note that Fig. 17 represents the dynamical history of a single molecule. The studies we conducted found equally interesting effects in almost every molecule, and with great variety. However, further analysis is beyond the scope of the current paper. Dynamic interchange among multiple conformational states may be important for the biological function of the protein.

## VI. CONCLUDING REMARKS

Single molecule data demand methods of analysis going beyond the tradition based on ensemble averages. We have shown that the concept of the *distribution of underlying FRET rates* is powerful and practical. Benchmark comparisons have shown that data-driven distributions of FRET rates are more efficient than some traditional Poisson estimators (i.e., the running average). The time-dependent, running distributions $P(\mu_1, \mu_2; t)$ also capture information about the history of correlations that cannot be expressed with an ensemble language. Time-cumulative distributions $P(\mu_1, \mu_2)$ and $P(E)$ can always be constructed as convenient summaries of single molecule states. Previous single molecule FRET studies of calmodulin have shown multiple states using more traditional analytical methods with large numbers of molecules. The studies here confirm multiple states by using vastly more information from each data run. It appears that the vast information about configurational dynamics contained in single molecule measurements is becoming possible to obtain in an entirely new and unexplored manner.

## APPENDIX A: ANNEALING

Bayesian methods are capable of incorporating and representing information about systems that is more general than textbook rules of "independent" probability. Here we give the details underlying our approach. Let us first show that naive updating with a strict Poisson model *does not* represent an
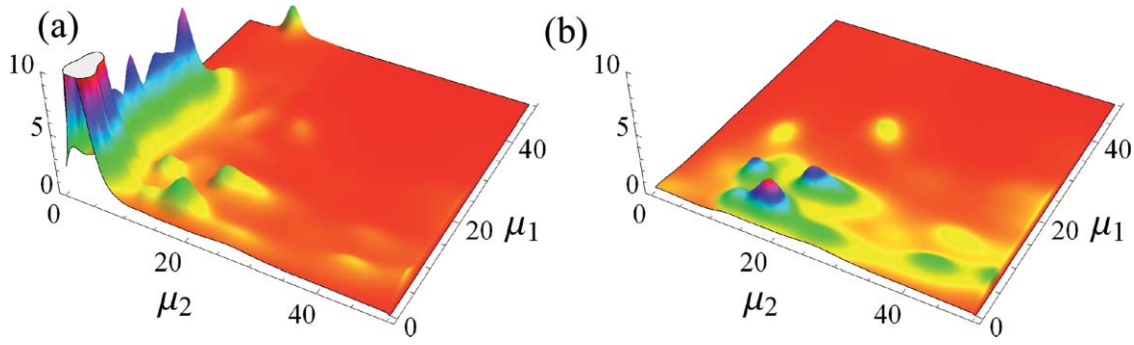
FIG. 17.  Left: The time-integrated distribution $P(\mu_a, \mu_b)$ of a typical molecule that includes a significant period of bleaching. Right: The same study cut off at the time bleaching set in. The numerous peaks seen in $P(\mu_a, \mu_b)$ appear to probe inherent dynamics of the molecule. Overall normalization in both plots is arbitrary.

accurate model of a time-dependent system, unless the system is so slowly varying that the approximation of a delta-function distribution would be good.

Given a data sample $(n_1, n_2, \ldots n_T)$, and a prior distribution $P_0(\mu)$, naive updating with the Poisson distribution produces

$$P(\mu|n_1, n_2, \ldots n_T) = \frac{\mu^{n_1} e^{-\mu}}{n_1!} \frac{\mu^{n_2} e^{-\mu}}{n_2!} \cdots \frac{\mu^{n_T} e^{-\mu}}{n_T!} P_0(\mu),$$

$$= \frac{\mu^{\sum_{i=1}^{T} n_i} e^{-T\mu}}{n_1! n_2! \ldots n_T!} P_0(\mu). \qquad (A1)$$

Let $\bar{n}_T = \sum_i^T n_i / T$ denote the *cumulative* average up to point $T$. Then for sufficiently large $T$, the distribution becomes arbitrarily sharp:

$$P(\mu|n_1, n_2, \ldots n_T) \sim e^{T\bar{n}_T \log(\mu) - T\mu} P_0(\mu)$$

$$\sim e^{-(\mu - \bar{n}_T)^2 T / (2\bar{n}_T^2)} P_0(\mu).$$

The second term is a saddle-point approximation near the distribution's peak at $\mu \sim \bar{n}_T$, and the overall normalization was dropped. The width of the Gaussian is $\bar{n}_T / \sqrt{T}$, and the Gaussian approaches a delta function for $\sqrt{T}/\bar{n}_T \gg 1$. This recapitulates the central limit theorem. With $\bar{n} \sim O(1)$ at good sample resolution, the strict Poisson updater freezes after $T \gg 1$ steps:

$$P(\mu|n_1, n_2, \ldots n_T) \to \delta(\mu - \bar{n}_T).$$

Upon reaching the stage where the probability $P(\mu)$ is completely localized, the weight given to the new data not agreeing with the localized model is driven to zero. Freezing of the distribution is quite difficult for the data to reverse, because it is driven by the *cumulative* history. Given a new data point, which is the transformation $(n_1, n_2, \ldots .n_T) \to (n_1, n_2, \ldots n_T, n_{T+1})$, the effect on the cumulative average is

$$\bar{n}_T \to \frac{T}{T+1}\bar{n}_T + \frac{1}{T+1}n_{T+1}.$$

Note the effect of order $1/T$.

The result is an effective bias *against new information simply because it is new*. Accumulating data for longer and longer periods tends to guarantee no response before an equally long period. The mathematical mishap comes with

the first step, Eq. (A1). It enforces a hidden assumption of *independence* onto the data set $(n_1, n_2, \ldots n_T)$ with a model

$$P(n_1, n_2, \ldots n_T|\mu) \to P(n_1|\mu)P(n_2|\mu) \ldots P(n_T|\mu)$$

Inverting that model literally and including a prior gives Eq. (A1). Independence is a common assumption of ensemble statistics not seeking time-correlated information. *Excess reliance on independence is a mistake in Bayesian terms*, because while we may believe $P(n_1|\mu)$ is nearly Poisson, we do not accept repeating the error of independence literally over a long string of data. Thus naive updating contradicts our information that the distribution of $\mu$ may change, which happens to be the focus of the time dependence being studied.

In order to make a fair updating procedure, our prior beliefs and interest in time dependence must be given an opportunity to enter. Many approaches are possible. One approach might make models of correlation in order to use the information possible in $P(n_1, n_2, \ldots n_T \ldots)$. Another approach might introduce a sharp distribution of a time-dependent parameter $\mu(t)$. While tempting, the notion of $\mu(t)$ is equivalent to a continuously infinite set of parameters $(\mu(t) = \mu_0, \mu_1, \mu_2, \ldots \mu_t)$, which leads to a distribution of an infinite set, namely, functional integrals.

We choose the simplest path. Consider a distribution of $n$ depending on two parameters, $\mu$ and $x$. Parameter $x$, restricted to $0 < x < 1$, describes "transition regions," where the uncertainties of $P(\mu)$ increase. There are many ways to do this. For simplicity let the condition $x \gtrsim x_0$ represent a transition, and $x \lesssim x_0$ be the converse, or steady-state conditions. We will show in a moment that the exact value of $x_0$ and the details defining it are irrelevant. The conditional distribution $P(n \,|\, x, \mu)$ then leads by Bayes to

$$P(x, \mu \,|\, n) = P(n \,|\, x, \mu)P(x, \mu). \qquad (A2)$$

The overall normalization is not written. The "priors" in $P(x, \mu) = P(\mu \,|\, x)P(x)$ describe how $\mu$ and $x$ are correlated. To represent our transition features, we have

$$P(\mu \,|\, x) = P(\mu)\theta(x \leq x_0) + P_{\text{broad}}(\mu)\theta(x > x_0). \qquad (A3)$$

Here $P(\mu)$ is the marginal $\mu$ distribution, updated step-by step, and $P_{\text{broad}}(\mu)$ is a broad distribution in $\mu$ appropriate for transition regions. It will become clear that any

reasonably disjoint functions can substitute for $\theta(x \leq x_0)$ and $\theta(x > x_0)$.

The $n$ distribution is also subdivided. During a transition some broad distribution $P_{\text{broad}}(n \mid \mu)$ is appropriate. Otherwise, the $n$-dependence will be represented by a local model $P_{\text{local}}(n \mid \mu)$ for which the Poisson distribution $P_{\text{local}}(n \mid \mu) \to \mu^n e^{-\mu}/n!$ is a good example. Thus,

$$P(n \mid x, \mu) = P_{\text{local}}(n \mid \mu)\theta(x \leq x_0)$$
$$+ P_{\text{broad}}(n \mid \mu)\theta(x > x_0). \quad \text{(A4)}$$

Combining gives

$$P(x, \mu \mid n) = P(x)(P_{\text{local}}(n \mid \mu)\theta(x \leq x_0)$$
$$+ P_{\text{broad}}(n \mid \mu)\theta(x > x_0))$$
$$\times (P(\mu)\theta(x \leq x_0) + P_{\text{broad}}(\mu)\theta(x > x_0)). \quad \text{(A5)}$$

Experiments don't observe $x$, so we integrate over it:

$$P(\mu \mid n) = \int dx\, P(x, \mu \mid n)$$
$$= (1 - \alpha)P_{\text{local}}(n \mid \mu)P(\mu)$$
$$+ \alpha P_{\text{broad}}(\mu)P_{\text{broad}}(n \mid \mu), \quad \text{(A6)}$$

where

$$\alpha = \int dx\, P(x)\theta(x > x_0); \quad \text{(A7)}$$

$$1 - \alpha = \int dx\, P(x)\theta(x \leq x_0). \quad \text{(A8)}$$

So long as $\alpha \ll 1$, both the derivation and the performance of the updater are not particularly sensitive to the details of $P_{\text{broad}}(\mu)P_{\text{broad}}(n)$, so that numerous models collapse into one parameter. Inserting the Poisson model for $P_{\text{local}}$ and a flat distribution $1/N_{\mu-\text{bins}}$ for $P_{\text{broad}}$ gives the formula used in our analysis,

$$P(\mu \mid n) = (1 - \alpha)\frac{\mu^n e^{-\mu}}{n!}P(\mu) + \frac{\alpha}{N_{\mu-\text{bins}}}. \quad \text{(A9)}$$

## APPENDIX B: EXAMPLE: SINGLE-CHANNEL UPDATING

We present an example calculation of single-channel updating. The example shows how to calculate $P(\mu \mid n(t))$, given $n(t)$ distributed by $P(n \mid \mu) = \mu^n e^{-\mu}/n!$, called "the model."

Consider data $n(t)$ given by

$$n(t) = (8, 3, 4, 10 \ldots). \quad \text{(B1)}$$

The data were generated by drawing two (2) random numbers from the model with $\mu_1 = 3$ followed by two (2) random numbers from the model with $\mu_2 = 7$. For the example, let the discrete $\mu$ values of the distribution range over $0 \leq \mu \leq 10$ in 5 steps $\mu = 1, 3, 5, 7, 9$, representing rather large steps $\Delta\mu = 2$. In practice, 20 or more steps will be preferred. The example uses $\alpha = 0.01$ for definiteness.

To calculate the distribution $P(\mu \mid n(t))$ using the four data points $n(t)$ and conditions given, perform the following steps:

- *Step 0:* Create a blank $5 \times 4$ array $P_{\mu t} = P(\mu \mid n(t))$.
- *Step 1:* Initialize $P_{\mu, 0} = 1/5$ for all $\mu$.
- *Step 2:* Follow the method of the text to calculate

$$P(\mu \mid n(1)) = (1 - \alpha)\frac{\mu^{n(1)}e^{-\mu}}{n(1)!}P(\mu \mid n(0)) + \alpha/5. \quad \text{(B2)}$$

With $\alpha = 0.01$ and $n(1) = 8$ we find the row

$$P(\mu \mid n(1)) = 0.99\left(\frac{1e^{-1}}{40320}, \frac{729e^{-3}}{4480}, \frac{78125e^{-5}}{8064},\right.$$
$$\left.\frac{823543e^{-7}}{5760}, \frac{4782969e^{-9}}{4480}\right)\frac{1}{5}$$
$$+ 0.01\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right).$$

Here $8! = 4480$, $3^8/8! = 729/4480$, etc.

- *Step 3:* Normalize $P_{\mu 1}$ by dividing the entire row $P_{\mu 1}$ by $\sum_{\mu} P_{\mu 1}$:

$$P_{\mu 1} = (0.0261902, 0.0471535, 0.195269, 0.363909,$$
$$0.367479). \quad \text{(B3)}$$

- *Step 5:* Repeat steps 2, 3 for each point $n(t)$ in the data set. At each subsequent step $t$, the distribution $P(\mu \mid n(t))$ will be determined by

$$P(\mu \mid n(2)) = (1 - \alpha)\frac{\mu^{n(2)}e^{-\mu}}{n(2)!}P(\mu \mid n(1)) + \alpha/5,$$

$$P(\mu \mid n(3)) = (1 - \alpha)\frac{\mu^{n(3)}e^{-\mu}}{n(3)!}P(\mu \mid n(2)) + \alpha/5,$$

$$\vdots \quad \text{(B4)}$$

For purposes of the debugging code, we include a list of 20 random numbers from the model with $\mu_1 = 3$ followed by 20 random numbers from the model with $\mu_2 = 7$.

$$n(t) = (5, 0, 3, 1, 5, 5, 6, 1, 3, 5, 2, 7, 4, 5, 2, 2, 2, 1, 5, 3, 9,$$
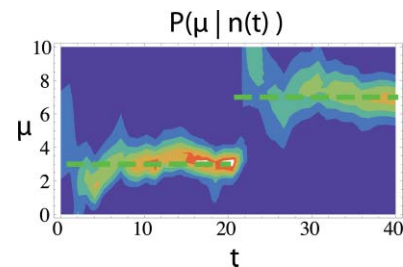$$12, 8, 4, 6, 5, 10, 9, 8, 10, 4, 8, 6, 7, 4, 7, 8, 7, 6, 8).$$
$$\text{(B5)}$$



FIG. 18. Contour plot of the distribution calculated in the example. Horizontal dashed lines represent the mean parameter values.

Let $\mu$ values run from 0.5 to 10 in 20 steps of $\Delta\mu = 0.5$. Then at $t = 10, 20$, the values rounded to two decimal place are:

$$P_{\mu\,10} = (0, 0.01, 0.02, 0.04, 0.12, 0.23, 0.24, 0.15, 0.07,$$
$$0.03, 0.02, 0.01, 0.01, 0.01, 0, 0, 0, 0, 0, 0),$$

$$P_{\mu,\,20} = 0, 0.01, 0.01, 0.03, 0.08, 0.24, 0.32, 0.16, 0.05,$$
$$0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01,$$
$$0.01, 0.01, 0.01). \tag{B6}$$

Note that $t$ is counted from $t = 0$. The resulting contour plot of the distribution $P_{\mu\,t}$ together with the generating parameters $\mu_1$, $\mu_2$ is shown in Fig. 18.

Extension to two channel updating is straightforward. At each point in time $t$, one calculates not an array $P_\mu$ but a matrix $P_{\mu_1,\,\mu_2}$.

[1] W. E. Moerner and M. Orrit, Science **283**, 1670 (1999).

[2] H. Yang and X. S. Xie, J. Chem. Phys. **117**, 10965 (2002).

[3] T. Förster, Ann. Phys. **437**, 55 (1948).

[4] D. C. Lamb, in *Single Particle Tracking and Single Molecule Energy Transfer*, edited by C. Bräuchle, D. C. Lamb, and J. Michaelis (Wiley-VCH, Weinheim, 2010), pp. 99–129.

[5] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss, Proc. Natl. Acad. Sci. U.S.A. **93**, 6264 (1996); A. A. Deniz, M. Dahan, J. R. Grunwell, T. Ha, A. E. Faulhaber, D. S. Chemla, S. Weiss, and P. G. Schultz, *ibid.* **96**, 3670 (1999).

[6] D. S. Talaga, W. L. Lau, H. Roder, J. Tang, Y. Jia, W. F. DeGrado, and R. M. Hochstrasser, Proc. Natl. Acad. Sci. U.S.A. **97**, 13021 (2000).

[7] B. Schuler, E. A. Lipman, and W. A. Eaton, Nature (London) **419**, 743 (2002); E. A. Lipman, B. Schuler, O. Bakajin, and W. A. Eaton, Science **301**, 1233 (2002).

[8] P. J. Rothwell, S. Berger, O. Kensch, S. Felekyan, M. Antonik, B. M. Wörhl, T. Restle, R. S. Goody, and C. A. M. Seidel, Proc. Natl. Acad. Sci. U.S.A. **100**, 1655 (2003); 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, Proc. Natl. Acad. Sci. U.S.A. **100**, 15516 (2003)

[9] L. P. Watkins, H. Chang, and H. Yang, J. Phys. Chem. A **110**, 5191 (2006).

[10] D. S. Talaga, J. Phys. Chem. A **110**, 9743 (2006).

[11] M. Antonik, S. Felekyan, A. Gaiduk, and C. A. M. Seidel, J. Phys. Chem. B **110**, 6970 (2006); S. Kalinin, S. Felekyan, M. Antonik, and C. A. M. Seidel, J. Phys. Chem. B **111**, 10253 (2007).

[12] E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, and S. Weiss, J. Phys. Chem. B **110**, 22103 (2006).

[13] I. Gopich and A. Szabo, J. Chem. Phys. **122**, 014707 (2005); J. Phys. Chem. B **113**, 10965 (2009).

[14] S. A. McKinney, C. Joo, and T. Ha, Biophys. J. **91**, 1941 (2006).

[15] S. C. Kou, X. S. Xie, and J. S. Liu, Appl. Stat. **54**, 469 (2005).

[16] J. B. Witkoskie and J. Cao, J. Chem. Phys. **121**, 6361 (2004); **121**, 6373 (2004); J. Phys. Chem. B **112**, 5988 (2008).

[17] See, e.g., "*Kendall's Advanced Theory of Statistics*" in *Classical Inference and the Linear Model*, edited by A. Stuart, K. Ord, and S. Arnold (Wiley, New York, 2009).

[18] J. P. Ralston 2010 (unpublished).

[19] D. Chin and A. R. Means, Trends Cell Biol. **10**, 322 (2000).

[20] X. S. Xie, J. Chem. Phys. **117**, 11024 (2002).

[21] J. N. Taylor, D. E. Makarov, and C. F. Landes, Biophys. J. **98**(1) 164 (2010).

[22] D. L. Ensign and V. S. Pande, J. Phys. Chem. B**114**(1) 280 (2010).

[23] J. E. Bronson *et al.*, Biophys. J. **97**(12) 3196 (2009).

[24] M. W. Allen, R. J. Bieber-Urbauer, and C. K. Johnson, Anal. Chem. **76**, 3630 (2004).

[25] M. W. Allen, R. J. Bieber-Urbauer, A. Zaidi, T. D. Williams, J. L. Urbauer, and C. K. Johnson, Anal. Biochem. **325**, 273 (2004).

[26] E. Rhoades, E. Gussakovsky, and G. Haran, Proc. Natl. Acad. Sci. U.S.A. **100**, 3197 (2003).

[27] B. Okumus, T. J. Wilson, D. M. Lilley, and T. Ha, Biophys. J. **87**, 2798 (2004).

[28] B. D. Slaughter, M. W. Allen, J. R. Unruh, R. J. Bieber-Urbauer, and C. K. Johnson, J. Phys. Chem. B **108**, 10388 (2004).

[29] B. D. Slaughter, R. J. Bieber-Urbauer, and C. K. Johnson, J. Phys. Chem. B **109**, 12658 (2005).

[30] C. K. Johnson, Biochemistry **45**, 14233 (2006).