

Efficient, nonparametric removal of noise and recovery of probability distributions from time series using nonlinear-correlation functions: Photon and photon-counting noise

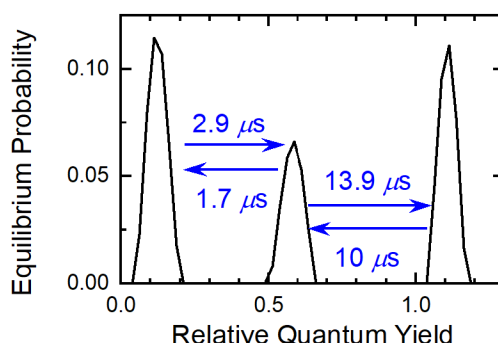
Mainak Dhar and Mark A. Berg^{a)}

Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, USA

^{a)}Author to whom correspondence should be addressed: berg@sc.edu

J. Chem. Phys. **161**(5), 034116 (2024). <http://doi.org/10.1063/5.0212157>

TABLE OF CONTENTS IMAGE



ABSTRACT

A preceding paper [M. Dhar, J. A. Dickinson, and M. A. Berg, J. Chem. Phys. **159**, 054110 (2023)] shows how to remove additive noise from an experimental time series, allowing both the equilibrium distribution of the system and its Green's function to be recovered. The approach is based on nonlinear-correlation functions and is fully nonparametric: no initial model of the system or of the noise is needed. However, single-molecule spectroscopy often produces time series with either photon or photon-counting noise. Unlike additive noise, photon noise is signal-size correlated and quantized. Photon counting adds the potential for bias. This paper extends noise-corrected-correlation methods to these cases and tests them on synthetic datasets. Neither signal-size correlation nor quantization is a significant complication. Analysis of the sampling error yields guidelines for the data quality needed to recover the properties of a system with a given complexity. We show that bias in photon-counting data can be corrected, even at the high count rates needed to optimize the time resolution. Using all these results, we discuss the factors that limit the time resolution of single-molecule spectroscopy and the conditions that would be needed to push measurements into the submicrosecond region.

I. INTRODUCTION

Onsager’s original regression hypothesis postulated that if a system is at equilibrium, all the information about its dynamics is encoded in its thermal fluctuations about that equilibrium.¹ In addition to its importance in theory and simulation, which are noise free, this hypothesis opens up new types of experiment,²⁻⁵ which do contain noise. The most prominent of these equilibrium-fluctuation experiments are fluorescence-correlation^{6,7} (FCS) and single-molecule spectroscopies.⁸⁻¹² They always have high levels of noise, either due to detector imperfections or simply due to the quantization of light detection. Modern instruments can partially compensate by collecting very large datasets, but detection noise remains a limiting factor in many single-molecule experiments.

The more common way to measure dynamics is by measuring a system’s nonequilibrium response to an external perturbation. Although these perturbation–response experiments are theoretically equivalent to equilibrium-fluctuation experiments, the noise removal problem is fundamentally different in the two types of experiment. In perturbation–response experiments, averaging repeated measurements is standard. In equilibrium-fluctuation data, both the signal and the noise are fluctuations; a simple average of repeated experiments removes both equally. However, in either experiment, we can assume that the correlation time of detector noise is fast relative to the rate of change of the signal. This principle underlies the use of fitting functions or smoothing routines for perturbation–response data. The preceding paper (Paper 1¹³) and this one show how to apply the same principle to equilibrium-fluctuation measurements by using nonlinear-correlation functions. The noise-corrected correlation functions can then be used to recover all the system’s properties, both static and dynamic.

This correlation approach offers a number of advantages over other classes of time-series analysis. Compared to parametric methods, correlation analysis avoids the need to know or correctly guess an initial model for the system. Correlation analysis is nonparametric: the data are first reduced to physically relevant system properties, and a model to account for those properties is developed later. Compared to methods that seek an ideal trajectory underlying the noisy data, correlation analysis is more noise tolerant. As a result, it can have improved time resolution. These advantages are shared by the low-order, linear-correlation function, but the addition of high-order functions is required to recover all the system’s properties.

Although Paper 1¹³ showed that a correlation approach works well with additive noise, single-molecule spectroscopy is subject to more complex noise: Poisson noise from linear photon detection or Bernoulli noise from photon counting. Issues of noise–signal correlation, quantization noise, and bias are introduced. This paper establishes that a correlation approach also works well with these types of noise.

This paper also addresses the fundamental limitations imposed by these more complex types of noise. It make useful statements about the highest experimental time resolution for a dataset with a given quality, that is, a given signal-to-noise ratio, time-series length, and number of series. Conversely, it also predicts the signal-to-noise ratio and quantity of data needed to recover the properties of a system of a given complexity. These results then allow a discussion of the current limitations on the time resolution of single-molecule experiments and how the resolution can be improved.

Although Onsager’s regression hypothesis in its broadest form applies to all measures of a system’s dynamics, attention has focused almost entirely on the linear-correlation function. For a time series of noisy data $D(t)$, this function is

$$M_{11}[D](\tau) = \langle D(\tau)D(0) \rangle. \quad (1)$$

It is fundamental to fluorescence-correlation spectroscopy^{6,7} and is often used in single-molecule spectroscopy.^{9,14-17} The effects of detector noise are limited to $\tau = 0$. As a result, the time resolution is limited only by the detector-response time and is higher than in any other analysis method. However, in general, this function alone carries limited information on the system’s properties, and attention has shifted away from linear-correlation analysis of single-molecule data.

However, the linear-correlation function is only the lowest member of a large set of high-order correlation functions,

$$M_{k...lm}[D](\tau_N, \dots, \tau_1) = \langle D(\tau_N + \dots + \tau_1)^k \dots D(\tau_1)^l D(0)^m \rangle. \quad (2)$$

This set contains all the information available in the time series. However, learning which functions provide what information and developing effective methods for extracting that information are ongoing projects. One-dimensional ($N = 1$) correlation functions up to $k + l = 4$ have been used in FCS experiments.^{18,19} The multidimensional members of the set—those functions with more than one time delay τ_i —are important when there are dynamical coordinates that are not distinguished by the observable $D(t)$, that is, when there are hidden coordinates.^{20,21} Both two-dimensional (2D, $N = 2$) and three-dimensional (3D, $N = 3$) correlation functions have been applied to both noise-free, computer simulations^{22,23} and to noisy, single-molecule and FCS data.²⁴⁻³¹ For simplicity, this paper will focus on systems without hidden coordinates and not requiring multidimensional-correlation functions. However, the methods developed here can be extended to these more complex cases (see Paper 1¹³).

Experiments produce data $D(t)$ that consist of an underlying, noise-free signal $S(t)$ with the addition of noise. Paper 1¹³ showed that detector noise creates strong effects in high-order correlation functions that are not present in the linear correlation, so it is not possible to use Eq.

(2) on experimental data without noise correction. However, computer simulations have direct access to a noise-free signal $S(t)$. Looking at them has shown how the high-order correlation functions can be used, once the noise is corrected.^{23,32} In a system without hidden coordinates, it is sufficient to look at the subset of 1D, but nonlinear, correlation functions. Different nonlinear functions of the signal can be used,³³ but here, we focus on powers, yielding the moment-correlation functions,

$$M_{kl}[S](\tau) = \left\langle S(\tau)^k S(0)^l \right\rangle. \quad (3)$$

The overall order of the correlation function is $n = k + l$. These functions can be transformed into the equilibrium-probability distribution $P_{\text{eq}}(S)$ of the signal S , and the 1D Green's function $G(S_1 | S_0; \tau)$, which is the conditional probability of observing S_1 at time $t_1 = t_0 + \tau$, if S_0 occurs at time t_0 . These two quantities fully define the properties of the system: its state space is defined by $P_{\text{eq}}(S)$, and its dynamics within that space are specified by $G(S_1 | S_0; \tau)$. This approach has been used to successfully analyze noise-free time series from computer simulations.^{23,32}

To include experimental data in this approach, Paper 1 introduced a simple noise-correction procedure for high-order correlation functions and detailed its behavior with additive noise.¹³ In this procedure, noise and signal are distinguished by the difference between their correlation times: detector-noise correlation decays rapidly compared to signal correlation. Full recovery of the system's properties from the set of noise-corrected, nonlinear-correlation functions was demonstrated. With an infinite amount of data, the system's properties could be recovered for any signal-to-noise ratio. In reality, property recovery from a finite dataset is limited by sampling error. Relationships describing the required data quality: the signal-to-noise ratio, the length of time series and the number of series; were derived for additive noise. High noise levels can be compensated by collecting more and longer series, until one reaches a "floor" in the signal-to-noise ratio of about one half. The need to bin the time series to reach this floor sets the effective time resolution of the experiment. For high noise levels, it can greatly exceed the time resolution of the detector.

This correlation approach can be contrasted with more established methods of single-molecule analysis.³⁴⁻³⁶ Many of the most popular methods are fully parametric: complete models of the system and the noise are proposed, and the unknown parameters of the model are fit to the data.³⁷⁻⁵¹ Procedures for continuous³⁷⁻³⁹ or discrete state spaces differ. For example, hidden Markov models are confined to discrete-state systems.⁴⁰⁻⁵¹ In contrast, correlation methods are agnostic about whether the states are continuous, discrete, or a mixture of the two. (The method is initially continuous but can be reduced to discrete states later, if the data support that interpretation. See Figs. 6 and 13 below.)

Parametric methods also face the problems of model generation and model selection. If the correct model for the system is not known *a priori*, multiple models must be proposed and fit, and then, the “best” model must be selected. Knowing whether the diversity of proposed models is sufficiently large, deciding which is the best, and determining the uniqueness of the results are all challenges. Bayesian nonparametric methods automate the generation and selection of models using statistically motivated criteria.⁵²⁻⁵⁸ In contrast, correlation methods can be fully nonparametric, and modelling can be postponed until after the data are analyzed.

Another set of existing methods are based on explicitly extracting the noise-free signal series $S(t)$ —often called the ideal trajectory—from the noisy data series $D(t)$. With the ideal trajectory in hand, analysis is easier. The 1D histogram of $S(t)$ gives $P_{\text{eq}}(S)$,^{48,59-62} and a 2D histogram gives $G(S_1 | S_0; \tau)$. In the case of discrete-state systems, the time series can be reduced to a sequence of derived quantities: dwell times, transitions times, or more general events.⁶³⁻⁶⁵ Parametric modeling or nonparametric statistics can then be applied to these derived quantities, rather than to raw intensities, as in the correlation approach. For example, dwell-time histograms are analogous to the linear intensity correlations [Eq. (1)].^{14,66} High-order statistics of derived quantities, such as 2D dwell-time histograms,^{5,67} constrained dwell-time correlation functions^{68,69} or event-averaged correlation functions,⁷⁰⁻⁷³ have the same goals as the high-order intensity correlations defined in Eq. (2).

However, extracting the ideal trajectory from noisy data is problematic in practice. The simplest and most general method of noise reduction is binning in time along the time series.⁷⁴⁻⁷⁶ The method requires no model, but dramatically reduces the time resolution when the signal-to-noise ratio is low. Improvements are possible by restricting the solution space to certain classes of model. For example, if solutions are confined to a small number of states undergoing sudden transitions, thresholding,^{77,78} filtering,⁷⁹ change-point detection,⁸⁰⁻⁸⁵ or machine learning⁸⁶⁻⁹⁰ can yield an ideal trajectory. Determining the correct number of states remains challenging.

In addition, the ideal trajectory $S(t)$ is problematic in concept. Knowledge of the state of a specific molecule at a specific time cannot be improved by collecting more data at other times or from other molecules: in practice, it is not a statistically consistent quantity. Insisting that $S(t)$ be known accurately places unnecessary demands on the signal-to-noise ratio. However, the ideal trajectory is never the ultimate goal of an analysis; only its statistical properties are important. Correlation methods bypass any explicit determination of the ideal trajectory, allowing noisier data to be used and improving the time resolution.

Another idea currently popular in single-molecule analysis is forgoing an ensemble average over multiple molecules, for example, by calculating single-molecule correlation functions or single-molecule diffusion constants. The limited amount of data available for one molecule can leave large sampling error in the results.⁹¹⁻⁹³ However, any system that has reached equilibrium

or a stable steady state by the start of the experiment is ergodic. An average over both time and molecules is possible and is implied by the angular brackets $\langle \dots \rangle$ in our correlation functions. Similarly, parametric methods often pool the results from many molecules.⁵¹ However, some investigators have argued that ensemble averaging is antithetical to single-molecule experiments.⁹⁴⁻⁹⁹ Many of their concerns are addressed by recognizing that (ensemble-averaged) multidimensional-correlation functions contain information about heterogeneity within the ensemble.^{20-22,27-29,31} The ability to remove noise by correlation methods without seeking an ideal trajectory also alleviates concerns about the loss of information when multiple times series are directly averaged. The correlation methods described here allow free ensemble averaging. As a result, the measurement of many molecules can compensate for the finite lifetime of any one molecule.

Paper 1¹³ demonstrated the potential of correlation analysis when the noise is additive. However, single-molecule spectroscopy is subject to photon noise, which is more complex, and those complexities are addressed in this paper. At high signal levels, the noise is Gaussian, but the width of the Gaussian depends on the signal level: there is noise–signal correlation. At low signal levels, the noise also becomes non-Gaussian and quantized. In photon counting, the data become biased at high count rates. Section II describes a model for each of these types of noise and a model system with which to test the proposed methods.

Section III examines photon noise. It first (Sec. III.A) uses simulated data to demonstrate that neither noise–signal correlation nor non-Gaussian noise nor quantization significantly degrades correlation analysis relative to additive noise. It then (Sec. III.B) analyses the data quality needed for a successful property recovery with photon noise. It shows that a floor to the signal level exists at a count rate of about $\frac{1}{2}$ count per time bin, analogous to what was seen for additive noise in Paper 1.¹³

Single-molecule data are often collected by photon counting, leading to binary (Bernoulli) signals. For linear-correlation functions, no change in methodology is needed, and the experimental time resolution can match the detector time resolution.^{9,15-17} Parametric methods have also been developed specifically for this “photon-by-photon” data, although the time resolution is not as high.^{24,38,40,42,43,47,55-58,100} Section IV looks at this problem for nonlinear-correlation functions. In Sec. IV.A, we identify the amount of time binning needed and the time resolution achievable in the typical case of low count rates and low bias.

However, general concepts argue that a time series will contain the most information at high count rates and high bias. It is difficult to directly correct for bias in photon-by-photon data, but Sec. IV.B shows that it is easy to correct the results of a correlation analysis. This ability opens up the analyzing data with high signal intensities and, therefore, high time resolution.

The fundamental limits to the time resolution in single-molecule spectroscopy remains an important issue.¹⁰¹ In many single-molecule experiments, the time resolution is limited to $\gtrsim 1$ ms. However, the time resolution of photon-counting detectors can be orders of magnitude lower, 25 ns or less.¹⁰²⁻¹⁰⁷ One goal of improved noise removal is to reduce the effective time resolution. Section V discusses this issue in light of the results of this paper. As an illustration, it presents a scenario in which the time resolution is submicrosecond and discusses the conditions needed to extend the current limits on time resolution.

II. MODELING DETECTOR NOISE IN EQUILIBRIUM-FLUCTUATION MEASUREMENTS

We assume that a molecule or other emitter is at equilibrium or in a stable steady state. Its quantum yield $X(t)$ is the observable of interest, and it varies stochastically with time t . The observable's statistics are defined by its joint-probability distribution $P(X_1, X_0; \tau)$, the probability of observing X_0 at time t_0 and then X_1 at time $t_1 = t_0 + \tau$. This quantity can be split into two parts,

$$P(X_1, X_0; \tau) = G(X_1 | X_0; \tau) P_{\text{eq}}(X_0), \quad (4)$$

with the static properties defined by the equilibrium distribution $P_{\text{eq}}(X_0)$, and the dynamics defined by the conditional probability or Green's function $G(X_1 | X_0; \tau)$. In our examples, these dynamics will be Markovian (that is, no coordinates other than the quantum yield affect the dynamics), but Markovian dynamics are not required. Similarly, discrete-state examples of $P_{\text{eq}}(X_0)$ will be presented, but continuous distributions can be treated in the same manner.

The molecule is excited with a constant light flux I_{ex} . The emitted photons are collected with an experimental sensitivity,

$$S^0 = \Omega \frac{I_{\text{ex}} \sigma_{\text{abs}}}{\hbar \omega_{\text{ex}}}, \quad (5)$$

where the fraction of emitted photons collected and detected is Ω , the absorption cross-section is σ_{abs} , and the excitation frequency is ω_{ex} . A noise-free detector would give a signal

$$S(t) = S^0 (X(t) + B). \quad (6)$$

with a background contribution B from the sample.

A real detector generates data $D(t_j)$ that include noise

$$D(t_j) \sim \text{Noise}(D | S(t_j)), \quad (7)$$

where $\sim$ indicates a random selection from the conditional-probability distribution $\text{Noise}(D | S)$. The data are generated at discrete time points t_j , which are separated by the time resolution of the

detector, $T_\varepsilon = t_{j+1} - t_j$. This time includes any dead time and temporal cross-talk in the detector. It must be long enough that the noise is uncorrelated in time, that is

$$E\left(D(t+T_\varepsilon)^m D(t)^n\right) = E\left(D(t+T_\varepsilon)^m\right)E\left(D(t)^n\right), \quad (8)$$

where $E(\dots)$ is the expectation value over many hypothetical datasets. In addition, we assume that the signal correlation decays slowly relative to the noise correlation, that is,

$$E\left(S(nT_\varepsilon)S(0)\right) \approx E\left(S^2\right) \quad (9)$$

for n not too large. This “fast-noise/slow-signal” assumption is the primary basis for noise removal.

The noise is characterized by its first two moments. If the detector is linear, the noise is unbiased, and

$$E(D | S) = S, \quad (10)$$

or using the law of total expectation,

$$E(D) = E(S). \quad (11)$$

In this case, the magnitude of the noise σ_ε can be measured by its variance for a given signal level, averaged over all the signal levels in the data,

$$\begin{aligned} \sigma_\varepsilon^2 &= E(\text{Var}(D | S)) \\ &= \text{Var}(D) - \text{Var}(S), \end{aligned} \quad (12)$$

where $\text{Var}(\dots)$ is the variance over many hypothetical datasets. The last line follows from the law of total variance. It is more useful to measure the noise’s magnitude relative to the average signal level, giving the signal-to-noise ratio,

$$\text{SNR} = \frac{\langle D \rangle}{\sigma_\varepsilon}. \quad (13)$$

A simple and common method to decrease the noise is to average the data over N_b time bins,

$$D(t_j, N_b) = \frac{1}{N_b} \sum_{i=j}^{j+N_b} D(t_i). \quad (14)$$

The experimental noise is reduced to

$$\sigma_\varepsilon(N_b) = N_b^{1/2} \sigma_\varepsilon. \quad (15)$$

Both $S(t)$ and $D(t)$ are defined to have units of photons/s, so their magnitudes are independent of time binning. The signal-to-noise ratio increases with binning, but the experimental time resolution T_r is reduced, $T_r = N_b T_\varepsilon$.

In general, knowing the noise distribution requires detailed characterization of a specific detector.¹⁰²⁻¹⁰⁷ Our preceding paper considered additive noise,¹³ for which

$$\text{Noise}(D | S) = P_\varepsilon(D - S). \quad (16)$$

where $P_\varepsilon(\varepsilon)$ is an unconditioned noise distribution. Additive noise is typical of electronic amplifiers. It has three properties that make it simple. First, additive noise is unbiased [Eq. (10)]. Second, the signal dynamics and the noise generation are independent processes. In particular, the variance of the data is independent of the signal size,

$$\text{Var}(D(t, N_b | S)) = N_b \sigma_\varepsilon^2. \quad (17)$$

The signal-to-noise ratio simplifies to

$$\text{SNR} = \frac{\sqrt{N_b} \langle D \rangle}{\sigma_\varepsilon}. \quad (18)$$

Lastly, additive noise is effectively continuous. Although real detectors digitize the data to discrete values, this effect is minor compared to σ_ε .

In general, the additive-noise distribution can have a variety of forms. For the examples presented here, Gaussian noise will be used,

$$\text{Gaussian}(X) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left(-\frac{X^2}{2\sigma_\varepsilon^2}\right) \quad (19)$$

and

$$\text{Noise}(D | S) = \text{Gaussian}(D - S) \quad (20)$$

An equivalent definition of additive noise is that

$$D(t_j) = S(t_j) + \varepsilon(t_j), \quad (21)$$

where $\varepsilon(t_j)$ is a second stochastic process with a distribution $P_\varepsilon(\varepsilon)$,

$$\varepsilon(t_j) \sim P_\varepsilon(\varepsilon = D - S), \quad (22)$$

which is independent of the signal. The lack of bias [Eq. (11)] implies that $P_\varepsilon(\varepsilon)$ has a zero mean. The independence of signal and noise processes implies that

$$E\left(\varepsilon(\tau)^k S(0)^l\right) = E(\varepsilon^k) E(S^l). \quad (23)$$

In particular, the noise magnitude σ_ε is the standard deviation of $P_\varepsilon(\varepsilon)$ [compare to Eq. (17)]. Although Eqs. (21) and (22) are more familiar ways to describe additive noise, Eqs. (7) and (16) are easier to generalize.

This paper will look at noise with more complex features: signal-correlated noise, quantized noise, and biased noise. Light detection is subject not only to noise from detector imperfections, but also to noise from the quantized nature of light. Photon noise occurs when this latter source is important. If the average number of collected photons in a time bin is $\bar{N}_{\text{ct}}$, the number of photons actually detected N_{ct} follows a Poisson distribution,¹⁰⁸

$$\text{Poisson}(N_{\text{ct}} | \bar{N}_{\text{ct}}) = e^{-\bar{N}_{\text{ct}}} \frac{\bar{N}_{\text{ct}}^{N_{\text{ct}}}}{N_{\text{ct}}!}; \quad N_{\text{ct}} = 0, 1, \dots \quad (24)$$

For photon noise, there are two, alternative descriptions of the data and signal, either as intensities $D(t_j)$ and $S(t_j)$, which have units of photons per second and binning-independent magnitudes, or as count levels $N_{\text{ct}}(t_j)$ and $\bar{N}_{\text{ct}}(t_j)$, which are unitless (photons per bin) and have binning-dependent magnitudes. These two descriptions are related by

$$\begin{aligned} N_{\text{ct}}(t_j) &= N_b T_\varepsilon D(t_j) \\ \bar{N}_{\text{ct}}(t_j) &= N_b T_\varepsilon S(t_j) . \end{aligned} \quad (25)$$

In terms of intensities, the noise is defined by

$$\text{Noise}(D | S) = \text{Poisson}(N_b T_\varepsilon D | N_b T_\varepsilon S) . \quad (26)$$

A special property of Poisson noise is that the sum of Poisson variables is also a Poisson variable. Upon binning, the shape of the distribution changes, but the form does not.¹⁰⁸ As a result, N_b does not appear as an independent variable in Eq. (26) [contrast Eq. (33) below].

Similar to additive noise, photon noise is unbiased [Eq. (10)]. Unlike additive noise, photon noise is correlated to the signal level. However, the correlation is simple to describe,

$$\text{Var}(D(t, N_b | S)) = \frac{S}{N_b T_\varepsilon} . \quad (27)$$

[compare to Eq. (17)]. These two properties make it possible to calculate the average noise magnitude, $\sigma_\varepsilon = (\langle D \rangle / N_b T_\varepsilon)^{1/2}$, and the average signal-to-noise ratio,

$$\text{SNR} = \sqrt{N_b T_\varepsilon \langle D \rangle} = \langle N_{\text{ct}} \rangle^{1/2} . \quad (28)$$

[compare to Eq. (18)]. Because the signal-to-noise ratio is well defined for both additive and photon noise, it is the best basis for comparing them.

Photon noise also differs from additive noise in that it is quantized: N_{ct} is limited to integers [Eq. (24)]. However, the effects of noise–signal correlation and quantization are distinguishable. The effects of quantization disappear for $\langle N_{\text{ct}} \rangle \gg 1$ or $\text{SNR} \gg 1$, whereas the effects of correlation remain at all signal levels.

In the linear photon detector just discussed, there is no upper limit to the number of photons that can be detected in one time bin. Current-detected photomultipliers, photodiodes, and photodiode arrays approximate this ideal. However, the highest sensitivity and time resolution and the lowest background are achieved by photon-counting detectors, such as avalanche photodiodes. In photon counting, any number of photons within the detector-time bin creates the same result. This behavior is idealized as Bernoulli noise,¹⁰⁸

$$\text{Bernoulli}(N_{\text{ct}} | \bar{N}_{\text{ct}}^{\text{o}}) = \begin{cases} e^{-\bar{N}_{\text{ct}}^{\text{o}}}; & N_{\text{ct}} = 0 \\ 1 - e^{-\bar{N}_{\text{ct}}^{\text{o}}}; & N_{\text{ct}} = 1 \end{cases}, \quad (29)$$

in which only the absence ($N_{\text{ct}} = 0$) or the presence of one or more photons ($N_{\text{ct}} = 1$) is registered. The true distribution of photons is assumed to be Poisson, and the true, average number of photons in an detector-time bin is $\bar{N}_{\text{ct}}^{\text{o}}$. The unbinned data are generated by

$$\text{Noise}(D | S) = \text{Bernoulli}(T_{\varepsilon} D | T_{\varepsilon} S(t_j)). \quad (30)$$

As with photon noise, it is correlated and quantized. However, the data are now biased,

$$E(D | S) = \frac{1}{T_{\varepsilon}} (1 - e^{-T_{\varepsilon} S}) \quad (31)$$

[compare to Eq. (10)]. As a result, neither σ_{ε} nor SNR are useful measures of the noise magnitude. The unbinned count level $\bar{N}_{\text{ct}}^{\text{o}}$ is more important.

Unlike additive or photon noise, the statistics of photon counting change upon time binning. The Bernoulli distribution becomes a binomial distribution,

$$\text{Binomial}(N_{\text{ct}} | \bar{N}_{\text{ct}}^{\text{o}}, N_b) = \binom{N_b}{N_{\text{ct}}} e^{-\bar{N}_{\text{ct}}^{\text{o}}(N_b - N_{\text{ct}})} \left(1 - e^{-\bar{N}_{\text{ct}}^{\text{o}}}\right)^{N_{\text{ct}}}. \quad (32)$$

$N_{\text{ct}} = 0, 1, \dots, N_b$

In terms of intensities,

$$\text{Noise}(D | S, N_b) = \text{Binomial}(N_b T_{\varepsilon} D | T_{\varepsilon} S(t_j), N_b). \quad (33)$$

The detector-time counts $\bar{N}_{\text{ct}}^{\text{o}}$ and the extent of binning N_b are independently relevant, not just their product [contrast with Eq. (24)].

One way to avoid the complications of photon-counting noise is to operate in the quasi-Poisson limit: the detector-count rate must be low $\bar{N}_{\text{ct}}^0 = T_{\text{e}}S \ll 1$, and the time binning must be extensive, $\bar{N}_{\text{ct}} = N_b \bar{N}_{\text{ct}}^0 \gtrsim 1$. In this limit, the noise approaches a Poisson distribution,

$$\lim_{\substack{\bar{N}_{\text{ct}}^0 \rightarrow 0 \\ \bar{N}_{\text{ct}}^0 N_b = \text{const}}} \text{Binomial}(N_{\text{ct}} | \bar{N}_{\text{ct}}^0, N_b) = \text{Poisson}(N_{\text{ct}} | N_b \bar{N}_{\text{ct}}^0). \quad (34)$$

The noise becomes unbiased, it can be described by a signal-to-noise ratio [Eq. (28)], and the detector is linear. These desirable properties come at the expense of an experimental time resolution much greater than the detector time resolution.

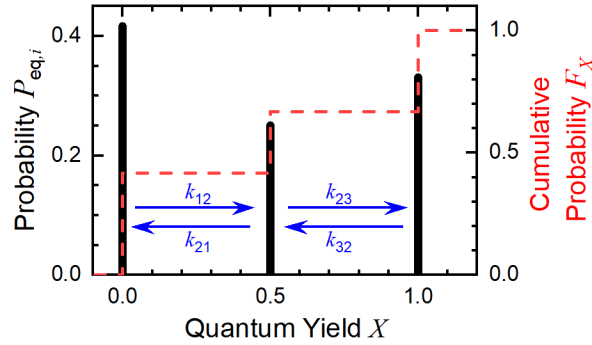


FIG. 1. Synthetic data were generated from this model system. The equilibrium probabilities of the states versus their quantum yields are shown as black bars (left scale), along with their state-to-state transition rates (blue arrows). The cumulative-distribution function is shown as a dashed, red curve (right scale).

To compare these different types of noise, ideal-signal time series $S(t)$ were generated from a model system, and then different types of noise were added to them to create data time series $D(t)$. The model is the same one used in our preceding paper on additive noise,¹³ except the observed property X is now explicitly a quantum yield. The model (Fig. 1) consists of three, discrete states with equilibrium probabilities $\mathbf{P}_{\text{eq}} = \{5/12, 1/4, 1/3\}$ and corresponding quantum yields $\mathbf{X} = \{0, 0.5, 1.0\}$. The cumulative probability,

$$F_X(X) = \int_0^X P_X(X') dX' , \quad (35)$$

is also shown.

The kinetics are ultimately defined by the state-to-state rates k_{ij} . However, with the population ratios already set, we only need to define the equilibration times, $(k_{ij} + k_{ji})^{-1}$. Being a simulation, all the rates can be uniformly scaled to any relevant time scale. However, we will quote specific times to help visualize the results. The low- and medium-quantum-yield states exchange rapidly

with a fast-equilibration time of $T_f = (k_{12} + k_{21})^{-1} = 1$ ms. The medium- and high-quantum-yield states exchange with a slow-equilibration time $T_{\text{er}} = (k_{23} + k_{32})^{-1} = 100$ ms. As the slowest relaxation in the system, it is also the ergodic time. There are no direct transitions between the low- and high-quantum-yield states, $(k_{13} + k_{31})^{-1} = 0$. In the initial simulations, the detector time resolution will be $T_\varepsilon = 10 \mu\text{s}$.

The length of a single series will be T_L . The correlation functions use an ergodic average $\langle \dots \rangle$, one that allows both averaging over the T_L / T_ε time points from one molecule and ensemble averaging over multiple series from N_{en} different molecules. The overall dataset size is $N_{\text{er}} = N_{\text{en}} T_L / T_\varepsilon$ data points.

The initial simulations will have one series ($N_{\text{en}} = 1$ molecule) and a length of $T_L = 1000$ s, resulting in $N_{\text{er}} = 10^8$ data points. This large size has the pedagogic advantage that we can see the effects of clearly separated fast and slow dynamics. At the same time, the effects of detector noise dominate, and the errors in sampling the slow dynamics and in resolving the fast dynamics remaining secondary. An example of a smaller dataset will be given in Sec. V.A.

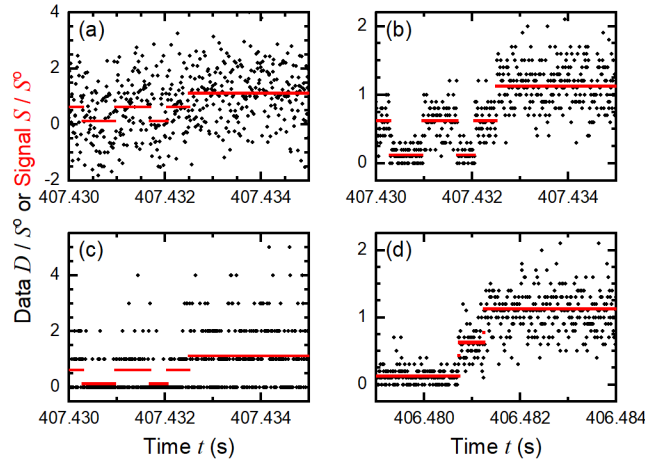


FIG. 2. Examples of time series with different types of noise: noisy data (black points) and the underlying, noise-free signal (red points). All series have a time bin of $10 \mu\text{s}$. Left column: Medium-intensity signal, $S^0 = 0.1 \mu\text{s}^{-1}$. Right column: High-intensity signal, $S^0 = 1 \mu\text{s}^{-1}$. (a) Medium-intensity signal with Gaussian-additive noise ($\sigma_\varepsilon = S^0$, $\text{SNR} = 0.58$, from Paper 1¹³). (b) High-intensity signal with photon noise ($\bar{N}_{\text{ct}} = 5.83$, $\text{SNR} = 2.41$). (c) Medium-intensity signal with photon noise ($\bar{N}_{\text{ct}} = 0.583$, $\text{SNR} = 0.76$). (d) High-intensity signal with photon-counting noise ($\bar{N}_{\text{ct}}^0 = 0.0583$ binned by $N_b = 100$, $\text{SNR} \approx 2.4$). See Table S1 in the supplementary material for details of the datasets.

Figure 2 shows samples of data with different types of noise. In each case, the noise-free signal is shown in red, and the noisy data are shown in black. [Figs. 2(a)–2(c) use the same signal

time series; Fig. 2(d) uses a separately generated series.] The data in Fig. 2(a) have Gaussian-additive noise with a low signal-to-noise ratio, $\text{SNR} = 0.583$ ($\sigma_\varepsilon = S^0 = 0.1 \mu\text{s}^{-1}$). These data were extensively analyzed in Paper 1¹³ and serve as a reference.

Figure 2(b) shows the effects of correlated noise. The signal was increased 10-fold ($S^0 = 1 \mu\text{s}^{-1}$ or $\langle S \rangle = 583 \text{ count/ms}$) to reduce quantization, and photon noise [Eq. (26)] was added. The noise for the high-quantum-yield states is noticeably higher than for the low-quantum-yield states. Figure 2(c) adds quantization effects. As in Fig. 2(b), the data has photon noise, but the signal level has been dropped ($S^0 = 0.1 \mu\text{s}^{-1}$ or $\langle S \rangle = 58.3 \text{ count/ms}$), until the signal-to-noise ratio ($\text{SNR} = 0.764$) is comparable to the additive example in Fig. 2(a). The quantization gaps are now comparable to the overall spread due to noise. The noise distribution also develops noticeable skewing toward high counts, especially when the signal is low.

Figure 2(d) shows an example of photon-counting noise just outside the quasi-Poisson limit. Data from a photon-counting detector with a time resolution of $T_\varepsilon = 0.1 \mu\text{s}$ has been binned to $10 \mu\text{s}$ ($N_b = 100$). A high signal level, the same as in Fig. 2(b) ($S^0 = 1 \mu\text{s}^{-1}$ or $\langle S \rangle = 583 \text{ count/ms}$), creates bias, in addition to noise-signal correlation and noise quantization. The small bias is discernable as a slight deficiency in the number of high values while in the high-quantum-yield state.

III. CORRECTING PHOTON NOISE

A. Recovering the system's properties

In the absence of detector noise, the properties of a system are easily extracted from a time series of its fluctuations at equilibrium $S(t)$ using histograms. The equilibrium distribution is well approximated by a 1D histogram of the series, and the Green's function can be approximated by a 2D histograms of values separated by specific times. We start with an alternative approach that is initially more complex. However, it will become more useful when noise is introduced.

We define a set of moment-correlation functions,³³

$$M_{kl}[S](\tau) = \left\langle S(\tau)^k S(0)^l \right\rangle, \quad (36)$$

which are written as functionals of the time series $S(t)$. They can be regarded as raw, bivariate moments of $P(S_1, S_0; \tau)$.¹⁰⁹ The univariate moments of $P_{\text{eq}}(S_0)$,

$$\mu_n[S] = \left\langle S^n \right\rangle = M_{kl}[S](0); \quad n = k + l, \quad (37)$$

can be measured as the zero-delay points of the moment-correlation functions of the same overall order n . These moments are sufficient to find $P_{\text{eq}}(S_0)$ (Paper 1¹³). After this distribution is found, the moment-correlation functions [Eq. (36)] are sufficient to find $G(S_1 | S_0; \tau)$.³³

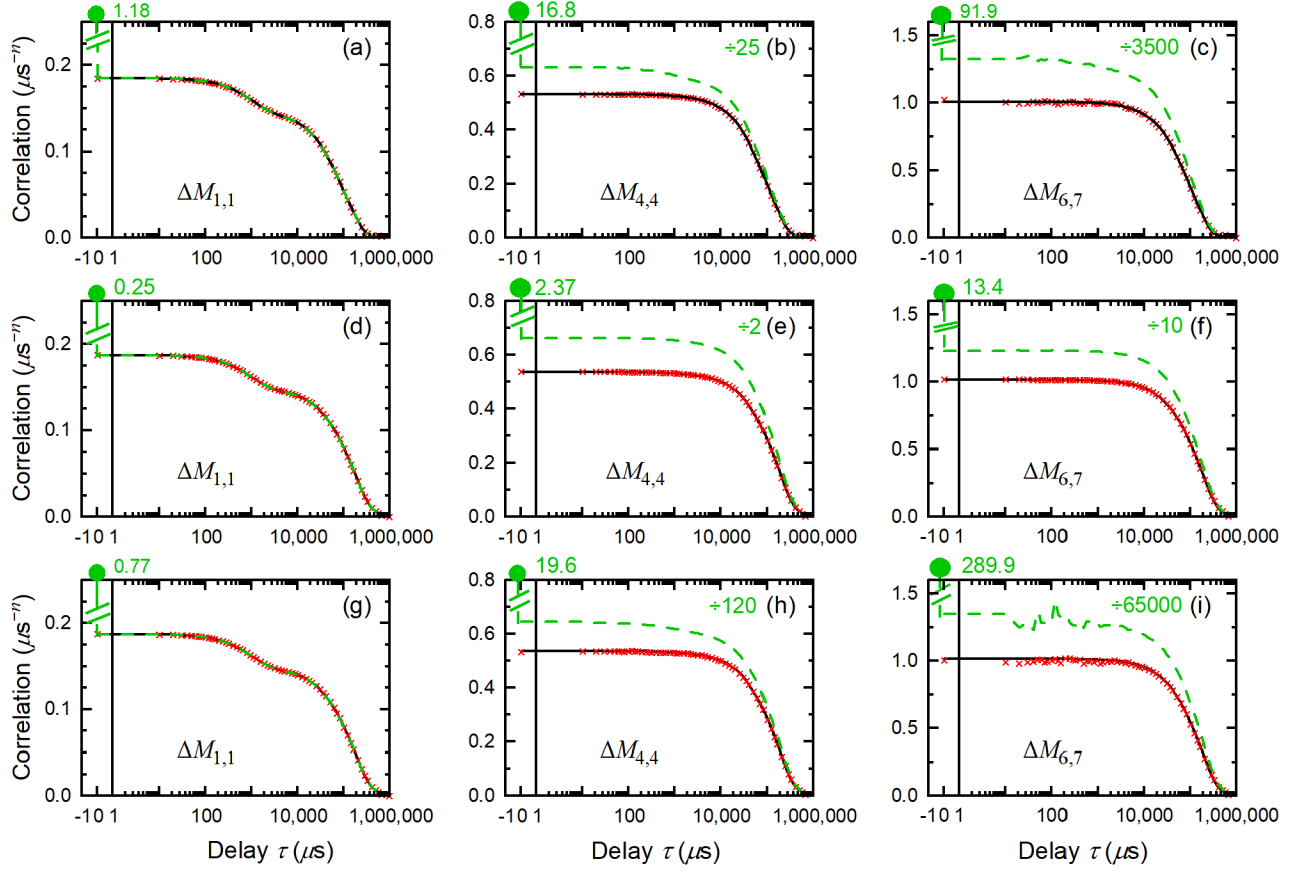


FIG. 3. Effect of noise and noise correction on moment-correlation functions. Top row: Medium-intensity signal and Gaussian-additive noise. Middle row: High-intensity signal and photon noise. Bottom row: Medium-intensity signal and photon noise. Black, solid curves: from noise-free signal $\Delta M_{kl}[S](\tau)$. Green, dashed curves: from noisy data without noise correction $\Delta M_{kl}[D](\tau)$. They have been reduced by the factors given within the figures to bring them onto the vertical scale. Their time-zero points (green circles) are still off-scale. The circles are labeled with their vertical value (after division). Red crosses: from noisy data with noise correction $\Delta M'_{kl}[D](\tau)$.

Figure 3 shows examples of noise-free correlation functions calculated from $S(t)$ (black curves, identical for all three rows). Because these functions do not decay to zero, we plot $\Delta M_{kl}(\tau) = M_{kl}(\tau) - M_{kl}(\infty)$. The lowest non-trivial, moment-correlation function $M_{1,1}[S](\tau)$ is the standard, linear-correlation function and is shown in the left-hand column. By itself, it does not have enough information to recover the Green's function. Higher orders add the required

information. The number of orders needed for a successful recovery increases with the complexity of the system (Paper 1¹³). Orders up to those of the center column, $n = 8$, are sufficient for a system of three discrete states. With four discrete states, orders at least up to those of the right-hand column, $n = 13$, are needed.

In the presence of strong noise, neither histogramming nor uncorrected moment-correlation functions work well. The green, dashed curves in Fig. 3 show the effects of strong Gaussian noise [top row of Fig. 3, also see Fig. 2(a)], moderate photon noise [middle row of Fig. 3, also see Fig. 2(b)] and strong photon noise [bottom row of Fig. 3, also see Fig. 2(c)] on the moment-correlation functions. Because detector noise is uncorrelated between different time points, the noise only affects the $\tau = 0$ point of the linear-correlation function, regardless of the nature of the noise (Fig. 3, left-hand column). As a result, the linear-correlation function has very good time resolution, even with noisy data.^{9,14-17} However, noise more strongly affects higher-order correlation functions, so much so that the green, dashed curves must be rescaled to remain on the plots in the center and right-hand columns of Fig. 3. Histogramming these data fails just as dramatically (Paper 1¹³). Some type of noise correction is needed to recover the system's properties.

Both the signal and the noise are stochastic processes. The fundamental feature that distinguishes between them is their correlation time; the noise has a short correlation time, and the signal has a long one. To exploit this difference, primed powers of the data are defined by using values separated by the noise-correlation time T_ϵ ,

$$D(t)^{k'} = D((k-1)T_\epsilon + t) \dots D(T_\epsilon + t) D(t). \quad (38)$$

The primed powers are then used to define the noise-corrected, moment-correlation functions,

$$M'_{kl}[D](\tau) = \left\langle D(\tau + lT_\epsilon)^{k'} D(0)^{l'} \right\rangle \approx M_{kl}[S](\tau), \quad (39)$$

and moments,

$$\mu'_n[D] = \left\langle D^{n'} \right\rangle \approx \mu_n[S]. \quad (40)$$

These definitions extend the noise rejection of the linear-correlation function to the nonlinear case. If the fastest signal-correlation time is still long compared to nT_ϵ , the noise-corrected results from the data will closely approximate those from the noise-free signal. These values can then be used to recover both the equilibrium distribution and Green's function.

The practicality of this approach was demonstrated for additive noise in Paper 1¹³ and is summarized by the top row of Fig. 3, which shows noise-corrected correlation functions for Gaussian-additive noise (red crosses). The agreement with the noise-free (black curves) result is

excellent, even at high orders. The accuracy of the noise-corrected moments can be judged from the $\tau = 0$ points.

The current question is whether the performance of this method degrades for photon noise. The middle row of Fig. 3 shows high-signal, photon-noise data, which is affected by noise–signal correlation [see Fig. 2(b)]. The bottom row shows low-signal, photon-noise data [see Fig. 2(c)], which is also affected by quantization. In both cases, Eq. (39) continues to accurately correct noise effects of many orders of magnitude, just as it did for additive noise.

However, close examination of Fig. 3 shows that the noise-correction is not perfect, especially at the highest orders. There is residual error due to imperfect sampling, despite the large amount of data used. Comparing the Gaussian–additive and the low-signal, photon-noise cases [Fig. 3(c) and 3(i)], which have similar signal-to-noise ratios, the photon noise leaves somewhat higher sampling error. The cause of this effect will be discussed below, but we first check to see whether it creates a problem in recovering the system’s properties.

We first focus on recovering the equilibrium distribution and briefly review the methods that were developed in Paper 1.¹³ A sequence of noise-corrected moments, $\boldsymbol{\mu}'[D] = \{\mu'_n[D]\}$ for $n = 1$ to N_μ , are calculated from the noisy data using Eqs. (38) and (40). A vector of N_P evenly spaced points is taken along the signal axis, $\mathbf{S} = \{S_i\}$ from $S_1 = S_{\min}$ to $S_{N_P} = S_{\max}$. The recovered equilibrium distribution is given by the vector of values at these points, $\mathbf{P}_r(\mathbf{S}) = \{P_r(S_i)\}$. These quantities satisfy the equation,

$$\boldsymbol{\mu}'[D] = \mathbf{V}(\mathbf{S}) \cdot \mathbf{P}_r(\mathbf{S}), \quad (41)$$

where the Vandermonde matrix $\mathbf{V}(\mathbf{S})$ is defined by¹¹⁰

$$\mathbf{V}(\mathbf{S}) = \begin{pmatrix} \mathbf{S} \\ \mathbf{S}^2 \\ \vdots \\ \mathbf{S}^{N_\mu} \end{pmatrix} = \begin{pmatrix} S_1 & S_2 & \dots & S_{N_P} \\ S_1^2 & S_2^2 & \dots & S_{N_P}^2 \\ \vdots & \vdots & \ddots & \vdots \\ S_1^{N_\mu} & S_2^{N_\mu} & \dots & S_{N_P}^{N_\mu} \end{pmatrix}. \quad (42)$$

The problem is to invert Eq. (41) to find $\mathbf{P}_r(\mathbf{S})$ from $\boldsymbol{\mu}'[D]$. As with many inverse problems, this one is underdetermined and ill-conditioned.¹¹¹ As a result, an infinite number of very different functions solve the problem exactly.

Moreover, due to the sampling error in $\boldsymbol{\mu}'[D]$, even the correct solution should not solve Eq. (41) exactly. Rather, the solution error,

$$\chi(\mathbf{P}_r)^2 = \frac{1}{N_\mu} \sum_{n=1}^{N_\mu} \left[\frac{(\mathbf{V} \cdot \mathbf{P}_r)_n - \mu'_n[D]}{\mu'_n[D]} \right]^2, \quad (43)$$

should have a nonzero value equal to the (unknown) size of the sampling error. If χ is too small, the solution is “fitting the error” instead of giving a more accurate result.

In such cases, “regularizing” conditions are chosen to select from among the possible solutions.¹¹¹⁻¹¹⁴ In our case, the constrained minimization,

$$\begin{aligned} \min_{\mathbf{P}_r} & \left[\chi(\mathbf{P}_r)^2 + \beta R_1(\mathbf{P}_r) + \alpha R_2(\mathbf{P}_r, P_{\text{sp}}) \right] \\ \text{subject to} & \sum_{i=1}^{N_P} P_{r,i} = 1 \\ \text{and} & P_{r,i} \geq 0; \quad i = 1, \dots, N_P, \end{aligned} \quad (44)$$

expresses our choices. The first constraint requires that $\mu_0[\mathbf{P}_r] = 1$, that is, the solution must have unit area. The second constraint requires the probability distribution to be positive everywhere. Because we are dealing with quantum yields, we also set $S_{\min} = 0$, implying that the probability of a negative signal is zero. The maximum possible signal $S_{\max}$ is less obvious. However, by checking that the recovered probability reaches zero before this point, we can verify that our choice was sufficiently large, after the fact.

The conditions so far reject many mathematically good solutions by using physical requirements that are nearly certain. The regularizing functions, $R_1(\mathbf{P}_r)$ and $R_2(\mathbf{P}_r, P_{\text{sp}})$, impose requirements that are also grounded in physical expectations, but that are more subjective. The main constraint is to favor a smoother solution over a jagged, noisier one. Such smoothing regularizers also favor solutions with broader peaks over ones with narrow peaks. Paper 1¹³ found that

$$R_1(\mathbf{P}_r) = \frac{1}{N_P - 1} \|\mathbf{D} \cdot \mathbf{P}_r\|^2, \quad (45)$$

where

$$\mathbf{D} = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix}, \quad (46)$$

works well as a smoothing regularizer in this problem. This function approximates the average first derivative of the distribution and is a standard choice.^{110,111} The magnitude of the regularizing parameter β reflects the importance given to smoothness relative to the importance of matching the data.

The sparsity regularizer,

$$R_2(\mathbf{P}_r, P_{\text{sp}}) = \sum_{P_{r,i} < P_{\text{sp}}} P_{r,i}, \quad (47)$$

discriminates against low probability areas, specifically probabilities below P_{sp} , and pushes them toward zero. It is justified by the expectation that data are poor at distinguishing between zero and low probability and that it is better not to infer probability where it may not exist. Its effects are to remove baseline noise and low-level background to narrow the wings of peaks and to reduce interpeak intensities (Paper 1¹³). In the examples in this paper, this regularizer will not be needed, but it is included for generality.

These regularizing functions should not push the solution outside the data's error range. Unfortunately, the correct error level is not known. A more useful measure is $\chi/\chi_{\min}$, the ratio of the error of a fully regularized solution χ , to one based on only high-confidence constraints $\chi_{\min}$ ($\alpha = \beta = 0$). The set of solutions with $\chi/\chi_{\min}$ only modestly larger than one defines the “error bars” of the result.

Finally, we must choose the number of moments to use N_μ . On one hand, the error in the moments increases exponentially with the order of the moment.¹³ Including moments beyond some number $N_{\mu,\text{max}}$ will degrade the quality of the solution. The value of $N_{\mu,\text{max}}$ depends on the quality of the data. On the other hand, there is a minimum number of moments that are needed to for an adequate recovery, $N_{\mu,\text{min}}$. This number increases with the complexity of the system and serves as a useful quantification of a system's complexity. As shown in Paper 1,¹³ for two discrete states, $N_{\mu,\text{min}} = 4\text{--}5$; for three, $N_{\mu,\text{min}} = 7\text{--}8$; and for four, $N_{\mu,\text{min}} \approx 16$. For good quality data, $N_{\mu,\text{max}} > N_{\mu,\text{min}}$, and nearly the same result is found for N_μ anywhere between these limits. For poor quality data, $N_{\mu,\text{max}} < N_{\mu,\text{min}}$, the solution changes erratically with N_μ , and a good recovery is not possible.

Do these methods work as well on photon noise as they did on additive noise? Given the good noise correction seen in Fig. 3(f), we expect them to, at least when the signal is high and noise–signal correlation is the only new issue. Figure 4 shows the equilibrium distribution (solid, black curve) recovered from the high-intensity, photon-noise dataset [see Fig. 2(b)]. The peaks are sufficiently narrow to show that the system consists of three, distinct states. There are error ranges on the state positions, which are expressed as peak widths. The widths vary across the signal range. Not surprisingly, the high-signal state, from which the most photons are collected, has the most precise position.

Comparing the cumulative probabilities of the original model (dashed, red curve) and of the recovered distribution (solid, red curve) shows that the peak positions and areas are accurately recovered. Note that the recovered signal differs from the absolute quantum yield by two factors that are generally unknown experimentally: the instrumental sensitivity S^0 and the background

level B [see Eq. (6)]. Figure 4 and subsequent figures correct for the sensitivity. However, they leave in the background as a reminder that it is not part of the noise correction and must be corrected with additional measurements on a blank sample. A large number of moments are accurate: $N_\mu = 15$ were used in Fig. 4 and many more can be used, $N_{\mu,\max} > 15$. This results reflects the relatively high signal-to-noise ratio of these data (SNR = 2.4). Overall, the methods of Paper 1¹³ can be used on signal-correlated noise without modification.

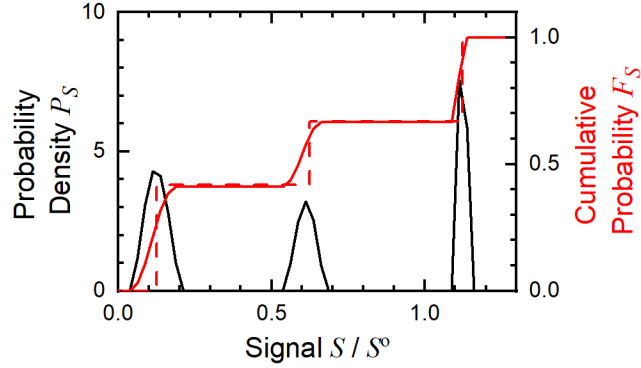


FIG. 4. A state space recovered from data with noise–signal correlation [high-intensity, photon-noise dataset, see Fig. 2(b)]. The equilibrium distribution (black, solid curve) and cumulative probability (red, solid curve) are shown. The correct cumulative probability is shown as the red, dashed curve. See Table S2 in the supplementary material for the regularization conditions.

The effect of noise quantization is introduced in Fig. 5. It compares results from the medium-intensity, photon-noise data to those from the additive-noise data. These datasets have similar signal-to-noise ratios (SNR = 0.76 and 0.58, respectively), which allows a more direct comparison of noise types. Figure 5(a) compares the distributions; Fig. 5(b) shows the corresponding cumulative probabilities. The true results are also shown in black. Almost identical distributions are recovered with either type of noise, and these distributions agree well with the system’s true properties. No change in methods is needed for quantized noise.

However, there is a discernable, quantitative difference in the results. For additive noise, nearly identical results were found over the range $N_\mu = 8$ –13, whereas the results for photon noise are only unchanging for $N_\mu = 8$ –10. That is, $N_{\mu,\max} = 13$ accurate moments are extracted with additive noise, but only $N_{\mu,\max} = 10$ accurate moments with photon noise. Fortunately, this number is higher than the minimum for three states, $N_{\mu,\min} = 7$ –8, and a good recovery is still possible (Fig. 5). However to achieve an entirely identical recovery, photon noise requires somewhat higher quality data than Gaussian–additive noise does. This result is consistent with the earlier

observation of higher sampling error in Fig. 2(f) versus Fig. 2(c). Thus, the demands for data quality with photon noise are higher than with additive noise, but only slightly higher.

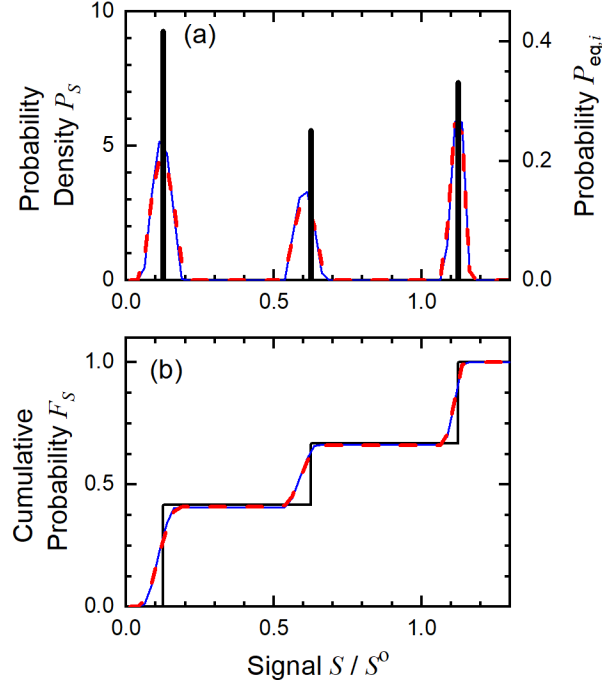


FIG. 5. Comparison of state distributions recovered from data with additive versus photon noise [see Figs. 2(a) and 2(c), respectively]. Equilibrium distributions (a) and cumulative probabilities (b) recovered from medium-intensity data with photon noise using $N_\mu = 10$ moments (red, dashed curves) or with Gaussian-additive noise using $N_\mu = 13$ moments (blue, thin curves). These datasets have similar signal-to-noise ratios (SNR = 0.76 or 0.58, respectively). The correct values are shown as solid, black bars in (a) or as the solid, black curve in (b).

We now turn to recovering the system's dynamics. The sampling error in the first point of the moment-correlation functions reflects the error along the entire correlation function (Fig. 3). Thus, we expect that the ability to recover equilibrium distributions will extend to the recovery of kinetics as well. A general discussion of this process is in Refs. 23,32,33; the specifics used here are described in the supplementary material, Sec. SI. We use the photon-noise data used in Fig. 5 and shown in Fig. 2(c).

The recovery of kinetics is simplified by ignoring the linewidths in the recovered distributions. Thus, the distribution is reduced to three discrete states. In the simplified calculation, only three moment-correlation functions are used: $M'_{1,1}[D](\tau)$, $M'_{1,2}[D](\tau) = M'_{2,1}[D](\tau)$ and $M'_{2,2}[D](\tau)$. (Higher moment functions would be useful if we wanted to look for dynamics within the linewidths.) By itself, the linear-correlation function $M'_{1,1}[D](\tau)$ identifies two time constants

[Fig. 3(g)], but it cannot assign them to specific state-to-state transitions. The two, additional correlation functions are sufficient to resolve this ambiguity.

Results are shown in Fig. 6. The correlation function is first converted to the Green's function, which is more easily plotted as the joint probability [Eq. (4), Figs. 6(a)–6(e)]. This function is symmetric in the two signals, S_0 and S_1 . At $\tau = 0$, this probability is diagonal, with the equilibrium distribution along the diagonal [Fig. 6(a)]. By $\tau = 1.1$ ms, a cross peak between the two low-signal peaks is clearly visible [Fig. 6(b)], and it continues to grow to its full height by $\tau = 10$ ms [Fig. 6(c)]. This peak indicates that the fast component of the autocorrelation is due to transitions between the two low-signal states. Additional cross peaks between the high-signal state and the pair of low-signal states are visible at $\tau = 100$ ms [Fig. 6(d)] and reach their full height by $\tau = 1.1$ s [Fig. 6(e)]. The timing of this peak growth shows that the slow component of the autocorrelation is due to transitions in and out of the highest-signal state. The fact that the system reaches equilibrium within the measurement time can be verified by the fact that the long time value of the joint probability is the cross product of the equilibrium distribution.

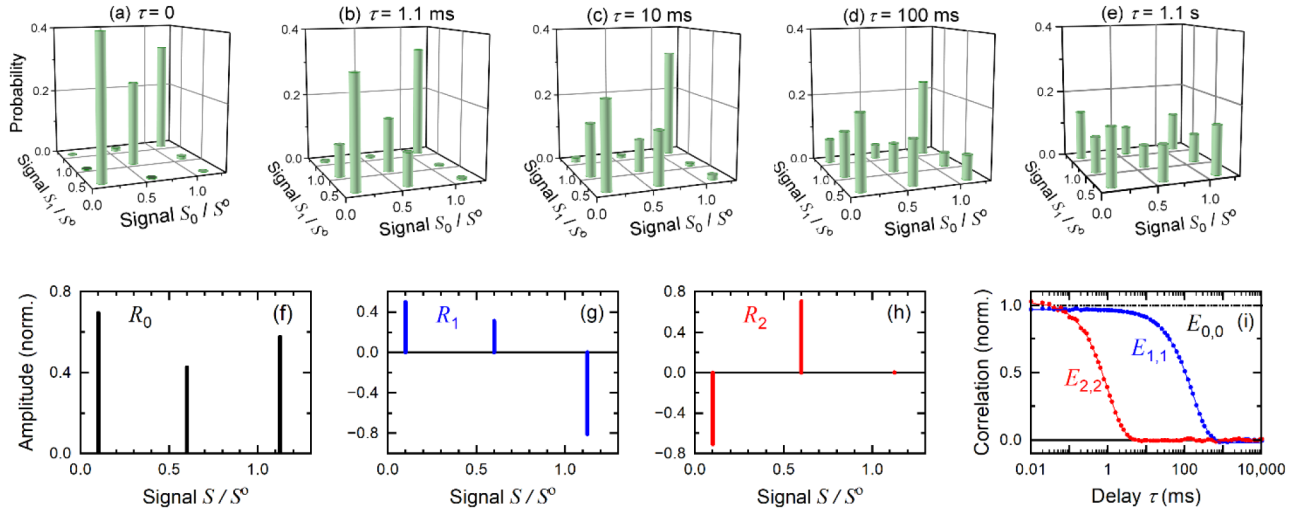


FIG. 6. Kinetics recovered from the noise-corrected, moment-correlation functions [Fig. 3(g)–3(i)]: (a)–(e) The joint-probability distribution $P(S_1, S_0; \tau)$. (f)–(h) The eigenstates $R_k(S)$ normalized to a mean-squared amplitude of one. (i) The corresponding eigendecays $E_{kk}(\tau)$ (points). The eigendecays are fit with single exponentials (solid curves, $T_f = 0.99$ ms from $E_{2,2}$, $T_{er} = 160$ ms from $E_{1,1}$). The cross-decays are zero (supplementary material, Fig. S2).

It is possible to further reduce the joint probability to eigenstates and eigendecays [Figs. 6(f)–6(i)]. Number the states from 0 for the lowest-signal state to 2 for the highest-signal state. In terms of discrete-state labels, $P_{ij}(\tau)$ is a real, symmetric matrix and can be diagonalized at any

single delay time. Applying the same transformation at all times gives the matrix $E_{ij}(\tau)$. If this matrix remains diagonal at all times, the system has right eigenstates $R_i(S)$. The diagonals $E_{ii}(\tau)$ are the eigendecays, and the eigenstates are extracted from the diagonalizing matrix.

In our example, diagonalizing at $\tau = 10$ ms makes the off-diagonal decays zero at all times (supplementary material, Fig. S2), so eigenstates and eigendecays can be found. One eigendecay $E_{0,0}(\tau)$ is constant [Fig. 6(i)], and its associated eigenstate $R_0(S)$ is proportional to the equilibrium distribution [Fig. 6(f)]. The fast eigendecay $E_{2,2}(\tau)$ [Fig. 6(i)] corresponds to the fast component of the autocorrelation function [Fig. 3(g)]. Its eigenstate $R_2(S)$ involves only exchange between the two low-signal states [Fig. 6(h)]. The intermediate eigendecay $E_{1,1}(\tau)$ [Fig. 6(i)] corresponds to the slow component of the autocorrelation [Fig. 3(g)]. Its eigenstate $R_1(S)$ shows that this time corresponds to the exchange between the highest signal state and the lower two [Fig. 6(g)]. The eigendecays fit to single exponentials [Fig. 6(i)], showing that the system is Markovian in the quantum yield. There are no hidden states, that is, no states with identical quantum yields, but different kinetics. The time constants (0.99 ms and 160 ms) recover the original rate constants of the system (1.00 ms and 100 ms) with reasonable accuracy.

In Paper 1, the same calculations were done with the same data, except with Gaussian-additive noise (see Fig. 12 in that paper).¹³ These results are nearly identical. In this example, the effects of noise quantization and noise-signal correlation are not important.

B. Theory of photon-noise sampling error

To understand noise in a more general context than the specific examples presented above, we need a more analytical treatment. This analysis will show how the results above generalize to other systems or to data with lower quality. The analysis focusses on the moments, that is, to the first point of the moment-correlation functions. However, similar considerations will hold at later delays as well.

Many of the results in Paper 1¹³ apply to any type of unbiased noise and can be used as a starting point. First, noise-correction does not introduce bias:

$$E(\mu'_n[D]) = E(\mu_n[P_S]). \quad (48)$$

[A finite time series $D(t)$ is sampled from a corresponding, ideal probability distribution $P_D(D)$. Thus, moments written as functionals of a time series include sampling error and functionals of a probability distribution do not.] The typical fractional error in a noise-corrected moment is

$$\omega_n^2 = \frac{\text{Var}(\mu'_n[D])}{\mu_n[P_S]^2}. \quad (49)$$

This total error has two contributions,

$$\omega_n^2 = \omega_S^2 + \omega_{\varepsilon,n}^2, \quad (50)$$

the signal-sampling error ω_S and the noise-sampling error $\omega_{\varepsilon,n}$.

The signal-sampling error exists even with a noise-free dataset due to its finite size. That size is characterized by the length of an individual time series T_L (the size of the time average in $\langle \dots \rangle$) compared to the longest relaxation time in the system (the ergodic time) T_{er} and by the number of molecules measured N_{en} (the size of the ensemble average in $\langle \dots \rangle$). In the large-order limit that is relevant for determining $N_{\mu,\text{max}}$, the signal-sampling error is independent of order:

$$\omega_S^2 = \frac{2a-1}{N_{\text{en}} \left(1 + \frac{\langle T_L \rangle}{T_{\text{er}}} \right)}. \quad (51)$$

The details of the equilibrium distribution determine a , which is approximately constant and is of order unity.

Signal-sampling noise is of secondary interest for the current discussion. First, it is independent of the nature of the noise. When the noise is significant, and the dataset is not very small, signal-sampling noise is smaller than the noise-sampling error and is less important in recovering the system's properties.¹³ As an example, the black curve in Fig. 3(i) has signal-sampling error, which is visually undetectable, whereas the red points also include noise-sampling error, which can be seen at high orders.

With simulated data, noise correction should be perfect for an infinite dataset [Eq. (48)]. However, in a finite dataset, errors in sampling the noise lead to imperfections in its removal. This noise-sampling error is¹³

$$\omega_{\varepsilon,n}^2 = \frac{1}{\mu_n[P_S]^2 N_{\text{er}}} \sum_{k=1}^n \binom{n}{k} E \left(\text{Var}(D|S)^k S^{2(n-k)} \right). \quad (52)$$

It depends on the noise process $\text{Noise}(D|S)$ through its variance $\text{Var}(D|S)$. For additive noise,

$$\text{Var}(D|S) = \sigma_{\varepsilon}^2 \quad (53)$$

[Eq. (16)]. To treat photon noise, we substitute¹⁰⁸

$$\text{Var}(D|S) = \frac{S}{N_b T_{\varepsilon}}. \quad (54)$$

[Eq. (26)]. Using the same large-order approximations used in Eq. (51), photon noise gives

$$\omega_{\varepsilon,n}^2 = \frac{2a}{N_{\text{er}}} \left[\left(1 + \frac{1}{2b\text{SNR}^2} \right)^n - 1 \right] \quad (55)$$

(supplementary material, Sec. SII.1), whereas additive noise previously gave (Paper 1¹³)

$$\omega_{\varepsilon,n}^2 = \frac{2a}{N_{\text{er}}} \left[\left(1 + \frac{1}{4b^2 \text{SNR}^2} \right)^n - 1 \right]. \quad (56)$$

Like a, b is a quantity that depends on the details of the equilibrium distribution, but that is constant and near unity at high orders. Thus, despite the significant differences between additive and photon noise, the differences in the sampling error are minor, when compared on the basis of their signal-to-noise ratios.

Note that Eqs. (55) and (56) predict the absolute size of the sampling error and have been tested against simulated data (see Paper 1¹³), but that $\omega_{\varepsilon,n}$ is not directly observable in real data. A more practical measure of data error is the maximum number of accurate moments $N_{\mu,\text{max}}$ that can be recovered from a dataset. To estimate this quantity for a certain dataset size and signal-to-noise ratio, we need $\omega_{\varepsilon,\text{max}}$, the maximum fractional error that a moment can have before it degrades the recovery. We previously found that $\omega_{\varepsilon,\text{max}} = 5 \times 10^{-3}$ is a reasonable value.¹³ Setting $\omega_{\varepsilon,n} = \omega_{\varepsilon,\text{max}}$ when $n = N_{\mu,\text{max}}$, Eq. (55) gives

$$N_{\mu,\text{max}} = \frac{\ln\left(\frac{1}{2} N_{\text{er}} \omega_{\varepsilon,\text{max}}^2\right)}{\ln\left(1 + \frac{1}{2} \text{SNR}^{-2}\right)} \quad (57)$$

for photon noise, and Eq. (56) gives

$$N_{\mu,\text{max}} = \frac{\ln\left(\frac{1}{2} N_{\text{er}} \omega_{\varepsilon,\text{max}}^2\right)}{\ln\left(1 + \frac{1}{4} \text{SNR}^{-2}\right)} \quad (58)$$

for additive noise. Here and in the rest of the paper, we set $a = b = 1$. This simplification is sufficient to predict trends in and to estimate magnitudes for generic datasets.

These formulas predict that a dataset with photon noise will need an approximately 1.4 time higher signal-to-noise ratio to be equivalent to a dataset with Gaussian noise. For the Gaussian–additive-noise and medium-intensity, photon-noise datasets, these formulas predict $N_{\mu,\text{max}} = 13$ and $N_{\mu,\text{max}} = 11$, respectively. These results are consistent with our fitting results, $N_{\mu,\text{max}} = 13$ and $N_{\mu,\text{max}} = 10$, respectively.

We can rephrase the problem to ask how large a dataset is needed to recover a given system. Clearly, a more complex system will require more data, even for a fixed signal-to-noise ratio. Here, a system's complexity is measured by the minimum number of accurate moments needed to recover its properties, $N_{\mu,\text{min}}$. It enters through the basic criterion for a successful recovery, $N_{\mu,\text{min}} \lesssim N_{\mu,\text{max}}$. For photon noise, it is useful to characterize the size of the dataset through the total

number of photons collected $N_{\text{ph}} = \bar{N}_{\text{ct}} N_{\text{er}}$. Similarly, the noise can be characterized through the average number of counts per time bin $\bar{N}_{\text{ct}}$.

Equation (55) gives different answers in two limits. In the high-count-rate limit, $\bar{N}_{\text{ct}} \gg \frac{1}{2}$,

$$N_{\text{ph}} \gtrsim \omega_{\varepsilon, \text{max}}^{-2} N_{\mu, \text{min}}. \quad (59)$$

This result is typical for noise analysis. The number of photons needed is a fixed value that depends on the complexity of the system being analyzed. If the count rate $\bar{N}_{\text{ct}}$ is lowered, the total number of data points N_{er} must increase in proportion, by increasing either the length of the measurement T_L or the number of molecules measured N_{en} . However, the count rate does not matter as an independent factor; noise level and dataset size can be freely traded.

This behavior changes in the low-count-rate regime. In the limit $\bar{N}_{\text{ct}} \ll \frac{1}{2}$, Eq. (55) gives

$$N_{\text{ph}} \gtrsim \frac{2}{\omega_{\varepsilon, \text{max}}^2 (2\bar{N}_{\text{ct}})^{N_{\mu, \text{min}}-1}}. \quad (60)$$

Now, as the count rate drops, the total number of photons or the total number of data points must increase by a high power of the drop, typically a power of 5–15. It is no longer practical to compensate for a low count rate by collecting more data. Thus, $\bar{N}_{\text{ct}} \approx \frac{1}{2}$ ($\text{SNR} \approx 1/\sqrt{2}$) is a “floor” to the signal level that can be successfully analyzed. A similar floor exists for additive noise at a similar level, $\text{SNR} \approx \frac{1}{2}$ (Paper 1¹³).

For a statistically consistent quantity, the results from an arbitrarily large dataset ($N_{\text{er}} \rightarrow \infty$) always converge to the correct answer, regardless of the signal-to-noise ratio. At high signal levels, the rate of convergence is slow (with the square root of the dataset size), but familiar. However, for a signal-to-noise ratio that is below the signal floor, the rate of convergence becomes so unfavorable that collecting sufficient data becomes impractical.

The existence of a signal floor has a major effect on the maximum time resolution of the experiment. Above the floor, correlation-based noise-removal preserves time resolution better than time binning. The blurring of time due to noise removal [Eq. (38)] can be estimated (Paper 1¹³) to give the time resolution of the experiment as

$$T_r = \frac{N_{\mu, \text{min}} + 1}{2} T_{\varepsilon}; \quad ST_{\varepsilon} > \frac{1}{2}. \quad (61)$$

The exact signal level is not relevant to the time resolution, and the time resolution can be improved by choosing a faster detector.

On the other hand, if the raw count rate is below this floor, correlation methods are inefficient, and the data must be binned until $\bar{N}_{\text{ct}} \approx \frac{1}{2}$. Accounting for the binning, the maximum experimental time resolution is

$$T_r = \frac{N_{\mu,\min} + 1}{4S}; \quad ST_\varepsilon < \frac{1}{2}. \quad (62)$$

Now the detector time resolution is irrelevant. The experimental time resolution can only be improved by increasing the signal from the sample, either by increasing the collection efficiency or by using more intense excitation.

These formulas resolve the apparent paradox posed in the introduction. The time resolution a single-molecule experiment can be much longer than the time resolution of its detector because the nature of signal averaging changes at the signal floor. Simply increasing the detector speed with a constant signal level will inevitably put the experiment below this limit.

IV. CORRECTING PHOTON-COUNTING NOISE

A. Low-bias conditions

The previous section (Sec. III) showed that noise from linear photon detectors can be corrected using moment-correlation functions. However, many single-molecules experiments are done with photon-counting detectors because of their high time resolution and low background. Often, the count rate is kept low, and the detector is quasi-linear. With extensive binning, photon-counting noise becomes quasi-Poisson, and the data can be treated as if they came from a linear photon detector. However, the binning causes the time resolution of the experiment to be much less than the detector time resolution.

On the other hand, “photon-by-photon” analyses directly treat unbinned photon-counting data. Photon-by-photon, linear-correlation functions give a time resolution equal to the detector time resolution,^{9,14-17,100} but they do not exhaust the information in the time series. In contrast, photon-by-photon, parametric analyses provide a complete interpretation of the data, but have a lower time resolution.^{24,38,40,42,43,47,55-58,100} This subsection (Sec. IV.A) looks at a complete, nonparametric analysis of photon-counting data in the unbiased regime and elucidates the factors limiting the time resolution. The analysis argues for using conditions with bias, a topic treated in the following subsection (Sec. IV.B).

We must first ask whether a moment-based method can work at all on unbinned photon-counting data. In its raw form, the data consist entirely of ones and zeros. Taking a power of the time series leaves it unchanged. Every moment-correlation function is identical to the linear-correlation function, and there is no advantage to using higher orders.

Primed powers solve this problem. Every value of $D^{k'}(t_i)$ actually samples a sequence of k points starting at t_i [Eq. (38)]. A nonzero $D^{k'}(t_i)$ occurs only when the entire sequence is nonzero. The number of such occurrences varies with the power, as shown in Fig. 7. The detailed shape of this variation is not exponential due to the distribution of signal levels, and this non-exponential

variation is exploited by an analysis using primed moments and moment-correlation functions. In the limit of an infinite amount of data, a nonlinear-correlation analysis could be completed with the same high-time resolution that a linear-correlation function has.

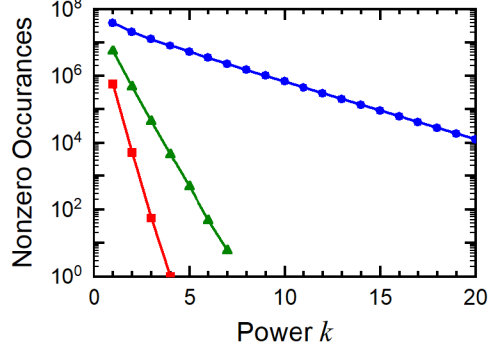


FIG. 7. The number of nonzero occurrences of $D^{k'}$ in unbinned, photon-counting data with medium intensity and low bias ($\langle \bar{N}_{ct}^0 \rangle = 5.83 \times 10^{-3}$, red squares), high intensity and moderate bias ($\langle \bar{N}_{ct}^0 \rangle = 5.83 \times 10^{-2}$, green triangles) or very high intensity and high bias ($\langle \bar{N}_{ct}^0 \rangle = 0.583$, blue circles).

Unfortunately, the sampling error in a finite dataset alters this conclusion. In a finite dataset with a low count rate, the number of nonzero values of a primed power drops very rapidly with the power. Figure 7 shows result from data with photon-counting noise [Eq. (30)] and with a detector 100-times faster ($T_e = 0.1 \mu s$) than the one used in the examples in Sec. III. The red line comes from a medium-intensity dataset ($\langle S \rangle = 58.3$ photon/ms). At this count rate ($\langle \bar{N}_{ct}^0 \rangle = 0.00583$), bias is low, but $D^{4'}(t_i) = 1$ occurs only once, and $D^{3'}(t_i) = 1$ occurs fewer than 100 times, even in a very large dataset. The very high sampling error makes these, or any higher powers, unusable.

The situation improves somewhat for a high-intensity dataset (Fig. 7, green line, $\langle S \rangle = 583$ photon/ms, $\langle \bar{N}_{ct}^0 \rangle = 0.0583$) at the expense of introducing some bias. However, it is still not possible to accurately calculate primed powers up to the order required to fully recover the system's properties, $N_{\mu, \min} = 7-8$. Thus, low-to-moderate bias data must be binned before a complete correlation analysis can be done. The high-time resolution of photon-by-photon linear correlations does not extend to higher correlation functions.

The same reasoning applied to moment-correlation functions. The need for progressively more binning at higher orders is illustrated in Fig. 8 using the medium-intensity, low-bias dataset ($\langle \bar{N}_{ct}^0 \rangle = 0.00583$). For the linear-correlation function $M_{1,1}[D](\tau)$, good results can be found without any binning [Fig. 8(a), black curve]. For $M_{2,2}[D](\tau)$, no binning produces extremely noisy results [Fig. 8(b), black curve]. However, binning by $N_b = 10$ (blue curve) produces an acceptable

correlation function [Fig. 8(b), blue curve]. (Our correlation algorithm uses quasi-logarithmic time bins, which creates additional signal averaging at later times. Thus, the noise at early times limits the overall analysis.) For $M_{4,4}[D](\tau)$, binning by $N_b = 10$ leaves excessive noise [Fig. 8(c), blue curve], but with $N_b = 100$, the result is acceptable [Fig. 8(c), red curve]. For $M_{8,8}[D](\tau)$, $N_b = 100$ is inadequate [Fig. 8(d), red curve], and further binning to $N_b = 400$ is needed [Fig. 8(d), green curve]. The amount of binning required depends on the number of moments needed and, thus, on the complexity of the system.

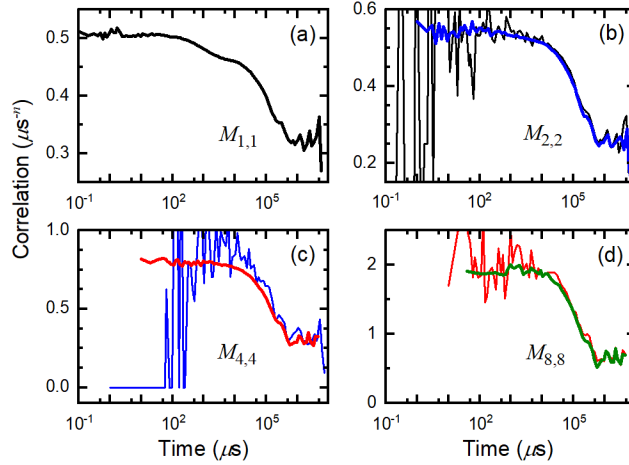


FIG. 8. Effect of time binning on moment-correlation functions from medium-intensity data with photon-counting noise [$\langle \bar{N}_{\text{ct}}^0 \rangle = 5.83 \times 10^{-3}$, also see Fig. 7, red squares]: (a) $M_{1,1}[D](\tau)$ with no binning ($N_b = 1$, black curve), (b) $M_{2,2}[D](\tau)$ with no binning ($N_b = 1$, thin black curve) or with 10-fold binning ($N_b = 10$; heavy, blue curve), (c) $M_{4,4}[D](\tau)$ with 10-fold ($N_b = 10$; thin, blue curve) or 100-fold binning ($N_b = 100$; heavy, red curve); (d) $M_{8,8}[D](\tau)$ with 100-fold ($N_b = 100$; thin, red curve) or 400-fold binning ($N_b = 400$; heavy, green curve).

Figure 9 looks at the consequences treating low-count-rate photon-counting data as if it came from a linear photon detector. Figure 9(a) shows equilibrium distributions recovered from data from a slow, linear photon detector [see also Figs. 2(c), 3(g)–3(i), 5, and 6]. Figure 9(b) is the same, except the data are from a fast, photon-counting detector, after binning to the same time resolution. Specifically, the photon-counting data have an unbinned count rate of $\langle \bar{N}_{\text{ct}}^0 \rangle = 5.8 \times 10^{-3}$, a detector time of $T_e = 100$ ns, and have been binned by a factor of $N_b = 100$ [see also Figs. 7; Fig. 8(c), red; and 8(d), red]. The recovery from the photon-counting data with quasi-Poisson noise is of marginal quality [Fig. 9(b)]. The distribution changes erratically depending on the exact number of moments used. This instability indicates that these low moments already have significant error. In comparison, linear-photon-detection data with true-Poisson noise give a stable

recovery under identical conditions [Fig. 9(a)]. Even at a count rate of less than 1%, photon-counting entails a loss of information with a detectable effect.

Fortunately, the loss is small and is easily corrected. Figure 9(c) shows good recovery with additional time binning by a factor of four. Alternatively, Fig. 9(d) shows a good recovery with a 10-times higher count rate. Overall, having a higher time-resolution detector has not improved the results of the experiment.

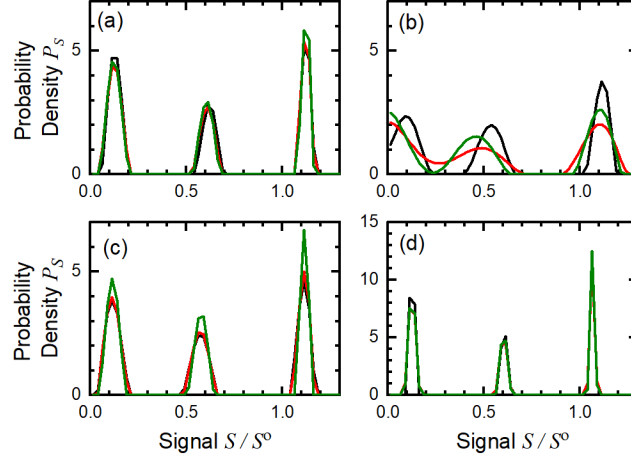


FIG. 9. Comparison of distribution recovery from data from a slow, linear photon detector with true Poisson noise ($T_\varepsilon = 10 \mu\text{s}$), or from data from a high-speed, photon-counting detector with quasi-Poisson noise ($T_\varepsilon = 0.1 \mu\text{s}$, $N_b = 100$). Within each subpanel, recoveries are shown using different numbers of moments: $N_\mu = 8$ (black), 9 (red), or 10 (green). (a) Linear-photon-detector data with medium intensity and 10- μs bins (see also Fig. 5). (b) Photon-counting data with medium intensity and 10- μs time bins. (c) Photon-counting data with medium intensity and 40- μs time bins. (d) Photo-counting data with high intensity and 10- μs time bins.

We can generalize this conclusion by estimating the effect of sampling error on unbinned, unbiased photon-counting data, as we did in Sec. III.B for photon noise. Details are given in the supplementary material (Sec. SII.2). For a given set of data, the estimated number of accurate moments is

$$N_{\mu,\max} = \frac{\ln\left(N_{\text{cr}} \omega_{\varepsilon,\max}^2\right)}{\ln\left[\left\langle \bar{N}_{\text{ct}}^{\text{o}} \right\rangle^{-1}\right]} \quad (63)$$

[compare to Eqs. (57) and (58)]. Conversely, if a certain number of moments are needed, the required number of collected photons can be estimated:

$$N_{\text{ph}} \gtrsim \frac{1}{\omega_{\varepsilon, \text{max}}^2 \left\langle \bar{N}_{\text{ct}}^o \right\rangle^{N_{\mu, \text{min}} - 1}}. \quad (64)$$

As expected, both of these equations are very similar to the photon-noise result in the limit of low count rates. In particular, Eq. (64) matches the result for photon noise below the signal floor [Eq. (60)]. Photon-counting data never enters the high-signal region where the more conventional $\text{SNR}^{1/2}$ dependence [Eq. (59)] holds.

Because they are effectively below the signal floor, unbiased photon-counting data always require time binning and always have an experimental time resolution similar to that of low-count-rate, photon-noise data:

$$T_r = \frac{N_{\mu, \text{min}} + 1}{4S}; \quad \left\langle \bar{N}_{\text{ct}}^o \right\rangle < \frac{1}{2}. \quad (65)$$

[see Eq. (62)]. The resolution depends critically on the signal level, but it is independent of the detector.

However, raising the signal level significantly will begin to introduce bias. This effect was seen in a close examination of Fig. 9. Using 10-times higher signal [Fig. 9(d)] gave improved time resolution [compare to Fig. 9(c)]. However, a small amount of bias is now evident: the position of the high-intensity peak is shifted from its correct position [see also Figs. 10(b) and 10(e) below]. Increasing the intensity to improve the time resolution will further increase the bias.

B. Bias correction and the optimal signal level

Given the trade-off between time resolution and bias, what is the optimal signal level to use in photon counting? Given that linear photon detectors have an optimum signal level of $\bar{N}_{\text{ct}}^o = 1/2$, it is tempting to suggest that the same level is optimal for a photon-counting detector. Information theory supports this answer.¹¹⁵ An unbinned, photon-counting time series has two outcomes when the system is generating signal S : a count of one occurring with probability $\bar{N}_{\text{ct}}^o$ and a count of zero with probability $1 - \bar{N}_{\text{ct}}^o$. Information entropy is gained at a rate $\mathcal{H}(D|S) = (1 - \bar{N}_{\text{ct}}^o) \log(1 - \bar{N}_{\text{ct}}^o) + \bar{N}_{\text{ct}}^o \log(\bar{N}_{\text{ct}}^o)$ as the time series is extended. No information is gained in the unbiased limit, $\bar{N}_{\text{ct}}^o \rightarrow 0$, or in the saturated limit, $\bar{N}_{\text{ct}}^o \rightarrow 1$. The maximum information rate occurs when $\bar{N}_{\text{ct}}^o = 1/2$. Not knowing the distribution of S beforehand, the optimal experimental strategy is to set the average count rate to $\langle \bar{N}_{\text{ct}}^o \rangle = 1/2$.

A dataset close to this condition was generated by increasing the signal another factor of 10, to $\langle S \rangle = 580$ photon/ms ($\langle \bar{N}_{\text{ct}}^o \rangle = 0.583$). With this very-high-intensity dataset, the blue line and circles in Fig. 7 show that powers beyond $k' > 15$ are well sampled with no binning. The distribution recovered without binning is shown as the green, dashed curve in Fig. 10(c). It has

the correct number of narrow peaks with the correct areas and with no compromise in the time resolution of the detector. However, the peak positions are strongly biased toward low values.

How can a dataset with the maximum amount of information give an erroneous result? The resolution comes from realizing that bias is deterministic and, therefore, correctable, so long as the nature of the bias is understood. For photon counting, the bias is given by Eq. (31). Defining a new variable $\hat{S}$ by

$$\hat{S} = \frac{1}{T_\varepsilon} \left(1 - e^{-ST_\varepsilon} \right) \quad (66)$$

creates a transformation that is nonlinear, but invertible,

$$S = \frac{1}{T_\varepsilon} \ln \left(1 - T_\varepsilon \hat{S} \right). \quad (67)$$

In terms of this variable, the detector is unbiased,

$$E(D | \hat{S}) = \hat{S}. \quad (68)$$

Recovery of the data with the methods used so far gives results in terms of this transformed variable, $\hat{P}_r(\hat{S}_0)$ and $\hat{G}_r(\hat{S}_1 | \hat{S}_0; \tau)$. Applying the inverse transformation [Eq. (67)] to these results gives the desired $P_r(S_0)$ and $G_r(S_1 | S_0; \tau)$.

To implement this scheme requires addressing one subtle point, which can be seen by rewriting Eq. (44) to include a preferred solution $\mathbf{P}_0$:

$$\begin{aligned} \min_{\mathbf{P}_r} & \left[\chi(\mathbf{P}_r)^2 + \beta R_1(\mathbf{P}_r - \mathbf{P}_0) + \alpha R_2(\mathbf{P}_r - \mathbf{P}_0, P_{\text{sp}}) \right] \\ \text{subject to} & \quad \sum_{i=1}^{N_P} P_{r,i} = 1 \\ \text{and} & \quad P_{r,i} \geq 0; \quad i = 1, \dots, N_P, \end{aligned} \quad (69)$$

This equation selects solutions that are both close to the data and close to $\mathbf{P}_0$. In a Bayesian perspective, $\mathbf{P}_0$ represents prior information about the solution, which is combined with the new data to select an optimum solution. Often no such prior information exists, and a flat, “no-information” prior should be used.¹¹⁶ A flat, continuous probability density $P_0(S) = \text{constant}$ discretizes to a constant probability vector $\mathbf{P}_0$, with $P_{0,i} = 1/N_P$ if the signal vector $\mathbf{S}$ has evenly spaced points. In this case, $R_1(\mathbf{P}_0) = 0$. Thus, even though the earlier Eq. (44) does not display $\mathbf{P}_0$ explicitly, it includes a flat prior implicitly. However, a flat prior is only appropriate for a variable with translational invariance, such as S .¹¹⁶ If such a variable is transformed nonlinearly, for example to $\hat{S}$, the no-information prior should also be transformed and will no longer be flat.

Failure to account for this change in the prior gives poor results for recovering the distribution (supplementary material, Fig. S3).

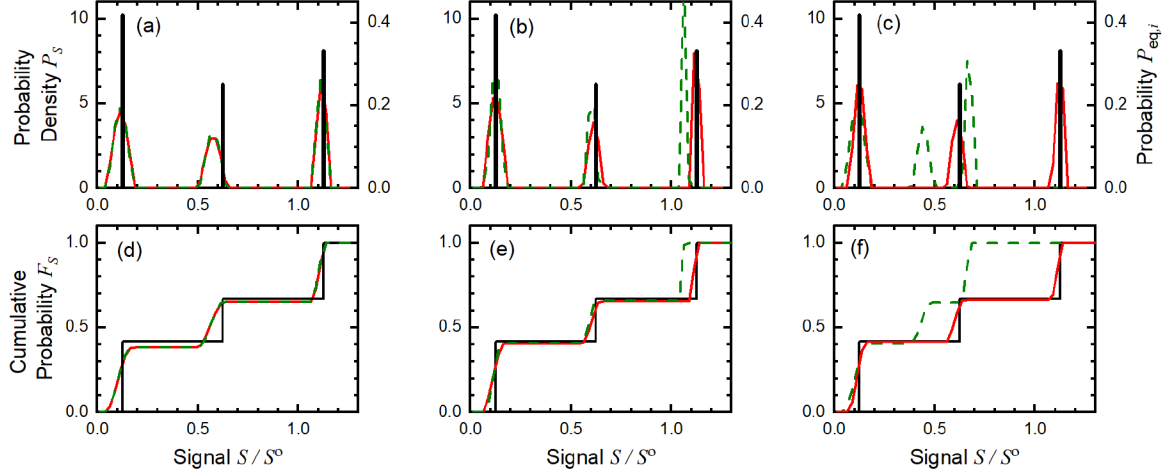


FIG. 10. Effect of bias correction. Distributions recovered from data with photon-counting noise with different count rates and levels of bias. Results without bias correction are shown as green, dashed curves; those with bias correction are shown as red, solid curves; and the correct results are shown as black, vertical bars or black, solid curves. (a) and (d): low-count-rate ($\langle \bar{N}_{ct}^0 \rangle = 5.83 \times 10^{-3}$, medium-intensity) and low-bias data. (b) and (d): medium-count-rate ($\langle \bar{N}_{ct}^0 \rangle = 5.83 \times 10^{-2}$, high-intensity) and medium-bias data. (c) and (e): optimal-count-rate ($\langle \bar{N}_{ct}^0 \rangle = 0.583$, very-high-intensity) and high-bias data. Top row – Distributions; bottom row – Cumulative probability functions.

To modify the prior, we could take unequal probabilities $P_{0,i}$ on a the original, evenly spaced $\mathbf{S}$ to give $\hat{\mathbf{P}}_0(\mathbf{S})$. Instead, we keep the probabilities $P_{0,i}$ equal but transform the evenly spaced $\mathbf{S}$ to an unevenly spaced $\hat{\mathbf{S}}$ using Eq. (66). The resulting $\mathbf{P}_0(\hat{\mathbf{S}})$ is also a discretization of the modified probability density. It satisfies $R_1(\mathbf{P}_0) = 0$, so Eq. (44) can be solved as before, without an explicit $\mathbf{P}_0$. Bias is removed from the solution $\mathbf{P}_r$ by pairing it with the original signal vector to give $\mathbf{P}_r(\mathbf{S})$.

The effectiveness of this process is shown in Fig. 10, which compares distributions without (green, dashed lines) and with (red, solids lines) bias correction. For the medium-intensity data [Figs. 10(a) and 10(d)], the bias is low, and the correction has little effect. Figures 10(b) and 10(e) for the high-intensity data show that a modest bias can be accurately corrected. Most encouragingly, the very strong bias in the very-high-intensity data can be corrected accurately [Fig. 10(c) and 10(f)]. Thus, accurate bias correction demands no higher data quality than the rest of the property recovery.

V. WHAT LIMITS THE TIME RESOLUTION IN SINGLE-MOLECULE EXPERIMENTS?

Single-molecule spectroscopy is regarded as a slow technique, often limited to millisecond times and longer. However, simulations have suggested that photon-by-photon parametric modeling should be able to recover decay times of tens of microseconds.⁴⁸ Autocorrelations are ultimately limited by the fluorescence lifetime of the molecules and thus can give decay times of tens of nanoseconds.⁹ At the same, the low time resolution of histogramming limits the exploration of the state space, even with high time-resolution experiments.^{30,31} This section attempts to clarify the fundamental factors that limit the time resolution of single-molecule experiments, in light of the results of this paper.

A. A submicrosecond-time-resolution scenario

We begin by posing a scenario for a high-time-resolution experiment. The realism of this scenario will be discussed in the following subsection (Sec. V.B). As a system of moderate complexity, we again take a molecule with the same three states shown in Fig. 1, and thus the same $\mathbf{P}$ and $\mathbf{S}$ vectors. However, we increase the rates to give $T_f = (k_{12} + k_{21})^{-1} = 1 \mu\text{s}$ and $T_{\text{cr}} = (k_{23} + k_{32})^{-1} = 4 \mu\text{s}$. The rate matrix is then

$$\mathbf{k} \cdot \mu\text{s} = \begin{pmatrix} -0.38 & 0.62 & 0 \\ 0.38 & -0.77 & 0.11 \\ 0 & 0.14 & -0.11 \end{pmatrix}. \quad (70)$$

Each molecule photobleaches with a time constant that allows an average of 4.7×10^5 photons to be collected from each molecule.

We take a photon-counting detector with a dead time of $T_e = 25 \text{ ns}$. Avalanche photodiodes this fast are commercially available,¹¹⁷ whereas research-level devices are reported with dead times down to 1 ns.¹⁰²⁻¹⁰⁷ To reach the optimal signal level for the detector, we need a very high signal rate of $\langle S \rangle = 2.3 \times 10^7 \text{ counts/s}$ (high-time-resolution photon counting in Table S1, supplementary material). Given the photobleaching rate, the time series are limited to an average length of 20 ms. To acquire sufficient data, $N_{\text{en}} = 10$ molecules are measured.

Figure 11(a) shows a sample of the raw data (black) from this model as well as the ideal signal from the system's true states (red). A standard, nonparametric analysis would focus on the time-binned histogram [$N_b = 40$, Fig. 11(b), black curve] and the autocorrelation [Fig. 11(c), black points] of these data. Even with optimal binning, the histogram has weak evidence of the three underlying states. The autocorrelation retains the full time resolution of the detector, but also has weak evidence for multiple states. It is fit moderately well by a single exponential (blue solid curve), which is consistent with only two states. It is more accurately fit with a stretched

exponential (not shown, $T = 4.97 \mu\text{s}$, $\beta_{\text{st}} = 0.80$), which is consistent with a continuous state space. Thus standard nonparametric methods are inadequate for this data.

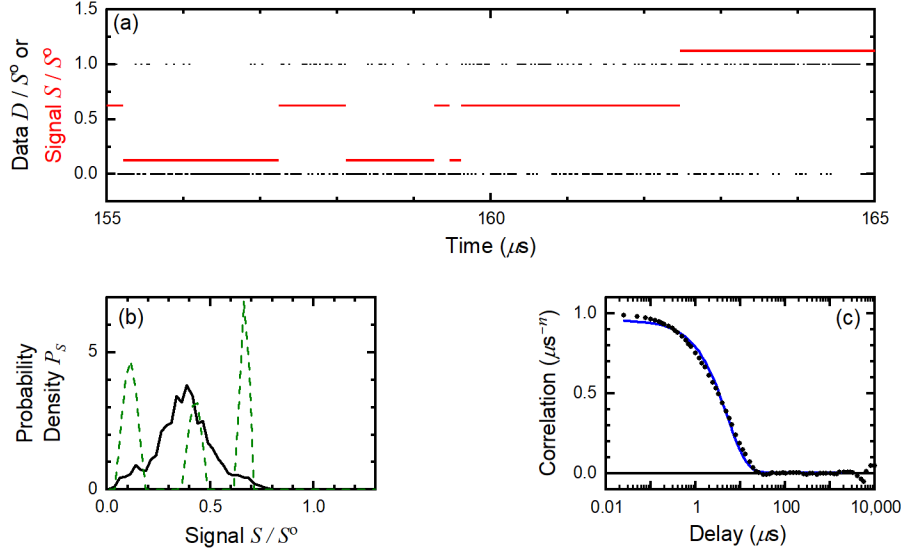


FIG. 11. Standard, nonparametric characterization of a very-high-signal, photon-counting dataset. (a) A sample of the time series: ideal signal (red points) and detected data (black points). (b) The histogram of the data with a bin time of $1 \mu\text{s}$ ($N_b = 40$; black, solid curve), and the biased distribution recovered by correlation methods (green, dashed curve). (c) Autocorrelation of the data (black points) along with a single-exponential fit (blue, solid curve; $T = 5.08 \mu\text{s}$).

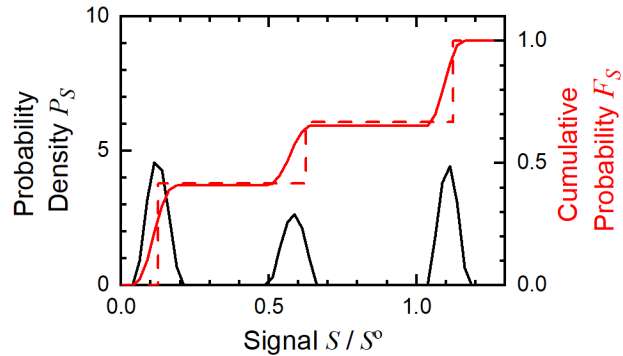


FIG. 12. Distributions recovered from the very-high-signal data after bias correction: probability densities (black, solid curve) and cumulative probabilities (red, solid curve) are compared to the correct cumulative probability (red, dashed curve). Compare to 11(b).

The suitability of a correlation analysis for this system can be estimated from our results. The complexity of a three-state system gives $N_{\mu, \min} = 8$ (Paper 1¹³), so the experimental time resolution

is estimated to be $T_r = 110$ ns [Eq. (61)]. If a decay is known to be exponential, its rate can be measured with a time resolution only 2–3 times faster than the decay time. However, we do not restrict our correlation functions to exponentials, so a more reasonable allowance is a factor of 10. Time constants greater than $1 \mu\text{s}$ should be measurable. Extrapolating Eq. (64) to these high count levels estimates that $N_{\text{ph}} = 5 \times 10^6$ photon need to be collected, close to the number in the simulation ($N_{\text{ph}} = 4.7 \times 10^6$). Thus, correlation methods should work on this dataset.

We first recover the distribution as described above. In Fig. 11(b), the biased distribution (green dashed) is compared to the simple histogram (black solid). After bias correction, the recovered distribution (Fig. 12, solid) matches the true distribution (Fig. 12, dashed), both in the positions of the states and in their populations. The correlation method avoids most of the peak broadening in the histogram. In particular, the time blurring of the states is small, implying an effective time resolution well under $1 \mu\text{s}$.

Reducing the recovered distribution to delta-function states (supplementary material, Sec. SI) yields recovered populations $\mathbf{P}_{\text{eq}} = \{0.416, 0.248, 0.336\}$ and signals $\mathbf{S} = \{0.125, 0.608, 1.114\}$, which can be compared to the input values $\mathbf{P}_{\text{eq}} = \{0.417, 0.250, 0.333\}$ and $\mathbf{S} = \{0.125, 0.625, 1.125\}$. The static features of the system are accurately recovered.

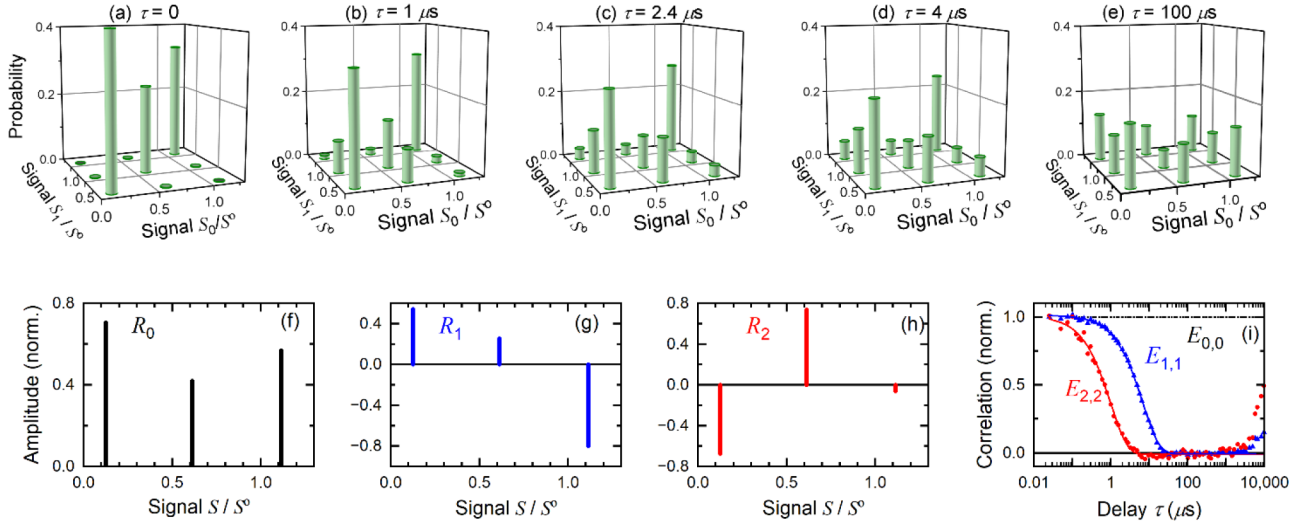


FIG. 13. Kinetics recovered from the very-high-signal, photon-counting data (compare to Fig. 6): (a)–(e) The joint-probability distribution $P(S_1, S_0; \tau)$. (f)–(h) The eigenstates $R_k(S)$ normalized to a mean-squared amplitude of one. (i) The corresponding eigendecays $E_{kk}(\tau)$ (points). $E_{1,1}$ (blue triangles) and $E_{2,2}$ (red circles) are fit with single exponentials (solid curves, $T_1 = 6.75 \mu\text{s}$ from $E_{1,1}$, $T_2 = 0.99 \mu\text{s}$ from $E_{2,2}$). The cross-decays are zero (supplementary material, Fig. S2).

The kinetic analysis used to generate Fig. 6 was repeated on this data to create Fig. 13 (supplementary material, Sec. SI). The joint-probability distribution [Figs. 13(a)–13(e)] visually identifies the connections between the states. The exchange between the low- and medium-signal states occurs first, as seen in the growth of their cross peaks in Figs. 13(b) and 13(c). Exchange with the high-intensity state only occurs later, as seen in the growth of its cross peaks in Figs. 13(d) and 13(e). With the two exchange rates being closer together in this example than in Fig. 6, the degeneracy in the growth of these two, later cross peaks is lifted. The cross peak between the high and medium-intensity states grows in first, identifying these as the directly coupled states. The low-intensity state is only indirectly coupled to the high-intensity state, delaying the growth of its cross peak.

The state-to-state connections are made more quantitative by extracting the eigenstates [Figs. 13(f)–13(h)]. The static eigenstate R_0 [Fig. 13(f)] is simply a repetition of the equilibrium distribution, but the fastest eigenstate R_2 [Fig. 13(h)] shows that the low- and medium-signal states are directly connected by the fast exchange. The intermediate eigenstate R_1 [Fig. 13(h)] shows that both the low- and high-intensity states feed into the medium-intensity state, minimizing its contribution to the eigenstate.

The eigendecays associated with these eigenstates [Fig. 13(i)] are well separated and are well fit by single exponentials, in contrast to the ambiguity of the autocorrelation function [Fig. 11(c)]. Exponential eigendecays assure us that the system is Markovian in the identified states and that there are no hidden or unresolved states affecting the kinetics. The eigenstates and their decay times can then be combined to recover the rate matrix,

$$\mathbf{k} \cdot \mu s = \begin{pmatrix} -0.37 & 0.58 & 0.03 \\ 0.35 & -0.68 & 0.07 \\ 0.02 & 0.10 & -0.10 \end{pmatrix}, \quad (71)$$

which closely reproduces the input matrix [Eq. (70)]. Thus in this scenario, all the static and kinetic features of the system are accurately recovered, with an effective time resolution in the submicrosecond regime.

B. Issues in real experiments

Real detectors have noise statistics that are more complex than the ideal models consider here.^{118,119} However, all the major forms of complexity; non-Gaussian statistics, signal-noise correlation, and quantization; have been treated, and they have only minor effects on the noise removal. That result holds because the only essential property of detector noise is that it has a faster correlation time than the signal. Further knowledge of the noise statistics is not needed, as long as the detector is linear and unbiased.

How does the noise-correlation time of these simulations T_c relate to real detectors? With an electronic autocorrelator, the detector pulsewidth or timing jitter limit the time resolution.⁹ However, it is unlikely that electronic correlators to many orders will become practical.^{18,19} The software correlation used here is dependent on the time between statistically independent measurements. Thus, T_c , as used here, is determined by the detector dead time and by any crosstalk between time bins.

With a biased detector, an accurate description of its nonlinearity is also needed. In the case of an ideal photon counter, the nonlinearity is known *a priori*. With a real detector, its nonlinearity needs to be measured, and Eq. (31) needs to be replaced by those empirical results. Otherwise, the bias correction will be unchanged.

In addition to fast detection noise, many real instruments also suffer from slow noise or baseline drift, which the current methods cannot remove. However, one can appeal to the broader premise of this work to suggest a way forward. High-order noise correlations can be measured on a blank sample. Noise and signal can then be separated on the basis of their differing correlation functions. Although a specific route to using this information efficiently has not been demonstrated, a basic idea should persist: sampling error limits the characterizing of the noise and, thereby, limits the signal recovery. Thus, the estimates of the required data quality presented here will remain relevant.

Current detector technology is compatible with our high-time-resolution scenario. However, real samples may pose problems. From the perspective of correlation methods, photoblinking, or any other reversible photophysical process, just creates an additional state with a different quantum yield. With an additional state, the complexity of the system and the required number of moments both increase. As a result, the quantity of data must be increased and/or the time resolution decreased, as described above. However, extra photoprocesses will not invalidate a correlation approach.

The most unusual feature of our high-time-resolution scenario is its very high signal level. Major improvements in optical-collection efficiency or detector quantum yield are unlikely, so very high excitation intensity would be required in a real experiment. High intensity will create rapid, permanent photobleaching, although the details are highly system dependent.¹²⁰⁻¹²³ However, so long as the bleaching is linear in the excitation intensity, the accessible time range simply shifts toward shorter times. High time-resolution experiments are feasible, at the expense of losing any slow dynamics. Similarly, linear heating can be dealt with by reducing the exposure time. In fact, any sample problem that is linear in the excitation intensity can be accommodated by appropriate experimental design.

Correlation methods do not require long time series. Measurements on multiple molecules are easily combined to increase the dataset size. As shown in detail in Paper 1,¹³ even the presence of relaxation times beyond the photobleaching time is not overly problematic.

We are left with the conclusion that nonlinear-excitation effects are the fundamental limitation on the time resolution of single-molecule experiments. We hope that by focusing attention on this issue, we can inspire better characterization of the nonlinear effects in specific samples and improved methods for minimizing them.

VI. COMPARISON TO OTHER METHODS

Some of the earliest, and still very popular, methods of single-molecule data analysis focus on removing noise from a noisy time series to reveal an ideal trajectory.⁷⁷⁻⁹⁰ The ideal trajectory is then used to construct histograms, dwell-time distributions, or other statistical quantities. Even with a sophisticated, change-point method, the data are initially binned to 10–30 photons/bin, rather than to $\frac{1}{2}$ photon/bin as in a correlation analysis.⁸⁰ Correlation methods require less binning by a factor of 40, with a proportional increase in the time resolution.

The rationale for this improvement is simple. Knowing the state of every molecule at every time demands a large amount of information from the data. However, in the end, we do not need the details of every realization of a stochastic process; we only need its average statistical properties. Correlation methods bypass determining these extra details and, thus, demand less of the data.

A place where an ideal trajectory might seem essential is in experiments that record multiple data channels and then perform a nonlinear calculation on them. For example, FRET (fluorescence resonant-energy transfer) records donor and acceptor channels and calculates a FRET efficiency from them.¹¹ Other experiments record the fluorescence intensity from two polarizations and use them to calculate the anisotropy.⁸ It might seem that noise-free signals are needed to do these nonlinear calculations.

However, one can envision extending the principles of correlation analysis to this problem. With multiple data series, one can calculate the entire set of nonlinear, auto- and cross-correlation functions. These can be related analytically to the correlation functions of the desired quantity, for example, to FRET-efficiency or anisotropy correlation functions. The nonlinear calculations are done on correlation functions, not time series. The techniques developed here could then be extended to extract the FRET-efficiency or anisotropy distributions and kinetics. Significant work needs to be done to demonstrate the feasibility of this idea, but one should not infer that correlation methods are fundamentally limited to single-channel experiments.

More recently, hidden-Markov models and other parametric models have become popular.^{37,38,40-51,86} We anticipate that these methods perform better than correlation methods, in

some cases. Because the range of solutions is systematically restricted in a parametric method, the range of acceptable solutions will be smaller, and the calculated uncertainty range of the results will be smaller. If the parametric model is based on reliable prior experiments, parametric methods allow that prior information to be incorporated into the analysis, improving its accuracy. On the other hand, if confidence in the model is low, or if it is chosen arbitrarily, potentially valid solutions are lost. The narrow uncertainty range calculated for the random error comes with the caveat that a systematic error may be undetected.

Even when a parametric method is desirable, existing methods typically use the large set of raw data points as their input and can become computationally intense.^{45,67} In these cases, a hybrid, parametric-correlation method may be useful. As discussed in Paper 1,¹³ one can also design a parametric analysis that uses the moments and moment-correlation functions as the target for fitting the model. (The delta-function models used to create Figs. 6 and 13 are examples. See the supplementary material, Sec. I.A.) Such an approach incorporates a well-established model into the analysis, while retaining the data-reduction features of the correlation approach.

Innovation in single-molecule analysis is ongoing; incorporating machine-learning techniques is a current example.⁸⁶⁻⁹⁰ Will a new method prove to be even more efficient than correlation methods? We do not expect so. The hierarchy of nonlinear and multidimensional-correlation functions defined in Eq. (2) is a complete description of a stochastic time series. The number of orders of this hierarchy accessible from a finite dataset is limited by sampling noise, an issue that will similarly limit any analysis method. The number of dimensions needed is limited to the number of hidden coordinates needed to reduce the system to Markovian dynamics. Again, the limitations are based on the nature of the system, not the method of analysis. Other approaches may be able to extract the same information and match the computational performance of our method, but it is hard to see how they could do significantly better.

The other feature that we can expect to be common to any type of analysis is the signal floor. It indicates a change in how noise averages and defines how much time binning is needed. The specific cases of additive, photon, and photon-counting noise have been analyzed in detail, but the unifying principle can also be seen. With additive noise, the signal floor occurs when the spread in the data from a single state (the noise width) exceeds the spacing between different states. The data distribution is exactly a convolution of the signal and noise. When a broad function is convolved with narrow one, there is a loss of information. When the noise and signal are correlated in magnitude, a convolution is not an exact description, but the same idea applies to some weighted measure of the noise width. When the noise is quantized, the literal noise width must be replaced by the quantum spacing, that is, by the noise width for an optimal signal level. In all cases, the signal floor occurs when the additional spread in the data exceeds the intrinsic spread in the signal. Thus, the existence of a signal floor is a general feature of noisy data, not of the analysis method.

Our estimates for the optimal time binning and the resulting estimates for the time resolution should remain applicable to any future analysis method.

VII. CONCLUSIONS

This paper has added another piece to a growing body of work that shows that times series can be completely analyzed using only correlation functions.^{13,22,23,27-29,31-33} For an experimental time series, the first task is removing the effects of detector noise. Whereas Paper 1 developed methods for removing simple, additive noise,¹³ this paper has extended those methods to more complex types of noise, including the noise encountered in single-molecule spectroscopy. Many noise-removal methods rely on the difference in correlation times of signal and noise to separate them, but correlation methods exploit this difference in a particularly direct and efficient manner.

The overall correlation analysis can be seen as a data-reduction process. In the first step, not only is the noise removed, but also the large set of raw data points is reduced to a smaller set of moments and moment-correlation functions. In the second step, these quantities are reduced to an even smaller set of physically relevant quantities—population distributions and transition rates. This second stage can also include correction of bias, in cases where the detector is not linear.

These final results are a reduced form of the data in the sense that a single model or assignment of states does not emerge. A range of solutions consistent with the data are generated by varying the regularizing conditions. The analyst retains the freedom to select from within this range to make a final interpretation of the data, based on their needs and their outside knowledge of the system. For example, they can decide after data reduction whether a peak in state space represents a single structure or a cluster of poorly resolved structures or whether a small rate constant means that the states are not kinetically connected at all. Model building comes after the data are analyzed; it is not a predicate to data analysis.

Experiments are not limited by noise alone; they are also limited by the finite size of the dataset. Thus, results in the limit of infinite amounts of data do not prove that a method is viable. Paper 1¹³ and this paper yield several insights into the effects of a finite dataset. First, with high-noise data, the inability to sufficiently sample the noise can be more limiting than the level of signal sampling. As a result, there is a floor to the signal level (or a ceiling to the noise level) beyond which it becomes impractical to collect enough data to compensate for a high noise level. Below this floor, the normal trend of converging with the square root of the number of data points does not apply. The common case of photon-counting always lies in this regime.

These ideas led to guidelines for the data quality (the number of data points and the signal-to-noise ratio) needed to recover the properties of a system. Very reasonably, the necessary data quality varies with the complexity of the system. These guidelines also lead to the conclusion about when time binning is needed and how wide the bins should be: the optimal signal level lies

near $\frac{1}{2}$ photon per bin. This condition then sets the time resolution possible under given experimental circumstances.

One can turn this problem around to ask what experimental circumstances will optimize the time resolution. For photon-counting experiments, signal levels in the strongly biased regime are needed to achieve an experimental time resolution approaching the detector's time resolution. The bias-correction methods demonstrated here offer a route to this high time resolution.

All these ideas have been tested against simulated data, which verifies the validity and accuracy of the methods with finite datasets. Needless to say, simulations cannot anticipate all the issues present in the many different samples that are of interest to single-molecule spectroscopy. We hope that by clarifying the issues involved, this paper will stimulate discussion and innovation around the experimental issues involved.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation – Chemical Measurement and Imaging program under Grant No. CHE-2003619 (with partial co-funding from the Chemical Structures, Dynamics, and Mechanisms – A program).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Mainak Dhar: Formal analysis (lead); Investigation (lead); Methodology (equal); Software (lead); Validation (lead); Visualization (lead); Writing – review & editing (equal). **Mark Berg:** Conceptualization (lead); Formal analysis (supporting); Methodology (equal); Project administration (lead); Supervision (lead); Writing – original draft (lead); Writing – review & editing (equal).

SUPPLEMENTARY MATERIAL

See the supplementary material for more detail on recovering the kinetics, for sampling-noise derivations, and for additional tables and figures.

DATA AVAILABILITY

The data that supports the findings of this study are available within the article and its supplementary material.

REFERENCES

- ¹ L. Onsager, "Reciprocal relations in irreversible processes. I," *Phys. Rev.* **37**, 405-426 (1931). <http://doi.org/10.1103/PhysRev.37.405>.
- ² W. J. Bruno, J. Yang, and J. E. Pearson, "Using independent open-to-closed transitions to simplify aggregated Markov models of ion channel gating kinetics," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6326-6331 (2005). <http://doi.org/10.1073/pnas.0409110102>.
- ³ B. J. Berne and R. Pecora, *Dynamic light scattering: With applications to chemistry, biology, and physics* (John Wiley & Sons, New York, 1976).
- ⁴ M. Sutton, "A review of x-ray intensity fluctuation spectroscopy," *C. R. Phys.* **9**, 657-667 (2008). <http://doi.org/10.1016/j.crhy.2007.04.008>.
- ⁵ F. Qin, "Principles of single-channel kinetic analysis," in *Patch-clamp methods and protocols*, edited by M. Martina and S. Taverna (Springer New York, New York, NY, 2014), pp. 371-399. http://doi.org/10.1007/978-1-4939-1096-0_23.
- ⁶ E. L. Elson, "Brief introduction to fluorescence correlation spectroscopy," in *Methods in enzymology*, edited by S. Y. Tetin (Academic Press, 2013) Vol. 518, pp. 11-41. <http://doi.org/10.1016/B978-0-12-388422-0.00002-9>.
- ⁷ C. Eggeling and C. Hellriegel, "Special issue on developments in fluorescence correlation spectroscopy and related techniques," *Methods* **140-141**, 1-222 (2018). <http://doi.org/10.1016/j.ymeth.2018.04.034>.
- ⁸ K. Paeng and L. J. Kaufman, "Single molecule rotational probing of supercooled liquids," *Chem. Soc. Rev.* **43**, 977-989 (2014). <http://doi.org/10.1039/c3cs60186b>.
- ⁹ J. Cui, A. P. Beyler, T. S. Bischof, M. W. B. Wilson, and M. G. Bawendi, "Deconstructing the photon stream from single nanocrystals: From binning to correlation," *Chem. Soc. Rev.* **43**, 1287-1310 (2014). <http://doi.org/10.1039/C3CS60330J>.
- ¹⁰ T. Kondo, W. J. Chen, and G. S. Schlau-Cohen, "Single-molecule fluorescence spectroscopy of photosynthetic systems," *Chem. Rev.* **117**, 860-898 (2017). <http://doi.org/10.1021/acs.chemrev.6b00195>.
- ¹¹ E. Lerner, A. Barth, J. Hendrix et al., "FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices," *eLife* **10**, e60416 (2021). <http://doi.org/10.7554/eLife.60416>.
- ¹² S. Adhikari and M. Orrit, "Progress and perspectives in single-molecule optical spectroscopy," *J. Chem. Phys.* **156**, 160903 (2022). <http://doi.org/10.1063/5.0087003>.
- ¹³ M. Dhar, J. A. Dickinson, and M. A. Berg, "Efficient, nonparametric removal of noise and recovery of probability distributions from time series using nonlinear-correlation functions: Additive noise," *J. Chem. Phys.* **159**, 054110 (2023). <http://doi.org/10.1063/5.0158199>.

- ¹⁴ R. Verberk and M. Orrit, "Photon statistics in the fluorescence of single molecules and nanocrystals: Correlation functions versus distributions of on- and off-times," *J. Chem. Phys.* **119**, 2214-2222 (2003). <http://doi.org/10.1063/1.1582848>.
- ¹⁵ M. Pelton, G. Smith, N. F. Scherer, and R. A. Marcus, "Evidence for a diffusion-controlled mechanism for fluorescence blinking of colloidal quantum dots," *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14249-14254 (2007). <http://doi.org/10.1073/pnas.0706164104>.
- ¹⁶ G. Hinze and T. Basché, "Statistical analysis of time resolved single molecule fluorescence data without time binning," *J. Chem. Phys.* **132**, 044509 (2010). <http://doi.org/10.1063/1.3303634>.
- ¹⁷ J. Houel, Q. T. Doan, T. Cajgfinger et al., "Autocorrelation analysis for the unbiased determination of power-law exponents in single-quantum-dot blinking," *ACS Nano* **9**, 886-893 (2015). <http://doi.org/10.1021/nn506598t>.
- ¹⁸ F. Abdollah-Nia, M. P. Gelfand, and A. Van Orden, "Artifact-free and detection-profile-independent higher-order fluorescence correlation spectroscopy for microsecond-resolved kinetics. 1. Multidetector and sub-binning approach," *J. Phys. Chem. B* **121**, 2373-2387 (2017). <http://doi.org/10.1021/acs.jpccb.7b00407>.
- ¹⁹ F. Abdollah-Nia, M. P. Gelfand, and A. Van Orden, "Artifact-free and detection-profile-independent higher-order fluorescence correlation spectroscopy for microsecond-resolved kinetics. 2. Mixtures and reactions," *J. Phys. Chem. B* **121**, 2388-2399 (2017). <http://doi.org/10.1021/acs.jpccb.7b00408>.
- ²⁰ M. A. Berg and J. R. Darvin, "Measuring the dynamics of a hidden coordinate: Rate exchange from 3D correlation functions," *J. Chem. Phys.* **145**, 054119 (2016). <http://doi.org/10.1063/1.4960186>.
- ²¹ M. A. Berg and H. Kaur, "Non-parametric analysis of nonexponential and multidimensional kinetics: I. Quantifying rate dispersion, heterogeneity and exchange," *J. Chem. Phys.* **146**, 054104 (2017). <http://doi.org/10.1063/1.4974508>.
- ²² S. D. Verma, S. A. Corcelli, and M. A. Berg, "Rate and amplitude heterogeneity in the solvation response of an ionic liquid," *J. Phys. Chem. Lett.* **7**, 504–508 (2016). <http://doi.org/10.1021/acs.jpcllett.5b02835>.
- ²³ S. R. Hodge, S. A. Corcelli, and M. A. Berg, "Nonlinear measurements of kinetics and generalized dynamical modes: II. Application to a simulation of solvation dynamics in an ionic liquid," *J. Chem. Phys.* **155**, 024123 (2021). <http://doi.org/10.1063/5.0053424>.
- ²⁴ H. Yang and X. S. Xie, "Statistical approaches for probing single-molecule dynamics photon-by-photon," *Chem. Phys.* **284**, 423-437 (2002). [http://doi.org/10.1016/S0301-0104\(02\)00672-9](http://doi.org/10.1016/S0301-0104(02)00672-9).

- ²⁵ W. K. Ridgeway, D. P. Millar, and J. R. Williamson, “The spectroscopic basis of fluorescence triple correlation spectroscopy,” *J. Phys. Chem. B* **116**, 1908-1919 (2012).
<http://doi.org/10.1021/jp208605z>.
- ²⁶ W. K. Ridgeway, D. P. Millar, and J. R. Williamson, “Quantitation of ten 30s ribosomal assembly intermediates using fluorescence triple correlation spectroscopy,” *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13614-13619 (2012). <http://doi.org/10.1073/pnas.1204620109>.
- ²⁷ S. D. Verma, D. A. Vanden Bout, and M. A. Berg, “When is a single molecule homogeneous? A multidimensional answer and its application to molecular rotation near the glass transition,” *J. Chem. Phys.* **143**, 024110 (2015). <http://doi.org/10.1063/1.4926463>.
- ²⁸ H. Kaur, S. D. Verma, K. Paeng, L. J. Kaufman, and M. A. Berg, “Biphasic rate exchange in supercooled *o*-terphenyl from an ensemble analysis of single-molecule data,” *Phys. Rev. E* **98**, 040603(R) (2018). <http://doi.org/10.1103/PhysRevE.98.040603>.
- ²⁹ C. Phelps, B. Israels, M. C. Marsh, P. H. von Hippel, and A. H. Marcus, “Using multiorder time-correlation functions (TCFs) to elucidate biomolecular reaction pathways from microsecond single-molecule fluorescence experiments,” *J. Phys. Chem. B* **120**, 13003–13016 (2016). <http://doi.org/10.1021/acs.jpcb.6b08449>.
- ³⁰ C. Phelps, B. Israels, D. Jose, M. C. Marsh, P. H. von Hippel, and A. H. Marcus, “Using microsecond single-molecule FRET to determine the assembly pathways of t4 ssDNA binding protein onto model DNA replication forks,” *Proc. Natl. Acad. Sci. U. S. A.* **114**, E3612-E3621 (2017). <http://doi.org/10.1073/pnas.1619819114>.
- ³¹ B. Israels, C. S. Albrecht, A. Dang, M. Barney, P. H. von Hippel, and A. H. Marcus, “Submillisecond conformational transitions of short single-stranded DNA lattices by photon correlation single-molecule förster resonance energy transfer,” *J. Phys. Chem. B* **125**, 9426-9440 (2021). <http://doi.org/10.1021/acs.jpcb.1c04119>.
- ³² H. Kaur and M. A. Berg, “Jump-precursor state emerges below the crossover temperature in supercooled *o*-terphenyl,” *Phys. Rev. E* **103**, L050601 (2021).
<http://doi.org/10.1103/PhysRevE.103.L050601>.
- ³³ S. R. Hodge and M. A. Berg, “Nonlinear measurements of kinetics and generalized dynamical modes: I. Extracting the one-dimensional green’s function from a time series,” *J. Chem. Phys.* **155**, 024122 (2021). <http://doi.org/10.1063/5.0053422>.
- ³⁴ *Single-molecule biophysics: Experiment and theory*, edited by T. Komatsuzaki, M. Kawakami, S. Takahashi, H. Yang, and R. J. Silbey (Wiley, Hoboken, New Jersey, 2012) Advances in chemical physics, Vol. 146. <http://doi.org/10.1002/9781118131374>.
- ³⁵ M. Tavakoli, J. N. Taylor, C.-B. Li, T. Komatsuzaki, and S. Pressé, “Single molecule data analysis: An introduction,” *Adv. Chem. Phys.* **162**, 205-305 (2017).
<http://doi.org/10.1002/9781119324560.ch4>.

- ³⁶ C. Du and S. C. Kou, “Statistical methodology in single-molecule experiments,” *Statist. Sci.* **35**, 75-91 (2020). <http://doi.org/10.1214/19-STS752>.
- ³⁷ J. B. Witkoskie and J. Cao, “Analysis of the entire sequence of a single photon experiment on a flavin protein,” *J. Phys. Chem. B* **112**, 5988-5996 (2008). <http://doi.org/10.1021/jp075980p>.
- ³⁸ K. R. Haas, H. Yang, and J.-W. Chu, “Expectation-maximization of the potential of mean force and diffusion coefficient in Langevin dynamics from single molecule FRET data photon by photon,” *J. Phys. Chem. B* **117**, 15591-15605 (2013). <http://doi.org/10.1021/jp405983d>.
- ³⁹ J. S. Bryan and S. Pressé, “Learning continuous potentials from smFRET,” *Biophys. J.* **122**, 433-441 (2023). <http://doi.org/10.1016/j.bpj.2022.11.2947>.
- ⁴⁰ 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**, 7454-7464 (2003). <http://doi.org/10.1021/jp035514+>.
- ⁴¹ S. A. McKinney, C. Joo, and T. Ha, “Analysis of single-molecule FRET trajectories using hidden Markov modeling,” *Biophys. J.* **91**, 1941-1951 (2006). <http://doi.org/10.1529/biophysj.106.082487>.
- ⁴² M. Hajdziona and A. Molski, “Maximum likelihood-based analysis of single-molecule photon arrival trajectories,” *J. Chem. Phys.* **134**, 054112 (2011). <http://doi.org/10.1063/1.3544494>.
- ⁴³ M. Pirchi, R. Tsukanov, R. Khamis, T. E. Tomov, Y. Berger, D. C. Khara, H. Volkov, G. Haran, and E. Nir, “Photon-by-photon hidden Markov model analysis for microsecond single-molecule FRET kinetics,” *J. Phys. Chem. B* **120**, 13065-13075 (2016). <http://doi.org/10.1021/acs.jpcb.6b10726>.
- ⁴⁴ B. E. Husic and V. S. Pande, “Markov state models: From an art to a science,” *J. Am. Chem. Soc.* **140**, 2386-2396 (2018). <http://doi.org/10.1021/jacs.7b12191>.
- ⁴⁵ J. Yoo, J.-Y. Kim, J. M. Louis, I. V. Gopich, and H. S. Chung, “Fast three-color single-molecule FRET using statistical inference,” *Nat. Commun.* **11**, 3336 (2020). <http://doi.org/10.1038/s41467-020-17149-w>.
- ⁴⁶ Z. Kilic, I. Sgouralis, W. Heo, K. Ishii, T. Tahara, and S. Pressé, “Extraction of rapid kinetics from smFRET measurements using integrative detectors,” *Cell Rep. Phys. Sci.* **2**, 100409 (2021). <http://doi.org/10.1016/j.xcrp.2021.100409>.
- ⁴⁷ P. D. Harris, A. Narducci, C. Gebhardt, T. Cordes, S. Weiss, and E. Lerner, “Multi-parameter photon-by-photon hidden Markov modeling,” *Nat. Commun.* **13**, 1000 (2022). <http://doi.org/10.1038/s41467-022-28632-x>.
- ⁴⁸ I. V. Gopich and H. S. Chung, “Theory and analysis of single-molecule FRET experiments,” in *Protein folding: Methods and protocols*, edited by V. Muñoz (Springer US, New York, NY, 2022), pp. 247-282. http://doi.org/10.1007/978-1-0716-1716-8_14.

- ⁴⁹ A. Barth, O. Opanasyuk, T.-O. Peulen, S. Felekyan, S. Kalinin, H. Sanabria, and C. A. M. Seidel, “Unraveling multi-state molecular dynamics in single-molecule FRET experiments. I. Theory of FRET-lines,” *J. Chem. Phys.* **156**, 141501 (2022).
<http://doi.org/10.1063/5.0089134>.
- ⁵⁰ O. Opanasyuk, A. Barth, T.-O. Peulen, S. Felekyan, S. Kalinin, H. Sanabria, and C. A. M. Seidel, “Unraveling multi-state molecular dynamics in single-molecule FRET experiments. II. Quantitative analysis of multi-state kinetic networks,” *J. Chem. Phys.* **157**, 031501 (2022).
<http://doi.org/10.1063/5.0095754>.
- ⁵¹ M. Götz, A. Barth, S. S. R. Bohr et al., “A blind benchmark of analysis tools to infer kinetic rate constants from single-molecule FRET trajectories,” *Nat. Commun.* **13**, 5402 (2022).
<http://doi.org/10.1038/s41467-022-33023-3>.
- ⁵² Keegan E. Hines, John R. Bankston, and Richard W. Aldrich, “Analyzing single-molecule time series via nonparametric Bayesian inference,” *Biophys. J.* **108**, 540-556 (2015).
<http://doi.org/10.1016/j.bpj.2014.12.016>.
- ⁵³ I. Sgouralis and S. Pressé, “An introduction to infinite HMMs for single-molecule data analysis,” *Biophys. J.* **112**, 2021-2029 (2017). <http://doi.org/10.1016/j.bpj.2017.04.027>.
- ⁵⁴ I. Sgouralis, S. Madaan, F. Djutanta, R. Kha, R. F. Hariadi, and S. Pressé, “A Bayesian nonparametric approach to single molecule förster resonance energy transfer,” *J. Phys. Chem. B* **123**, 675-688 (2019). <http://doi.org/10.1021/acs.jpcc.8b09752>.
- ⁵⁵ M. Tavakoli, S. Jazani, I. Sgouralis, O. M. Shafraz, S. Sivasankar, B. Donaphon, M. Levitus, and S. Pressé, “Pitching single-focus confocal data analysis one photon at a time with Bayesian nonparametrics,” *Phys. Rev. X* **10**, 011021 (2020).
<http://doi.org/10.1103/PhysRevX.10.011021>.
- ⁵⁶ A. Saurabh, M. Safar, M. Fazel, I. Sgouralis, and S. Pressé, “Single-photon smFRET: II. Application to continuous illumination,” *Biophys. Rep.* **3**, 100087 (2023).
<http://doi.org/10.1016/j.bpr.2022.100087>.
- ⁵⁷ A. Saurabh, M. Fazel, M. Safar, I. Sgouralis, and S. Pressé, “Single-photon smFRET. I: Theory and conceptual basis,” *Biophys. Rep.* **3**, 100089 (2023).
<http://doi.org/10.1016/j.bpr.2022.100089>.
- ⁵⁸ M. Safar, A. Saurabh, B. Sarkar, M. Fazel, K. Ishii, T. Tahara, I. Sgouralis, and S. Pressé, “Single-photon smFRET. III. Application to pulsed illumination,” *Biophys. Rep.* **2**, 100088 (2022). <http://doi.org/10.1016/j.bpr.2022.100088>.
- ⁵⁹ D. A. Koster, C. H. Wiggins, and N. H. Dekker, “Multiple events on single molecules: Unbiased estimation in single-molecule biophysics,” *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1750-1755 (2006). <http://doi.org/10.1073/pnas.0510509103>.

- ⁶⁰ S. Kalinin, E. Sisamakakis, S. W. Magennis, S. Felekyan, and C. A. M. Seidel, "On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits," *J. Phys. Chem. B* **114**, 6197-6206 (2010). <http://doi.org/10.1021/jp100025v>.
- ⁶¹ I. V. Gopich and A. Szabo, "Theory of single-molecule FRET efficiency histograms," *Adv. Chem. Phys.* **146**, 245-297 (2011). <http://doi.org/10.1002/9781118131374.ch10>.
- ⁶² Toma E. Tomov, R. Tsukanov, R. Masoud, M. Liber, N. Plavner, and E. Nir, "Disentangling subpopulations in single-molecule FRET and ALEX experiments with photon distribution analysis," *Biophys. J.* **102**, 1163-1173 (2012). <http://doi.org/10.1016/j.bpj.2011.11.4025>.
- ⁶³ J. Cao, "Event-averaged measurements of single-molecule kinetics," *Chem. Phys. Lett.* **327**, 38-44 (2000). [http://doi.org/10.1016/S0009-2614\(00\)00809-5](http://doi.org/10.1016/S0009-2614(00)00809-5).
- ⁶⁴ J. B. Witkoskie and J. Cao, "Testing for renewal and detailed balance violations in single-molecule blinking processes," *J. Phys. Chem. B* **110**, 19009-19017 (2006). <http://doi.org/10.1021/jp061471w>.
- ⁶⁵ J. Cao, "Correlations in single molecule photon statistics: Renewal indicator," *J. Phys. Chem. B* **110**, 19040-19043 (2006). <http://doi.org/10.1021/jp061302b>.
- ⁶⁶ I. V. Gopich and A. Szabo, "Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis," *J. Chem. Phys.* **124**, 154712 (2006). <http://doi.org/10.1063/1.2180770>.
- ⁶⁷ E. Oikonomou, T. Gruber, A. R. Chandra, S. Höller, C. Alzheimer, G. Wellein, and T. Huth, "2D-dwell-time analysis with simulations of ion-channel gating using high-performance computing," *Biophys. J.* **122**, 1287-1300 (2023). <http://doi.org/https://doi.org/10.1016/j.bpj.2023.02.023>.
- ⁶⁸ M. A. Ochoa, X. Zhou, P. Chen, and R. F. Loring, "Interpreting single turnover catalysis measurements with constrained mean dwell times," *J. Chem. Phys.* **135**, 174509 (2011). <http://doi.org/10.1063/1.3657855>.
- ⁶⁹ M. A. Ochoa, P. Chen, and R. F. Loring, "Single turnover measurements of nanoparticle catalysis analyzed with dwell time correlation functions and constrained mean dwell times," *J. Phys. Chem. C* **117**, 19074-19081 (2013). <http://doi.org/10.1021/jp4065246>.
- ⁷⁰ S. Yang and J. Cao, "Two-event echos in single-molecule kinetics: a signature of conformational fluctuations," *J. Phys. Chem. B* **105**, 6536-6549 (2001). <http://doi.org/10.1021/jp004349k>.
- ⁷¹ J. Cao and R. J. Silbey, "Generic schemes for single-molecule kinetics. 1: Self-consistent pathway solutions for renewal processes," *J. Phys. Chem. B* **112**, 12867-12880 (2008). <http://doi.org/10.1021/jp803347m>.

- ⁷² T. R. Avila, D. E. Piephoff, and J. Cao, "Generic schemes for single-molecule kinetics. 2: Information content of the poisson indicator," *J. Phys. Chem. B* **121**, 7750-7760 (2017).
<http://doi.org/10.1021/acs.jpcb.7b01516>.
- ⁷³ D. E. Piephoff and J. Cao, "Generic schemes for single-molecule kinetics. 3: Self-consistent pathway solutions for nonrenewal processes," *J. Phys. Chem. B* **122**, 4601-4610 (2018).
<http://doi.org/10.1021/acs.jpcb.7b10507>.
- ⁷⁴ L. P. Watkins and H. Yang, "Information bounds and optimal analysis of dynamic single molecule measurements," *Biophys. J.* **86**, 4015-4029 (2004).
<http://doi.org/10.1529/biophysj.103.037739>.
- ⁷⁵ D.-H. Lee, C.-T. Yuan, M. Tachiya, and J. Tang, "Influence of bin time and excitation intensity on fluorescence lifetime distribution and blinking statistics of single quantum dots," *Appl. Phys. Lett.* **95**, 163101-163103 (2009). <http://doi.org/10.1063/1.3236772>.
- ⁷⁶ C. H. Crouch, O. Sauter, X. Wu, R. Purcell, C. Querner, M. Drndic, and M. Pelton, "Facts and artifacts in the blinking statistics of semiconductor nanocrystals," *Nano Lett.* **10**, 1692-1698 (2010). <http://doi.org/10.1021/nl100030e>.
- ⁷⁷ P. A. Frantsuzov, S. Volkán-Kacsó, and B. Jankó, "Model of fluorescence intermittency of single colloidal semiconductor quantum dots using multiple recombination centers," *Phys. Rev. Lett.* **103**, 207402 (2009). <http://doi.org/10.1103/PhysRevLett.103.207402>.
- ⁷⁸ T. G. Terentyeva, H. Engelkamp, A. E. Rowan, T. Komatsuzaki, J. Hofkens, C.-B. Li, and K. Blank, "Dynamic disorder in single-enzyme experiments: Facts and artifacts," *ACS Nano* **6**, 346-354 (2011). <http://doi.org/10.1021/nn203669r>.
- ⁷⁹ J. N. Taylor, D. E. Makarov, and C. F. Landes, "Denoising single-molecule FRET trajectories with wavelets and Bayesian inference," *Biophys. J.* **98**, 164-173 (2010).
<http://doi.org/10.1016/j.bpj.2009.09.047>.
- ⁸⁰ 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).
<http://doi.org/10.1021/jp0467548>.
- ⁸¹ H. Yang, "Change-point localization and wavelet spectral analysis of single-molecule time series," *Adv. Chem. Phys.* **146**, 217-243 (2012). <http://doi.org/10.1002/9781118131374.ch9>.
- ⁸² D. L. Ensign and V. S. Pande, "Bayesian detection of intensity changes in single molecule and molecular dynamics trajectories," *J. Phys. Chem. B* **114**, 280-292 (2009).
<http://doi.org/10.1021/jp906786b>.
- ⁸³ B. Shuang, D. Cooper, J. N. Taylor, L. Kisley, J. Chen, W. Wang, C. B. Li, T. Komatsuzaki, and C. F. Landes, "Fast step transition and state identification (STaSI) for discrete single-molecule data analysis," *J. Phys. Chem. Lett.* **5**, 3157-3161 (2014).
<http://doi.org/10.1021/jz501435p>.

- ⁸⁴ J. N. Taylor, C.-B. Li, D. R. Cooper, C. F. Landes, and T. Komatsuzaki, “Error-based extraction of states and energy landscapes from experimental single-molecule time-series,” *Sci. Rep.* **5**, 9174 (2015). <http://doi.org/10.1038/srep09174>.
- ⁸⁵ Y. J. Bae, N. A. Gibson, T. X. Ding, A. P. Alivisatos, and S. R. Leone, “Understanding the bias introduced in quantum dot blinking using change point analysis,” *J. Phys. Chem. C* **120**, 29484-29490 (2016). <http://doi.org/10.1021/acs.jpcc.6b09780>.
- ⁸⁶ H. Li and H. Yang, “Statistical learning of discrete states in time series,” *J. Phys. Chem. B* **123**, 689-701 (2019). <http://doi.org/10.1021/acs.jpcc.8b10561>.
- ⁸⁷ J. Thomsen, M. B. Sletfjerding, S. B. Jensen, S. Stella, B. Paul, M. G. Malle, G. Montoya, T. C. Petersen, and N. S. Hatzakis, “DeepFRET, a software for rapid and automated single-molecule FRET data classification using deep learning,” *eLife* **9**, e60404 (2020). <http://doi.org/10.7554/eLife.60404>.
- ⁸⁸ J. Li, L. Zhang, A. Johnson-Buck, and N. G. Walter, “Automatic classification and segmentation of single-molecule fluorescence time traces with deep learning,” *Nat. Commun.* **11**, 5833 (2020). <http://doi.org/10.1038/s41467-020-19673-1>.
- ⁸⁹ D. S. White, M. P. Goldschen-Ohm, R. H. Goldsmith, and B. Chanda, “Top-down machine learning approach for high-throughput single-molecule analysis,” *eLife* **9**, e53357 (2020). <http://doi.org/10.7554/eLife.53357>.
- ⁹⁰ A. Bandyopadhyay and M. P. Goldschen-Ohm, “Unsupervised selection of optimal single-molecule time series idealization criterion,” *Biophys. J.* **120**, 4472-4483 (2021). <http://doi.org/10.1016/j.bpj.2021.08.045>.
- ⁹¹ C. Y. Lu and D. A. Vanden Bout, “Effect of finite trajectory length on the correlation function analysis of single molecule data,” *J. Chem. Phys.* **125**, 124701 (2006). <http://doi.org/10.1063/1.2352748>.
- ⁹² C.-Y. J. Wei, C.-Y. Lu, Y. H. Kim, and D. A. Vanden Bout, “Determining if a system is heterogeneous: The analysis of single molecule rotational correlation functions and their limitations,” *J. Fluoresc.* **17**, 797-804 (2007). <http://doi.org/10.1007/s10895-007-0234-9>.
- ⁹³ S. A. Mackowiak and L. J. Kaufman, “When the heterogeneous appears homogeneous: Discrepant measures of heterogeneity in single-molecule observables,” *J. Phys. Chem. Lett.* **2**, 438-442 (2011). <http://doi.org/10.1021/jz1016872>.
- ⁹⁴ X. S. Xie, “Single-molecule spectroscopy and dynamics at room temperature,” *Acc. Chem. Res.* **29**, 598-606 (1996). <http://doi.org/10.1021/ar950246m>.
- ⁹⁵ W. E. Moerner and M. Orrit, “Illuminating single molecules in condensed matter,” *Science* **283**, 1670-1676 (1999). <http://doi.org/10.1126/science.283.5408.1670>.

- ⁹⁶ W. P. Ambrose, P. M. Goodwin, J. H. Jett, A. Van Orden, J. H. Werner, and R. A. Keller, "Single molecule fluorescence spectroscopy at ambient temperature," *Chem. Rev.* **99**, 2929-2956 (1999). <http://doi.org/10.1021/cr980132z>.
- ⁹⁷ E. Barkai, Y. Jung, and R. Silbey, "Theory of single-molecule spectroscopy: Beyond the ensemble average," *Annu. Rev. Phys. Chem.* **55**, 457-507 (2004). <http://doi.org/10.1146/annurev.physchem.55.111803.143246>.
- ⁹⁸ H. Yang, "Detection and characterization of dynamical heterogeneity in an event series using wavelet correlation," *J. Chem. Phys.* **129**, 074701 (2008). <http://doi.org/10.1063/1.2969074>.
- ⁹⁹ F. Kulzer, T. Xia, and M. Orrit, "Single molecules as optical nanoprobe for soft and complex matter," *Angew. Chem. Int. Ed.* **49**, 854-866 (2010). <http://doi.org/10.1002/anie.200904858>.
- ¹⁰⁰ H. Yang and X. S. Xie, "Probing single-molecule dynamics photon by photon," *J. Chem. Phys.* **117**, 10965-10979 (2002). <http://doi.org/10.1063/1.1521154>.
- ¹⁰¹ H. S. Chung and I. V. Gopich, "Fast single-molecule FRET spectroscopy: Theory and experiment," *Phys. Chem. Chem. Phys.* **16**, 18644-18657 (2014). <http://doi.org/10.1039/C4CP02489C>.
- ¹⁰² J. Jiang, J. Jia, T. Zhao, K. Liang, R. Yang, and D. Han, "Recovery time of silicon photomultiplier with epitaxial quenching resistors," *Instruments* **1**, 5 (2017).
- ¹⁰³ F. Ceccarelli, G. Acconcia, A. Gulinatti, M. Ghioni, and I. Rech, "Fully integrated active quenching circuit driving custom-technology SPADs with 6.2-ns dead time," *IEEE Photonics Technol. Lett.* **31**, 102-105 (2019). <http://doi.org/10.1109/LPT.2018.2884740>.
- ¹⁰⁴ F. Severini, I. Cusini, D. Berretta, K. Pasquinelli, A. Incoronato, and F. Villa, "SPAD pixel with sub-ns dead-time for high-count rate applications," *IEEE J. Sel. Top. Quantum Electron.* **28**, 1-8 (2022). <http://doi.org/10.1109/JSTQE.2021.3124825>.
- ¹⁰⁵ S. Farina, I. Labanca, G. Acconcia, M. Ghioni, and I. Rech, "10-nanosecond dead time and low afterpulsing with a free-running reach-through single-photon avalanche diode," *Rev. Sci. Instrum.* **93** (2022). <http://doi.org/10.1063/5.0086312>.
- ¹⁰⁶ A. Giudici, G. Acconcia, I. Labanca, M. Ghioni, and I. Rech, "4 ns dead time with a fully integrated active quenching circuit driving a custom single photon avalanche diode," *Rev. Sci. Instrum.* **93** (2022). <http://doi.org/10.1063/5.0087341>.
- ¹⁰⁷ X. Li, D. Li, C. Tang, and S. Liu, "Fully-integrated SPAD-based receiver with nanosecond dead time for optical wireless communication," *Journal of Lightwave Technology* **41**, 653-661 (2023). <http://doi.org/10.1109/JLT.2022.3217075>.
- ¹⁰⁸ N. L. Johnson and S. Kotz, *Discrete distributions* (Houghton Mifflin, Boston, 1969).
- ¹⁰⁹ A. Stuart and J. K. Ord, *Kendall's advanced theory of statistics* (Oxford University Press, New York, 1987).

- ¹¹⁰ W. H. Press, S. A. Teukolsky, W. T. Vetterling, and P. P. Flannery, *Numerical recipes in C* (Cambridge University Press, Cambridge, 1992).
- ¹¹¹ A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola, *Numerical methods for the solution of ill-posed problems* (Kluwer Academic, Dordrecht, 1995).
<http://doi.org/10.1007/978-94-015-8480-7>.
- ¹¹² M. E. S. Arashi, M. Arashi, and B. M. G. Kibria, *Theory of ridge regression estimation with applications* (John Wiley & Sons, 2019). <http://doi.org/10.1002/9781118644478>.
- ¹¹³ R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B* **58**, 267-288 (1996). <http://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- ¹¹⁴ P. C. Hansen, "Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems," *Numer. Algorithms* **6**, 1-35 (1994). <http://doi.org/10.1007/BF02149761>.
- ¹¹⁵ T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, New York, 1991).
- ¹¹⁶ G. E. P. Box and G. C. Tiao, *Bayesian inference in statistical analysis* (Addison-Wesley Pub. Co., Reading, Mass, 1973).
- ¹¹⁷ *Excelitas SPCM-AQRH family datasheet* <https://www.excelitas.com/product/spcm-aqrh>, maintained by Excelitas Technologies, accessed 3/5/2024.
- ¹¹⁸ D. P. Ryan, M. K. Dunlap, M. P. Gelfand, J. H. Werner, A. K. Van Orden, and P. M. Goodwin, "A gain series method for accurate EMCCD calibration," *Sci. Rep.* **11**, 18348 (2021). <http://doi.org/10.1038/s41598-021-97759-6>.
- ¹¹⁹ R. Börner, D. Kowerko, M. C. A. S. Hadzic, S. L. B. König, M. Ritter, and R. K. O. Sigel, "Simulations of camera-based single-molecule fluorescence experiments," *PLOS ONE* **13**, e0195277 (2018). <http://doi.org/10.1371/journal.pone.0195277>.
- ¹²⁰ M. Haase, C. G. Hübner, F. Nolde, K. Müllen, and T. Basché, "Photoblinking and photobleaching of rylene diimide dyes," *Phys. Chem. Chem. Phys.* **13**, 1776-1785 (2011).
<http://doi.org/10.1039/C0CP01814G>.
- ¹²¹ H. Piwoński, A. Sokołowski, and J. Waluk, "In search for the best environment for single molecule studies: Photostability of single terrylenediimide molecules in various polymer matrices," *J. Phys. Chem. Lett.* **6**, 2477-2482 (2015).
<http://doi.org/10.1021/acs.jpcllett.5b01060>.
- ¹²² A. P. Demchenko, "Photobleaching of organic fluorophores: Quantitative characterization, mechanisms, protection," *Methods Appl. Fluoresc.* **8**, 022001 (2020).
<http://doi.org/10.1088/2050-6120/ab7365>.
- ¹²³ M. Hirano, R. Ando, S. Shimozone et al., "A highly photostable and bright green fluorescent protein," *Nature Biotechnology* **40**, 1132-1142 (2022). <http://doi.org/10.1038/s41587-022-01278-2>.

Supplementary Material for Efficient, nonparametric removal of noise and recovery of probability distributions from time series using nonlinear- correlation functions: Photon and photon-counting noise

Mainak Dhar and Mark A. Berg^{a)}

Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, USA

I. RECOVERING THE KINETICS

A. Simplified distribution

The general theory needed to recover kinetics from nonlinear correlation functions was presented in Ref. 33. This section presents the details of how the specific examples in Figs. 6 and 13 and Eq. (71) were generated. The state space was first simplified by eliminating the linewidths in the distributions shown in Figs. 5 and 12(a). A parametric model consisting of three delta functions was fit to the same set of moments used to generate the distributions. The probabilities were constrained to a sum of one, giving five parameters to be fit: three state positions and two population ratios. These parameters were fit to $N_\mu = 10$ or 13 moments, respectively, using Eq. (44). The zero-linewidth model gave good results: for the system of Fig. 5, with $\mathbf{P}_{\text{eq}} = \{0.409, 0.251, 0.340\}$ and $\mathbf{S} = \{0.125, 0.600, 1.125\}$, and for the system of Fig. 12, with $\mathbf{P}_{\text{eq}} = \{0.416, 0.248, 0.336\}$ and $\mathbf{S} = \{0.125, 0.608, 1.114\}$. These values can be compared to the input values, $\mathbf{P}_{\text{eq}} = \{0.417, 0.250, 0.333\}$ and $\mathbf{S} = \{0.125, 0.625, 1.125\}$ (Fig. 1).

B. Continuous formulation

The simplified distributions were used to calculate the normalized, orthogonal modes of the system $Q_n(S)$, as defined in Ref. 33. For a system with three discrete states, there are only three independent modes, $n = 0, 1, 2$. They are polynomial functions that depend on the first four moments of the distribution:

$$\begin{aligned} Q_0(S) &= 1 \\ Q_1(S) &= \left(\frac{S}{\sigma}\right) - \hat{\mu}_1 \\ Q_2(S) &= (\hat{\mu}_4 + 3\hat{\mu}_3 - 1)^{-1/2} \left[\left(\frac{S}{\sigma}\right)^2 - (\hat{\mu}_3 + 2\hat{\mu}_1) \left(\frac{S}{\sigma}\right) + \hat{\mu}_3\hat{\mu}_1 - 1 \right], \end{aligned} \quad (\text{S1})$$

where the standardized moments,

$$\begin{aligned}
 \sigma^2 &= \mu_2 - \mu_1^2 \\
 \hat{\mu}_1 &= \mu_1 / \sigma \\
 \hat{\mu}_3 &= \frac{1}{\sigma^3} \left(\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \right) \\
 \hat{\mu}_4 &= \frac{1}{\sigma^4} \left(\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4 \right),
 \end{aligned} \tag{S2}$$

are used. For consistency with the remainder of the calculation, the moments were calculated from the simplified distributions.

With the transformation from signal to modes defined, the mode-correlation matrix $\mathbf{C}(\tau)$ is defined by its elements

$$C_{mn}(\tau) = \langle Q_m(\tau) Q_n(0) \rangle. \tag{S3}$$

Using Eq. (S1), they can be calculate from the moment-correlation functions. For the current system,

$$C_{11}(\tau) = \hat{M}_{11}(\tau), \tag{S4}$$

$$C_{12}(\tau) = (\hat{\mu}_4 + 3\hat{\mu}_3 - 1)^{-1/2} \left[\hat{M}_{12}(\tau) - (\hat{\mu}_3 + 2\hat{\mu}_1) \hat{M}_{11}(\tau) \right], \tag{S5}$$

and

$$\begin{aligned}
 C_{22}(\tau) &= (\hat{\mu}_4 + 3\hat{\mu}_3 - 1)^{-1} \left[\hat{M}_{22}(\tau) - 2(\hat{\mu}_3 + 2\hat{\mu}_1) \hat{M}_{12}(\tau) \right. \\
 &\quad \left. + (\hat{\mu}_3 + 4\hat{\mu}_3\hat{\mu}_1 + 4\hat{\mu}_1^2) \hat{M}_{11}(\tau) \right].
 \end{aligned} \tag{S6}$$

where the standardized, moment-correlation functions,

$$\hat{M}_{mn}(\tau) = \frac{M'_{mn}[D](\tau) - \mu'_m \mu'_n}{\sigma^{m+n}}, \tag{S7}$$

are used. The simplification of the static distribution to three states lead to the simplification of the dynamics to these three functions. (Higher orders would be needed to test for dynamics between unresolved states within the linewidths.) The mode-correlation matrix has a simpler structure than the moment-correlation matrix: its off-diagonal elements all start at zero and decay to zero at long time. Results from our examples are shown in Fig. S1.

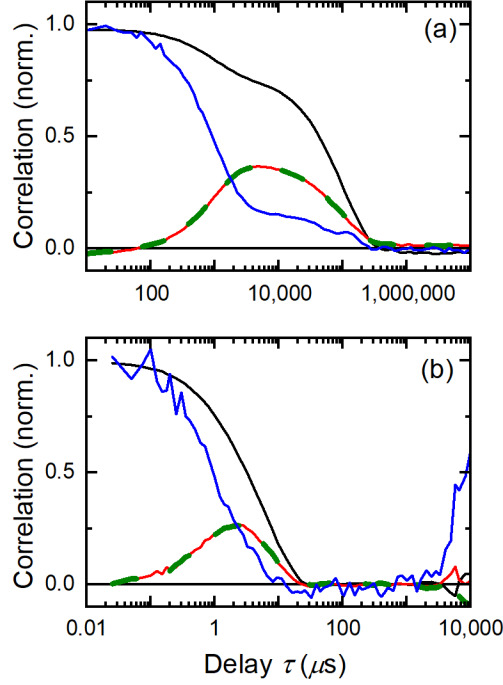


FIG. S1. Mode-correlation functions $C_{11}(\tau)$ (black), $C_{21}(\tau)$ (red), $C_{12}(\tau)$ (green dashed) and $C_{22}(\tau)$ (blue) from the data used in (a) Fig. 6 and (b) Fig. 13.

From Ref. 33, the Green's function is related to the mode-correlation functions by

$$G(S_1 | S_0) = P_{\text{eq}}(S_1) \sum_{n,m} Q_m(S_1) C_{mn}(\tau) Q_n(S_0). \quad (\text{S8})$$

Equation (4) then gives the time-dependent, joint-probability distribution.

C. Matrix formulation

For a discrete distribution, the equations in Sec. SI.B take on a matrix form. A population distribution is a vector $\mathbf{P}$. The values of the mode equations [Eq. (S1)] are only needed at the positions in the vector $\mathbf{S}$. Thus, we obtain a matrix $\mathbf{Q}$ with elements $Q_{mn} = Q_m(S_n)$. Then $\mathbf{QP}$ is a description of the population in terms of mode populations. That is, $\mathbf{Q}$ is a transformation matrix between signal and mode representations. For our two examples,

$$\mathbf{Q} = \begin{pmatrix} 1.00 & 1.00 & 1.00 \\ -1.07 & 0.04 & 1.26 \\ 0.56 & -1.73 & 0.61 \end{pmatrix} \quad (\text{S9})$$

for the data in Fig. 6, and

$$\mathbf{Q} = \begin{pmatrix} 1.00 & 1.00 & 1.00 \\ -1.11 & 0.28 & 1.17 \\ 0.40 & -1.72 & 0.78 \end{pmatrix} \quad (\text{S10})$$

for the data in Fig. 13. The differences are only due to the different sampling error in each dataset.

Continuing with a discrete-state system, Eq. (S8) reduces to a matrix equation for the time-dependent joint-probability distribution,

$$\begin{aligned} \mathbf{P}(\tau) &= \text{diag}(\mathbf{P}_{\text{eq}}) \mathbf{Q}^\top \mathbf{C}(\tau) \mathbf{Q} \text{diag}(\mathbf{P}_{\text{eq}}) \\ &= \left(\mathbf{P}_{\text{eq}} \otimes \mathbf{P}_{\text{eq}} \right) \odot \left(\mathbf{Q}^\top \mathbf{C}(\tau) \mathbf{Q} \right), \end{aligned} \quad (\text{S11})$$

where $\text{diag}(\mathbf{V})$ is the diagonal matrix with diagonal elements corresponding to the vector $\mathbf{V}$, $\otimes$ is the outer product and $\odot$ is the Hadamard (element-wise) product. Each element of the time-dependent-probability matrix $\mathbf{P}_{mn}(\tau)$ is the probability of starting in state n and ending in state m at a time τ later. Results have been shown Figs. 6(a) and 13(a).

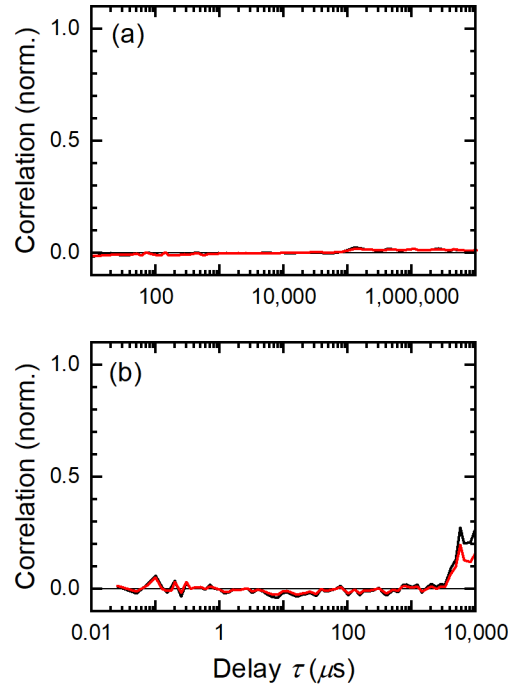


FIG. S2. Cross-correlation functions $E_{1,2}(\tau)$ (black) and $E_{2,1}(\tau)$ (red) from the correlation function in (a) Fig. S1(a) and (b) Fig. S1(b). The functions have been rotated to exactly zero at $\tau = 10.2$ ms and $\tau = 2.4$ μs , respectively. Compare to Figs. 6(i) and 13(i).

The matrix representation makes it straightforward to calculate a rotation matrix $\mathbf{A}$,

$$\mathbf{A}\mathbf{C}(\tau)\mathbf{A}^{-1} = \mathbf{E}(\tau), \quad (\text{S12})$$

such that $\mathbf{E}(\tau)$ is diagonal for a specific value of τ . If the system has eigenstates, $\mathbf{E}(\tau)$ will be diagonal at all other times, as well. Figure S2 shows the off-diagonal elements from our examples, and they are indeed zero, within experimental error. The diagonal elements of $\mathbf{E}(\tau)$ are then the eigendecays. These are shown in Figs. 6(i) and 13(i). The corresponding right eigenstates are represented by population vectors $\mathbf{R}_n$ that form the columns of a matrix given by

$$(\mathbf{R}_0, \mathbf{R}_1, \dots) = \mathbf{A}\mathbf{Q}\text{diag}(\mathbf{P}_{\text{eq}}). \quad (\text{S13})$$

These are shown in Figs. 6(f)–6(h) and 13(f)–13(h).

If the system is Markovian in the observable states, the eigendecays will be exponential,

$$\mathbf{E}_{n,n}(\tau) = e^{-\lambda_n \tau}, \quad (\text{S14})$$

with eigenvalues λ_n . These fits have been shown in Figs. 6(i) and 13(i). Alternative fits to stretched exponentials, $\exp[-(\lambda_n/\tau)^{\beta_{\text{st}}}]$ yield nearly exponential result ($1 - \beta_{\text{st}} < 0.015$). The vector of eigenvalues λ can be rotated to give the matrix of state-to-state rates,

$$\mathbf{k} = \mathbf{R}\text{diag}(\lambda)\mathbf{R}^{-1}. \quad (\text{S15})$$

An example is given in Eq. (71).

II. SAMPLING-NOISE DERIVATIONS

Starting with the general Eq. (52) and inserting the photon-noise condition [Eq. (54)] yields

$$\omega_{\varepsilon,n}^2 = \frac{1}{N_{\text{er}}} \sum_{k=1}^n \binom{n}{k} \left(\frac{1}{N_b T_{\varepsilon}} \right)^k \frac{\mu_{2n-k}}{\mu_n^2}. \quad (\text{S16})$$

The asymptotic approximation for the moments (Ref. 13, Eq. S34) gives

$$\frac{\mu_{2n-k}}{\mu_n^2} = \frac{2a}{(2b\mu_1)^k}, \quad (\text{S17})$$

where a and b are constants characteristic of the distribution. Inserting this result into Eq. (S16) and adding $k = 0$ to the summation leads to

$$\omega_{\varepsilon,n}^2 = \frac{2a}{N_{\text{er}}} \left[\sum_{k=0}^n \binom{n}{k} \left(\frac{1}{2bN_b T_{\varepsilon} \mu_1} \right)^k - 1 \right]. \quad (\text{S18})$$

Applying the binomial theorem and simplifying with Eqs. (25) and (28) results in Eq. (55).

To derive Eqs. (63) and (64), assume photon-counting data with a count rate of $\bar{N}_{\text{ct}}^0$. The probability that a time bin starts a sequence of n non-zero values is $(\bar{N}_{\text{ct}}^0)^n$. At low count rates, these sequences do not overlap, and the total number in the dataset is $N_{\text{er}}(\bar{N}_{\text{ct}}^0)^n$. The number of these sequences determines the value of the n th noise-corrected moment. Thus, the fractional sampling error in that moment [Eq. (49)] will be the inverse of the square root of this number,

$$\omega_n = \frac{1}{\sqrt{N_{\text{er}} (\bar{N}_{\text{ct}}^0)^n}}. \quad (\text{S19})$$

For the largest moment with acceptable error, $n = N_{\mu, \text{max}}$ and $\omega_n = \omega_{\varepsilon, \text{max}}$. Thus,

$$\omega_{\varepsilon, \text{max}} = \frac{1}{\sqrt{N_{\text{er}} (\bar{N}_{\text{ct}}^0)^{N_{\mu, \text{max}}}}}. \quad (\text{S20})$$

Solving for $N_{\mu, \text{max}}$ gives Eq. (63).

On the other hand, if we know that $N_{\mu, \text{min}}$ accurate moments are needed, then the number of photons collected, $N_{\text{ph}} = N_{\text{er}} N_{\text{ct}}^0$, must be large enough to ensure that $N_{\mu, \text{max}} > N_{\mu, \text{min}}$. Inserting these conditions into Eq. (S20) leads to Eq. (64).

III. ADDITIONAL FIGURE AND TABLES

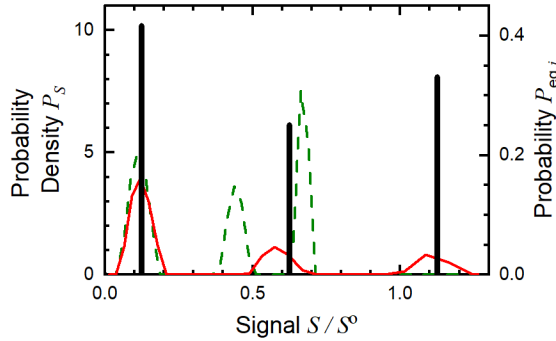


FIG. S3. Distributions recovered from the very-high-intensity data without transforming the prior. Before bias correction: green, dashed curve. After bias correction: red, solid curve. Input model: black bars. Compare to Fig. 10(c).

Table S1. Parameters used to generate the synthetic datasets used in the paper.

Dataset	S° (photon/ μ s)	T_e (μ s)	$\langle \bar{N}_{\text{ct}}^\circ \rangle$	$\langle S \rangle$ (photon/ms)	$\langle D \rangle$ (count/ms)	T_L (s)	N_{ph} (photons/ molecule)	N_b	$\langle \bar{N}_{\text{ct}} \rangle$	SNR	Figures
Medium-intensity, Gaussian-additive ^b	0.1 ^a	10	—	58.19 ^a	58.19 ^a	1000	—	1		0.58 ^b	2(a), 3(a–c), 5
Medium-intensity, photon-noise	0.1	10	0.583	58.26	58.26	1000	5.83×10^7	1	0.583	0.76	2(c), 3(g–i), 5, 6, 9(a), S1 (a), S2(a)
High-intensity, photon-noise	1	10	5.83	582.6	582.6	1000	5.83×10^8	1	5.83	2.41	2(b), 3(d–f), 4
Medium-intensity, photon-counting	0.1	0.1	0.00583	59.216	58.85	1000	5.92×10^7	1	0.00583	—	7
								1–400	various	—	8
								100	0.583	$\sim 0.76^c$	9(b)
								400	2.33	$\sim 1.5^c$	9(c), 10(a&d)
High-intensity, photon-counting	1	0.1	0.0583	592.16	565.9	1000	5.92×10^8	1	0.0583	—	7
								100	5.83	$\sim 2.4^c$	2(d), 9(d), 10(b&e)
Very-high- intensity, photon-counting	10	0.1	0.583	5,826	3,898	10	5.83×10^7	1	0.583	—	7, 10(c&f), S3
High-time- resolution, photon- counting ^d	40	0.025	0.583	23,300	15,484	0.020	4.70×10^5	1	0.583	—	11, 12, 13, S1(b), S2(b)

^a Substitute a continuous quantity, e.g., charge, for photon or count.^b $\sigma_e = S^\circ$.^c Calculated with the photon-noise formula, Eq. (28).^d The number of molecules $N_{\text{en}} = 10$, and total number of photons in the dataset is 4.70×10^6 .

Table S2: Regularization parameters for the distribution recoveries shown in the paper.
For all datasets, $S_{\min} = 0.0125$, $S_{\max} = 2.0125$, and $\alpha = 0$ (no sparsity regularization).

Dataset	N_μ	χ_0^a	β	χ/χ_0	Figures
Medium-intensity, Gaussian-additive	13	8.06×10^{-5}	1.40×10^{-3}	1.23	5
Medium-intensity, photon-noise	8	1.52×10^{-4}	2.22×10^{-3}	1.21	9(a)
	9	2.28×10^{-4}	1.80×10^{-3}	1.26	9(a)
	10	2.29×10^{-4}	3.00×10^{-3}	1.24	5, 9(a)
High-intensity photon-noise	15	1.69×10^{-4}	4.10×10^{-3}	1.27	4
Medium-intensity, photon-counting, $N_b = 100$	8	5.81×10^{-4}	1.10×10^{-2}	1.22	9(b)
	9	9.06×10^{-4}	6.70×10^{-2}	1.23	
	10	9.74×10^{-4}	4.49×10^{-2}	1.24	
Medium-intensity, photon-counting, $N_b = 400$	8	1.84×10^{-4}	2.70×10^{-3}	1.24	9(c)
	9	1.77×10^{-4}	2.70×10^{-3}	1.20	9(c)
	10	1.80×10^{-4}	1.80×10^{-3}	1.23	9(c), 10(a&d)
High-intensity, photon-counting	8	8.42×10^{-6}	1.11×10^{-4}	1.30	9(d)
	9	3.99×10^{-5}	3.25×10^{-4}	1.29	9(d)
	10	3.79×10^{-5}	3.25×10^{-4}	1.19	9(d)
	15	1.22×10^{-4}	2.20×10^{-3}	1.22	10(b&e)
Very-high- intensity, photon-counting	15	1.69×10^{-4}	2.20×10^{-3}	1.23	10(c&f)
		3.30×10^{-4}	7.40×10^{-3}	1.51	S3
High-time- resolution, photon- counting	13	6.54×10^{-4}	7.70×10^{-3}	1.12	12

^a $\beta = 0$.