Information Theoretical Approach to Single-Molecule Experimental Design and Interpretation

David S. Talaga*

Rutgers, The State University of New Jersey, New Brunswick, Department of Chemistry and Chemical Biology and BIOMAPS Institute, 610 Taylor Road, Piscataway, New Jersey 08854

Received: April 7, 2006; In Final Form: June 12, 2006

We use Shannon's definition of information to develop a theory to predict a photon-counting-based single-molecule experiment's ability to measure the desired property. We treat three phenomena that are commonly measured on single molecules: spectral fluctuations of a solvatochromic dye; assignment of the azimuthal dipole angle; determination of a distance by fluorescence resonant energy transfer using Förster's theory. We consider the effect of background and other "imperfections" on the measurement through the decrease in information.

1. Introduction

A. Single Molecules as State-to-Photon Transducers. Single-molecule measurements are rapidly changing the way scientists approach and understand complicated heterogeneous systems. Such a *system* typically includes three elements:

• a *molecule*, or assembly of several molecules, that is the focus.

• a reporter *dye*, the fluorescence of which is modulated by the molecule.

• the local *environment* surrounding the molecule(s) and fluorescent reporter(s).

From glasses,[1] to polymers,[2−4] to enzymes,[5−7] to nucleic acids,[8] to the folding of proteins,[9−11] to the infectious nature of viruses,[12] the paradigm of following the trajectory of single molecular systems has given new insight into systems that are problematic to study by conventional means.[3,13−17]

Single-molecule spectroscopy has revealed that the behavior of a complex system can be influenced by past events through persistent conformational changes, i.e., molecular memory. For example, individual enzymatic turnovers show memory effects where the reaction turnover rate is time-dependent.[6,7] Protein conformational fluctuations show evidence of subdiffusive dynamics.[18] Subdiffusion occurs when the kernel of the correlation function is not a δ function; that is, when the dynamics have memory and are non-Markovian at the given level of the dynamics description. Understanding of such effects requires being able to translate the signal stream from a single molecule into a state trajectory.

Single-molecule spectroscopy presents a new set of challenges and opportunities as compared to traditional bulk measurements. Sensitivity to rare events is a great benefit of single-molecule measurements. A consequence of this sensitivity is that the single-molecule luminescence measurements can be complicated by undesirable rare events such as intersystem crossing, photoionization, photoreduction, and photooxidation that result in the reporter chromophore undergoing transient or permanent passage to nonemitting states.[19] No information can be obtained about the system so long as the fluorescent reporter dwells in the nonemissive state. If the system later returns to an emissive state, then information can be obtained about the dark state of the fluorescent reporter based on the recovery statistics. However, the dynamics of the molecule to which it is coupled will remain hidden unless it is the dynamics of the molecule itself that cause the recovery.[20] For this reason, the system should be designed such that the molecular state can be determined from the photon stream before either the state of the molecule or the state of the reporter dye changes.

Sensitivity to unlikely events also means that even low-probability background events can obscure the signal. Methods to reduce background include confocal microscopy, near-field illumination of samples, and total internal reflection microscopy. Each method is amenable to observing photon streams. Each uses a high-numerical-aperture microscope objective for efficient light collection.

Single-molecule luminescence measurements require efficient detection of photons. Several detection schemes have been successfully employed.[21] These include the use of cameras based on charge-coupled devices (CCD) of various designs (e.g., front- or back-illuminated, intensified, electron multiplying/cascade). CCD signals can be related to statistics of photon emission; however, this is not always trivial. Avalanche photodiodes (APD) operated in Geiger mode have had widespread use in single-molecule measurements. Such devices have high quantum yields, low dark signal, and allow photon counting and timing within tens of picoseconds.[21]

Control of the molecule's environment can be characterized in two broad categories: diffusive and immobilized measurements. Diffusive measurements are limited by molecular diffusion, typically providing brief (100 μs to 10 ms) "snapshots" of the molecular state. Molecular dynamics can be extracted with careful analysis of many molecular transits,[22,23] making this method the equivalent of an extremely low concentration fluorescence correlation measurement. Immobilized measurements have the advantage of allowing long-time dynamics to be measured providing a true photon trajectory for each molecule measured.

Measurement of the single-molecule trajectory is the experimental equivalent of a molecular dynamics simulation but is of use only if the stochastic signal can be connected to the underlying and unobservable state of the system. The fluctuating

* E-mail talaga@rutgers.edu; URL http://talaga.rutgers.edu.

state trajectory is used to learn new information such as direct identification of the state space (including system heterogeneity),[24,25] the probability distribution function for the state space,[26] a potential of mean force,[27] a connectivity diagram connecting the states,[28] and exchange times between the states including memory effects.[28] These analyses require that the state of the molecule be determined with confidence.

The photon stream is, in effect, a noisy coded signal from the nonstationary source that is the thermally fluctuating molecular state. Decoding this signal to reproduce the molecular state is the goal of single-molecule data analysis schemes. As has been previously discussed,[28] the stochastic nature of classical spontaneous photon emission must be explicitly taken into account, or exchange rates between states can be overestimated because of misassigned states implying transitions that have not actually occurred.

There have been many approaches to this problem of distinguishing between states in the presence of what is often called "shot noise". These include thresholding either with or without the data first being filtered using ad hoc,[26,29] optimal,[27] or nonlinear[30] methods. Complete use of the information present in the data requires a photon-by-photon approach.[24,28,31] However, even a photon-by-photon approach will not always successfully assign states in the presence of shot noise. A full treatment using probability theory is needed to determine when a given approach to state assignment is feasible. Information theory will show whether state assignment is, in general, possible.

B. Information Theory Connects Signal and Inference. Probability theory is indispensable for making inferences in physical sciences. Rigorous adherence to its principles is particularly important for single-molecule measurements, since they are stochastic on several levels. The state of a single molecule may be fluctuating in a stochastic manner during the measurement. The property that directly modulates the signal, though presumably coupled to the state of the molecule, may also fluctuate independently of the state of the molecule. The signal itself is stochastic and consists of a stream of a randomly arriving photons. Random arrival of photons results in what is often called "shot noise". All of these coupled stochastic processes must be properly characterized in order to fully realize the potential of single-molecule measurements.

In the context of single-molecule spectroscopy, shot noise arises from inadequate sampling of a stochastic observable. The influence of shot noise on the variance of a parameter can be understood readily from a different definition of information due to Fisher using the Cramér–Rao inequality[32–34]

$$\langle (S - \bar{S})^2 \rangle_{\mathbb{Y}} \geq \mathscr{I}_f(Y)^{-1} = \left(\left\langle \left[\frac{d}{dY} \ln \mathscr{P}(O|Y) \right]^2 \right\rangle_{\mathbb{O}} \right)^{-1}$$

where $\mathscr{P}(O|Y)$ is the probability of the random observation, O, given the parameter to be estimated, Y. The angle brackets represent averaging over $\mathbb{O}$, the set of all possible observations, or $\mathbb{Y}$, the set of all possible parameter values. Fisher information considers the parameter to be single-valued and continuous with uncertainty in its estimation arising from only the stochastic nature of the measurement. In its standard formulation, Fisher information theory does not distinguish the contributions of "shot noise" broadening and the underlying width of the distribution. For single-molecule photon measurements, this is equivalent to considering the width of the distribution of the property being measured to be entirely due to "shot noise". A conceptual framework allowing separation of inherent distributions from stochastic sampling noise would be more flexible.
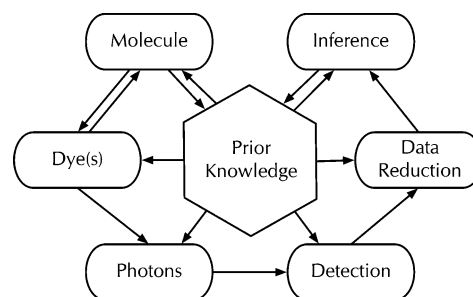

Figure 1. From an information theory point of view, the molecule *encodes information* into photons using the dyes as a transducer. The photons are converted into raw data by the detection apparatus and then *decoded* into a useful form by some data analysis procedure. From the reduced data, we draw inferences about the molecule on the basis of the data and our prior knowledge of the system.

Bayesian statistics and Shannon information as inspired by Jaynes[35,36] provide such a conceptual framework and lead to a unifying principle for single-molecule experimental design and interpretation. Jaynes used Shannon's theory of communication to develop an alternate formulation of statistical mechanics.[37] Jaynes combined this with Bayesian statistics to develop a consistent way to characterize scientific experiments.[36] The information being communicated in the experiment is a single quantity characterizing the experiment's quality. This quantity can, therefore, be optimized subject to the constraints of the system of interest and the measurement methods available to improve the quality of inference obtainable from the experiment.

Both Shannon's formulation of information and Bayesian statistics are nonparametric and consider both the states of the system and the photon measurements to have probability distributions. Furthermore, if one is interested in making inference between two (or more) states directly from the photon stream, rather than indirectly through a parameter, Shannon's formulation of information is especially useful. Shannon's definition of information provides a more straightforward interpretation of how multiple sources of information serve to determine knowledge of the state of a system. Practical modern experiments result in digital information which is immediately amenable to analysis with Shannon's Information. Shannon information theory is readily applied to both discrete and continuous distributions of state-spaces. Finally, under this formulation, entropy has a unique definition and a direct connection to statistical thermodynamics.

Viewed as a coding problem—where the molecule encodes its state into the noisy channel of the photon stream—the inference of the molecular state from the photon stream becomes an application of Shannon information theory.[36,38] The experiment opens a communication channel between the molecule and the investigator. Figure 1 illustrates an abstraction of this process where information flows counterclockwise from the molecule to the final inference through the experiment. Experimental design uses prior knowledge of the system (obtained through control experiments) to optimize the elements of the communication channel that are under the investigator's control such that the final inference will most closely resemble the state of the molecule that is being encoded in the experiment.

Shannon information theory provides a consistent and rigorous way to evaluate the fundamental limit of our ability to distinguish between events on the basis of our measurements. The amount of information in any conclusion that we draw from the data arriving from a single molecule cannot exceed the fraction of information in the data that came from the single molecule. Data processing may decrease, but it cannot increase,
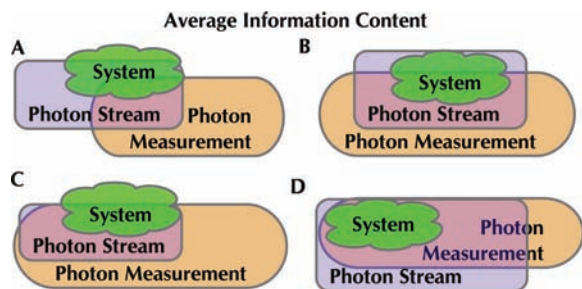
Single-Molecule Design and Interpretation

J. Phys. Chem. A, Vol. 110, No. 31, 2006 **9745**

Figure 2. Illustration of four abstract experiments where area represents information. The overlapping areas represent joint information. (A) The photon stream misses relevant information about the system. The photon measurement stores information (e.g., excessive resolution) that is not present in the photon stream yet misses relevant information that is present. (B) The photon stream now includes all the relevant information from the molecule. There is still irrelevant information stored, and there is relevant information in the photon stream not being measured. (C) The measurement records all the relevant information in the photon stream. However, there is not enough information in the photon stream to completely determine the molecular state. (D) The state information can be completely determined to arbitrary precision from the photon stream, and all relevant information in the photon stream is recorded in the experiment.

this information. If the information in the data cannot distinguish between the events of interest, then the investigator must either improve the measurements or accept that the technique is inadequate to address the problem. As applied to single-molecule spectroscopy, information theory allows us to determine if the stream of photons will provide enough information about the system to determine its properties within the above limitations; it provides a framework for interpreting and quantifying the uncertainty in single-molecule measurements.

Entropy ($\mathscr{H}$) is the expectation value of information

$$\mathscr{H}(\mathbb{O}) = -\sum_{k=1}^{n} \mathscr{P}(O_k) \log_2 \mathscr{P}(O_k) \qquad (1)$$

$\mathbb{O} = \{O_1, \cdots, O_m\}$ represents all possible outcomes of the experiment. The *conditional entropy*, $\mathscr{H}(\mathbb{S}|\mathbb{O})$ is the information expected to be still undetermined in the state of the system, $\mathbb{S} = \{S_1, \cdots, S_n\}$, once an observation $\mathbb{O}$ has been made. In Figure 2, this represents the area of the system that does not overlap with the photon measurement. In a good experiment, this area can be made arbitrarily small as in Figure 2D.

The average amount of information that the measurement $\mathbb{O}$ conveys about the system $\mathbb{S}$ is the mutual information, $\mathscr{T}(\mathbb{S}, \mathbb{O})$. In Figure 2, it is the area of overlap between the system and the photon measurement. The amount of information that the observation communicates about the system is equal to the decrease in system entropy that occurs as a result of the measurement.

$$\mathscr{T}(\mathbb{S}, \mathbb{O}) = \mathscr{H}(\mathbb{S}) - \mathscr{H}(\mathbb{S}|\mathbb{O}) \qquad (2)$$

This relationship has particular relevance for our analysis. It is the quantitative expression of the idea that the amount of information delivered by an experiment is the difference between the uncertainty before and after the observation is made.

C. Purpose of this Paper. The present work is motivated by a desire to be able to know prior to performing an experiment whether that experiment, as designed, can provide enough information to properly classify photons by state and therefore to be able to characterize state dwell time distributions and be able to reconstruct the hidden molecular-state trajectory from the photon stream.

We recently discussed a method by which molecular kinetic parameters may be robustly estimated from single-molecule photon arrival time trajectories.[28] For the case of common two-color experiments, we found that reasonable estimates could be obtained even in the limit where the kinetic rates were comparable to or faster than the mean interphoton time for the experiment. Estimation of kinetic parameters remained robust even when the classification of photons by hidden state became unreliable. This method, based on the application of hidden Markov models to the photon-timing trajectory, also provided a way to statistically decide between competing kinetic models. The particular strength of the hidden Markov model is its ability to use all the valid information before and after the observation of a particular photon.

We now examine the fundamental problem of state assignment in single-molecule spectroscopy from measurement of photons. In this paper, we only consider assignment of states from a distribution that is not exchanging on the time scale of the measurement. If a molecule can change states during the observation, the temporal sequence of states (the trajectory) must be reconstructed from the data. The information theoretical treatment of full trajectory reconstruction and model selection will be the topic of a future paper.

We start with a discussion of how to apply information theory to photon counting and photon timing experiments. We will then analyze the information required to make inferences from three different types of single-molecule spectroscopic measurements. Our examples include both discrete and psuedo-continuous state spaces. By pseudo-continuous, we mean that while the real variable is continuous in nature the computer representation is discrete. The computer representation can be made to arbitrary precision and is chosen such that the uncertainty in the measurement exceeds the digital precision of the discrete representation.

In the first example, we will treat the resolution of two discrete states by a spectral measurement. We will then examine the loss of information that results from reducing the spectral resolution until we are left with the classic two-color measurement. This is a common starting point for the interpretation of equilibrium dynamics of single molecules.

The second example will treat the assignment of a pseudo-continuous variable for the determination of the azimuthal angle of a radiating fixed dipole from a finite number of photons detected with polarization sensitivity.

In the third example, we will treat the assignment of a pseudo-continuous variable that we will call a distance for the purposes of our analysis. It is a common goal in single-molecule spectroscopy to attempt to measure a fluctuating distance via the distance dependence of fluorescence resonant energy transfer (FRET). In this treatment, we will not include the use of lifetime information.

We examine the effects that experimental limitations will have on the information. We include the effects of background and detector "cross-talk". For the case of FRET, we also include the effects of direct excitation of the acceptor and the presence of photobleached donor and acceptors. We will assume in our analyses that a reasonable number of simple bulk or bulk-equivalent measurements have been performed to provide basic information (prior knowledge in Figure 1) regarding the changes that can occur in the system of interest.

Specifics of our information theory notation, definitions, and derivations are in the Appendix.

II. Methods

Numerical solutions were obtained using Mathematica 5.0 (Wolfram Research) and Igor Pro 5.04b (Wavemetrics) using custom-written functions and procedures. Discrete inequality equations were solved by a robust binary search method also programmed in Mathematica.

III. Results

A. Characterizing Photon Streams and Experiments with Information Theory. Information theory can measure the efficiency of an experiment. A perfectly efficient experiment would record the information available in the photon stream that conveys knowledge of the state of the system and nothing else. In that case, the resulting data contains exactly the amount of information that arrived to the observer from the system. More commonly, experiments are inefficient and record phenomena at significantly higher resolution than is necessary to completely describe the phenomena present (Figure 2D) In general, experimental resolution can be increased arbitrarily (and unnecessarily) to the bounds of the uncertainty principle and the transform limit. Experiments can also be lossy in that they discard information, relevant or irrelevant, that is present in the phenomena being measured (Figure 2A–C). These losses can occur before the photons have been emitted (Figure 2A,C) or at point of detection (Figure 2A,B). Information theory can establish the amount of information required to fully describe the phenomenon without loss and without redundancy or irrelevance. In this section, we consider the amount of information delivered from a single-molecule photon stream and analyze some common experiments used to measure photon streams so as to evaluate the loss of information, if any, in the methods commonly implemented to study single molecules.

Since there are multiple observable properties per photon, the total amount of information in photon streams can be enormous. Experimental photon streams from single-molecule systems contain substantially less information than this maximum. The physical properties of dyes appropriate for single-molecule fluorescence measurements greatly reduce the expected experimental information. Moreover, the information recorded from a photon stream is typically a small fraction of the total information that is theoretically available from it. Laboratory measurements can determine the interphoton times $\mathbb{T}$, the photon frequencies $\mathbb{F}$ (or equivalently, wavelength), and polarization $\mathbb{P}$ of the photon. These measurements can be made independently to the extent that the transform limit allows.

$$\mathscr{P}(t_i, f_j, P_k) = \mathscr{P}(t_i)\mathscr{P}(f_j)\mathscr{P}(P_k)$$

The spatial pattern of emission contains information about the colatitude orientation of the emission dipole. Since current experimental methods do not typically resolve the emission direction of the photon, we will omit this information from our analysis.

If N_t, N_f, and N_P represent the number of possible outcomes of $\mathbb{T}$, $\mathbb{F}$, and $\mathbb{P}$ that are distinguishable by a practical measurement, then the total entropy (expected information) of a stream of n photons with all outcomes equally likely is

$$\mathscr{H}(\mathbb{T}, \mathbb{F}, \mathbb{P}|n) = n[\log_2(N_t \times N_f \times N_P)]$$

The combined temporal and spectral resolution is restricted by the transform limit to $\Delta t \times \Delta f \gtrsim (2\pi)^{-1}$. If the spectral range of interest is f_{tot} and the arrival time range of interest is t_{tot},
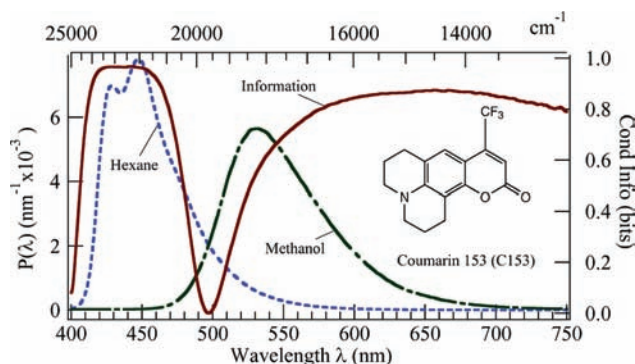

Figure 3. The probability-normalized ($\sum_{\lambda}\mathscr{P}(O_{\lambda}|S_i) = 1$) fluorescence spectra of C153 in hexane (blue) and in methanol (green). The mutual information in bits between the state (polar vs nonpolar) and each photon emitted (red).
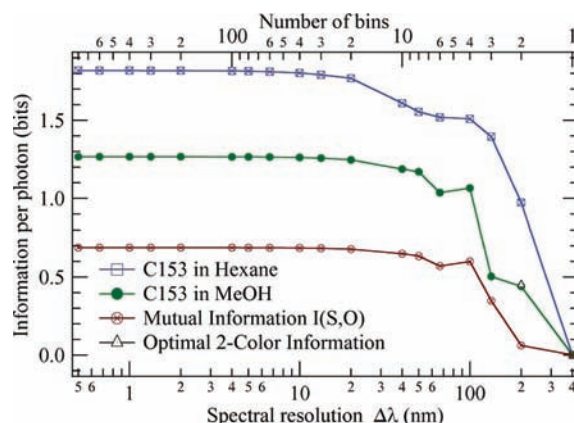

Figure 4. The expected spectral information vs resolution (and number of bins) for C153 in hexane (solid circles) and in methanol (squares). The mutual information between the photon observation and the two-color system of C153 in either a hexane or methanol environment (crossed circles). The mutual information for an *optimized* two-color experiment with two bins is marked with an open triangle.

then the combined information available from these observables is

$$-\log_2 \Delta f/f_{tot} - \log_2 \Delta t/t_{tot} = \log_2(2\pi f_{tot}t_{tot})$$

For a typical visible range of 400–700 nm and total experimental time of 100 seconds, it would take 57 bits per photon to record all the information present in the uncertainty-principle limited stream.

The physical limitations of organic dyes prevent this limit from being reached. A typical dye used for single-molecule measurements emits visible or near-infrared photons. Fluorescence lifetimes for these dyes are usually in the nanoseconds. So measurement of a dye emitting at ~500 nm (600 THz) with a ~5 ns lifetime is restricted by the transform limit to a spectral resolution of 2 ppm. Coverage of the 400–700 nm spectral range of coumarin 153[39] (C153, inset of Figure 3) at this resolution would require 19 bits of spectral information. However, this limit will not be relevant at room temperatures. The lack of spectral structure (see Figure 3) in commonly used dyes substantially reduces the average amount of information present. Figure 4 shows the dependence of the mutual information per photon from C153 in two different solvents as a function of the resolution. The hexane spectrum has a higher initial entropy than the methanol spectrum because it is sharper. The hexane spectrum loses information first as the resolution decreases because of the vibrational structure present on the

Single-Molecule Design and Interpretation

J. Phys. Chem. A, Vol. 110, No. 31, 2006 **9747**

spectrum. The methanol spectrum loses a larger fraction of its information when reduced to two channels because it is broader than the hexane spectrum. The dependence of information on spectral widths quantifies our intuition that more dyes (or quantum dots) can be distinguished if their spectra are narrower.

The amount of information for a specific spectrum as the resolution becomes low depends on where the bins are placed. For this example, they were evenly spaced across the spectral range (400−800 nm). This is the source of the non-monotonic parts of the curve. The information approaches an asymptotic limit as the resolution is increased. This limit is effectively reached when the resolution is 10 nm. For the entire spectral range, this gives a required amount of information to be recorded of 5.3 bits, or about 40 spectral channels. The information for the range required for a single dye environment is closer to 3 bits. This is a substantial reduction in the available information at the transform limit. This also implies that there is not much to be gained from measuring a high-resolution spectrum of a typical single-molecule dye at room temperature. This should not be surprising when one considers the lack of structure on the band. In section 3.2, we consider the information available from making a measurement on a dye that can undergo spectral fluctuations. We will come to similar conclusions there.

Note that the asymptotic limit in Figure 4 is the mutual information between a single photon and the spectrum. The amount of information that must be recorded in terms of spectral resolution is dependent on the information in the spectrum. Observation of increasing numbers of photons will asymptotically approach the information present in the spectrum.

If the dye randomly emits with a constant rate, then the time between photons will be exponentially distributed, $\mathscr{P}(t = i\,\delta t) = (1 - e^{-k\,\delta t})\,e^{-k\,i\,\delta t}$, with δt, the temporal resolution of the measurement, and k, the emission rate, then the total expected information (entropy) per photon is

$$\mathscr{H}(\mathbb{T}) = \frac{k\,\delta t}{(1 - e^{-k\,\delta t})\ln(2)} - \log_2(e^{k\,\delta t} - 1) \approx$$
$$\frac{1 + k\,\delta t}{\ln(2)} - \log_2(k\,\delta t) \quad (3)$$

The observed emission rate depends on the excitation rate, and as the excitation rate increases, it will ultimately be limited by the finite fluorescence lifetime (e.g., 5 ns). With a 10 nm effective spectral resolution, the highest temporal resolution possible would be ∼13 fs. Putting these numbers into eq 3 gives 20 bits. This gives a maximum of ∼25 bits of spectral and temporal information.

The polarization of a photon can be described in terms of the relative phase of the Cartesian components of the electric field with respect to the propagation direction and their mutual orientation relative to the laboratory frame. This implies two angles that run from 0 to π. With modern instrumentation, the polarization of a light source can be characterized to at least one part in 10^5 for each angle giving a total information content of 33 bits. This quantity of information is only relevant, however, for a stable continuous light source. Measurements of this precision require many more than a single photon. In measuring a single photon, a single bit of information regarding the polarization can be recorded if the photon passed though a polarizing beam splitter. This is because once the photon has passed through the beam splitter it has a polarization state that is determined probabilistically on the basis of the polarization angle of the incident light. The reduction in entropy of the unknown polarization of the light source can be no more than

a single bit. $\mathscr{H}(\mathbb{O}) = 1$. The entropy of linearly polarized light of unknown polarization angle ϕ is

$$\mathscr{H}(\mathbb{S}_\phi) = -\int_0^\pi \frac{1}{\pi} \log_2 \frac{1}{\pi}\,\mathrm{d}\phi = \log_2 \pi \quad (4)$$

where we have replaced the summation in eq 1 with an integral because of the continuous nature of the variable. (Picking an arbitrary resolution for a discrete summation gives the same result for the information delivered by the photon. An increment of $\pi/8$ gives the information to within 1%.) The mutual information that is delivered per photon regarding the polarization angle is

$$\mathscr{I}(\mathbb{S}_\phi, \mathbb{O}) = \mathscr{H}(\mathbb{S}_\phi) - \mathscr{H}(\mathbb{S}_\phi | \mathbb{O}) = \frac{1}{\ln(2)} - 1 \quad (5)$$

Thus, a single photon delivers 0.44 bits of information regarding the polarization state of the light source.

In total, there is a minimum of 58 bits of information available per photon that could potentially be recorded from an uncertainty-principle limited source. The use of organic dyes reduces this total amount of information to a minimum of 26 bits. An experiment that stores less than this number of bits per photon is losing information that is potentially available from the photon stream. However, the amount of information that is actually relevant to the hidden state of the system is necessarily less than this number. As a result, experiments that appear to be throwing away much of the information present in the photon stream may, in fact, be capturing the majority of the information that is available *about the system.*

Current state-of-the-art single-molecule fluorescence measurements can record much of the information present in the photon stream. The experiments that currently generate the most information from single molecules are single-photon timing experiments with pulsed lasers using time-correlated single-photon counting. These experiments record the time elapsed between the laser excitation and the arrival of the photon with as much as 12 bits of information dedicated to the excited state lifetime of the dye. The coarse arrival time of photons is also recorded with a resolution that depends on the repetition rate of the laser. For a 100 MHz laser and a 100 s experiment, this corresponds to 33 bits. These 2 pieces of information locate the photon with temporal resolution limited by the instrument response time of avalanche photodiodes to 20 ps. However, 45 bits is more information than is actually being delivered by the photons. If only the interphoton times are recorded, then the amount of information required is limited by the detector dark counts ∼50 Hz to 31 bits. The maximum count rate is limited by the detector dead time of ∼50 ns to 20 MHz. Physical limitations of light collection and detection quantum yield further reduce this to ≤5% of the laser repetition rate or ≤5 MHz. Using these limiting numbers in eq 3 gives 14.7 bits.

These experiments can also record the data from several detectors. A pair of 10×1 avalanche photodiode arrays could provide another 4.3 bits of information. This would typically be 1 bit of polarization information and 3.3 bits of spectral information.

Time-correlated single-photon counting loses information from two sources. The temporal resolution eliminates the ability to distinguish excited-state lifetimes shorter than ∼20 ps. Only a small fraction of the photons generated by the molecule are actually collected and detected by a typical experiment. Nevertheless, the maximum information transfer rate, or bandwidth, of a single-molecule experiment is ∼100 Mbit/s.

TABLE 1: Total Information Available from an Arbitrary Photon Stream, a Single Dye Photon Stream, Recorded in a Typical Single-Molecule Measurement[a]

observable	$\mathcal{S}_{h\nu}^{*}$	$\mathcal{S}_{dye}^{*}$	$\mathcal{S}_{rec}^{*}$
T	b	20.	33.
K		20.	12.
F	$57. - b$	5.3	3.3
P	0.44	0.44	1.0
total	58.	26.	50.

[a] Contributions from the interphoton time (intensity) T, the fluorescence rate (lifetime), K, the photon frequency (color), F, and the polarization, P, are all listed by their contributions to the total information. $\mathcal{S}_{h\nu}^{*}$ is the maximum information per photon for an arbitrary photon stream where b indicates tradeoff in resolution due to the transform limit. $\mathcal{S}_{dye}^{*}$ is the maximum information per photon for a photon stream arising from a typical dye that would be appropriate for a single-molecule measurement. $\mathcal{S}_{rec}^{*}$ is the maximum information per photon that can be typically recorded by current instrumentation.

However, in the following section, we will show that much of these data are not carrying information regarding the system. Most of it is, in fact, what is often called "shot noise".

B. State Information from Spectral Measurements. Spectrally resolved measurements represent one of the most common types of single-molecule experiments. Assignment of the molecular state is done by measuring the relative intensities of two or more spectral channels. We will illustrate the principle of using information theory to analyze single molecules using the simple case of a two-state system where each state produces photons that are detected in separate channels. A simple physical example of this would be an environmentally sensitive dye that can exist in two different local environments causing a spectral shift in the fluorescence spectrum. Another example where this analysis would be appropriate is a FRET colocalization experiment where the localized state is well inside of the Förster radius.[40,41] Information theory will allow us to determine the number of photons that will be required, on average, to distinguish the two states on the basis of their spectra. To ground our ideas, we will treat the specific example of C153 potentially existing in two environments: hydrophobic and hydrophilic.

In a spectrally resolved experiment, the amount of information delivered per photon will be wavelength-dependent. An examination of the two spectra in Figure 3 suggests that, while photons of wavelengths far from the overlap region will readily distinguish the two states, photons in the overlap region will not. This can be quantified with information theory. Evaluating the conditional entropy term in the mutual information expression over just the i states in eq A12 for individual photons observed at each wavelength, λ, gives the red curve in the top panel of Figure 3. Evaluating the sum over all wavelengths gives the mutual information for all wavelengths, which, for this example, is 0.69 bits. This level of information is the equivalent of being 94% confident in the state of the system based on a single photon (in the absence of background).

As noted in section A, the smooth spectra of organic dyes substantially reduce the average amount of information available from spectrally resolving the photon stream. Figure 4 shows the effect that reducing the number of wavelength channels has on the mutual information, $\mathcal{S}(\mathbb{O}_\lambda, \mathbb{S})$, between the stream of single photons and the state of the system. The loss of information is negligible until the spectral resolution has been reduced to 100 nm. This suggests that an experiment with only modest number of spectral channels will perform nearly as well as one at high spectral resolution when trying to distinguish the two states. As the number of bins becomes smaller, their placement becomes more important. For example, when two
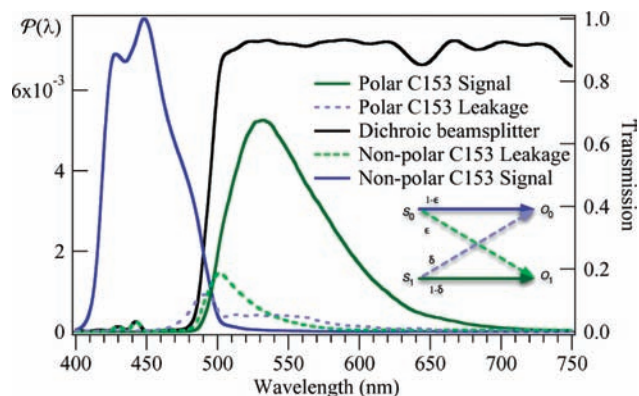

Figure 5. Cross-talk arises from spectral overlap and nonideal dichroic beam splitters. The inset illustrates the relationship between the leakage parameters, δ and ϵ, and the actual cross-talk between channels.

bins are left, the information is only 0.061 bits. However, in this case the bin boundary was located at the very nonoptimal location of 600 nm. The last evenly split resolution in this analysis with a bin boundary at ~500 nm was 100 nm or 4 bins. This is why the information increases slightly from 6 bins to 4 bins. The information will improve dramatically with proper experimental design.

To detect two discrete states with two detectors, a beam splitter should be introduced to separate the respective spectra at approximately their crossing point. The spectrum of a typical commercial dichroic beam splitter is shown in Figure 5. Spectral overlap between the two states (also called "cross-talk" or "leakage" by analogy to traditional analogue electronic communication circuits) is typically unavoidable, so when a photon is detected, there will be some uncertainty as to which state generated the signal. Spectral components of each state arrive at both detectors as illustrated in Figure 5. The fraction of photons that arrive at the wrong detector is the leakage parameter

$$\mathcal{P}(O_i | S_j) = \begin{cases} 1 - \epsilon: & j = 0, i = 0 \\ \epsilon: & j = 0, i = 1 \\ \delta: & j = 1, i = 0 \\ 1 - \delta: & j = 1, i = 1 \end{cases} \quad (6)$$

The crossover of the signals between the channels is characterized by the leakage parameters δ and ϵ. The two-detector experimental scheme using a commercial beam splitter shown in Figure 5 gives leakage parameters of $\delta = 0.120$ and $\epsilon = 0.137$. Leakage parameters can also be determined empirically from control experiments.

The goal is to determine the number of photons required to reduce the uncertainty in the state of the system to a desired level (e.g., 5%, 1%, 0.1%). By calculating the mutual information as a function of the number of photons observed, we can readily accomplish this. A single photon therefore conveys 0.45 bits of information on average. At this level, the state would be uncertain ~13% of the time. Therefore, multiple events are necessary to determine the state of the molecule, and we must calculate the information present as a function of the number of photons observed. $\mathcal{S}(\mathbb{S}, \mathbb{O}_n) = \mathcal{H}(\mathbb{S}) - \mathcal{H}(\mathbb{S}|\mathbb{O}_n)$. If there is no change in count rate or lifetime upon change of environment, no further information is available from those measurements. If n photons are observed, l of which are in a given channel and $n - l$ of which are the other channel, then this can be treated as the mutual information for a binomial random variable where the Bernoulli parameter (leakage parameter) is given by eq 6.
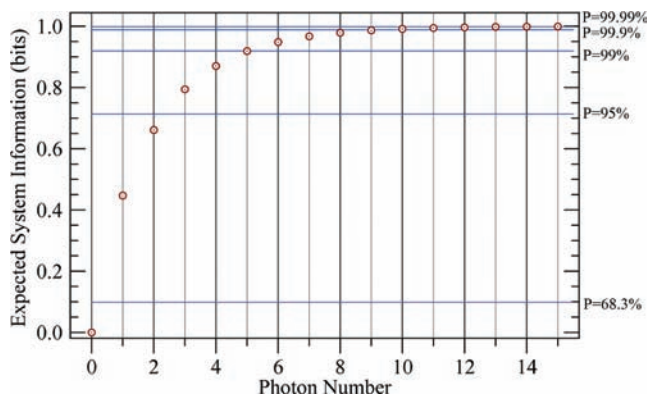
Single-Molecule Design and Interpretation

J. Phys. Chem. A, Vol. 110, No. 31, 2006 **9749**

Figure 6. Illustration of the increase of information about the system with photon number for C153 using the beam splitter and two detectors. The horizontal lines represent probability values of 95%, 99%, 99.9%, and 99.99% from bottom to top.
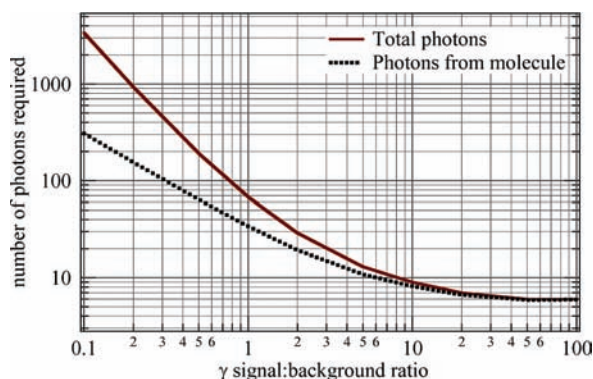

Figure 7. Plot of the total number of photons (solid line) and the average number of signal photons originating from the molecule (dotted line) as a function of signal-to-background ratio that are required to determine the state of C153 with 99% confidence.

A plot of the mutual information between the signal and the system vs the total number of photons observed, n, is in Figure 6.

Note the increase of the system information with the number of photons observed (Figure 6). This indicates that on average it will take between 5 and 6 photons to gain enough information to assign the state of the C153 molecule in this example with better than 99% confidence. Of course, for a given observation, it may not be the case that enough information has been transferred to make the inference regarding the state to the same level of confidence.

Background photons in the detection system do not convey any information regarding the system. As a result, on average, the presence of background reduces the average amount of mutual information provided per photon. Mathematically, this has the same effect as increasing the leakage parameter. If the ratio of signal to background, $\gamma = S/B$, and the ratio of the background in each channel, $\beta = B_0/B_1$, are included, the new leakage parameter is

$$\epsilon^* = \frac{1 + (1 + \beta)\gamma\epsilon}{(1 + \beta)(1 + \gamma)} \qquad (7)$$

Figure 7 plots the log of S/B, with $\beta = {}^1\!/_2$, versus the number of photons required to determine the state of the molecule to the 99% confidence for C153 as in the prior example. The information theory analysis demonstrates the rapid degradation of the experiment with increasing background levels. Even considering only the number of photons from the molecule—
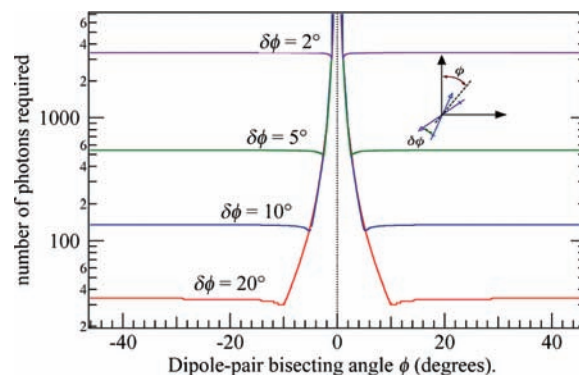

Figure 8. Plot of the number of photons required to distinguish, with 99% confidence, between two orientations of the dipole moment vs the angle of the bisector of the dipole moments with respect to the detection polarization axis. Several values of the dipole moment angle difference are plotted. From top to bottom, they are 2°, 5°, 10°, and 20°. The inset shows the definitions of ϕ and $\delta\phi$.

which is the critical parameter because it is most closely related to the likelihood of photodegradation—the information content is significantly degraded because of the background photons. This analysis also suggests that for this experiment there is little benefit in increasing signal-to-background ratio much beyond 10.

A two-color experiment can be performed many ways including simultaneous imaging of immobilized molecules in the two spectral regions using a CCD camera. These information theory results can be used to optimize CCD exposure time and illumination intensity to provide optimal resolution of the two states. State assignments are likely to fail when the molecule is fluctuating more rapidly than the appropriate number of photons can be emitted. Confocal microscopy can measure immobilized as well as freely diffusing molecules. A freely diffusing molecule would need to emit enough photons during its transit of the illumination volume. Selection of bursts with inadequate numbers of photons to provide enough information to resolve the states will result in frequent misassignment of the states and will tend to bias the ratio of state populations toward the maximum entropy result of unity.

C. Two Polarization Channels: Emission Dipole Assignments. Assignment of an emission-dipole azimuthal angle is a common goal of single-molecule measurements, as it can be exploited to determine molecular-level motions and geometries.[25,42−44] The number of photons required to determine the dipole angle depends not only on the desired resolution but also on the angle itself.

The first experiment we examine is distinguishing between two configuration states that have a difference, $\delta\phi$, in the azimuthal angle, ϕ, of the emitting dipole moment in the lab frame of reference. We can treat this problem with the same formalism we developed above for molecules undergoing spectral fluctuations due to different local environments. If the resulting photon stream is resolved with a polarizing beam splitter into two detectors, the leakage parameters are now related to the angle of the dipole in each state relative to the polarizing beam splitter, $\epsilon = \cos^2(\phi - \delta\phi/2)$ and $1 - \delta = \cos^2(\phi + \delta\phi/2)$. Without specific orientation, ϕ will be randomly distributed.

Figure 8 shows the number of photons required to gain enough information, on average, to distinguish between the states with emitting dipoles separated by different angles $\delta\phi$ with 99% confidence. A bulk experiment or a single-molecule experiment with low time resolution would only resolve the
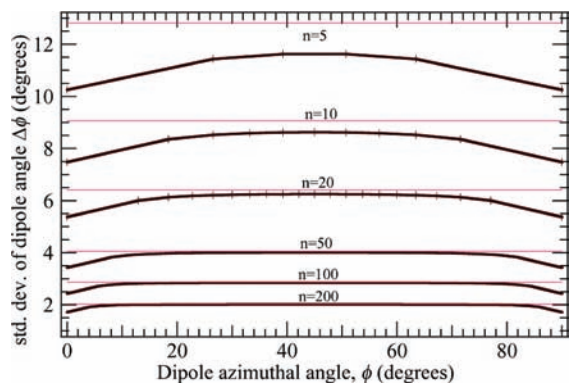
9750 *J. Phys. Chem. A, Vol. 110, No. 31, 2006*

Talaga

Figure 9. The uncertainty in azimuthal angle as predicted by the standard deviation of the posterior likelihood function. The thin lines are the corresponding estimates from a Gaussian approximation to the posterior likelihood.

Figure 10. The information (right axis) regarding the dipole azimuthal angle is plotted as a function of the total number of photons observed. The left axis is the bin width corresponding to the same amount of information. The thin line is the Gaussian approximation to the standard deviation of the posterior angle distribution.

average emission dipole orientation, ϕ, the bisector between the two dipoles. To resolve the angle-states, the experiment must obtain enough photons prior to the fluctuation of the system. Notice that when the dipoles are on the opposite sides of the vertical (or horizontal) polarization axis—when the mean angle approaches either of the polarization axes to within half the angle change $\delta\phi/2$—the number of photons required increases until it diverges in the case that the dipoles are symmetrically spaced on either side of a polarization axis, and thus indistinguishable in this experiment. This would suggest that the experimentalist should examine molecules with average ϕ angles outside of these areas when looking to follow fluctuating polarization trajectories. When $\delta\phi$ is large, there is a greater range of angles that will reduce the sensitivity of the experiment to the angle changes.

Another common single-molecule experiment experiment is to determine the distribution of dipole angles in a sample where that angle is stationary. Angle distributions are usually reported as distribution histograms. Information theory can guide the experimental design on the basis of the desired histogram bin width or resolution. To distinguish a dipole angle to arbitrary resolution, the experiment must reduce the likelihood that some other angle is consistent with the observed data. To evaluate the information required for a given resolution, we use the Bayesian posterior likelihood distribution

$$\mathscr{P}(\phi|\,l,n) = \frac{2\cos^2(\phi)^l \sin^2(\phi)^{n-l}}{\mathbf{B}\left(l+\frac{1}{2},\,n-l+\frac{1}{2}\right)} \qquad (8)$$

in the expression for mutual information, where $\mathbf{B}(l + {}^1\!/_2,\, n - l + {}^1\!/_2)$ is the complete β function.

Figure 9 shows the error in the measurement of the azimuthal angle estimated from the standard deviation of the Bayesian posterior likelihood distribution. The uncertainty for a given number of photons reaches its maximum at 45°. As the number of photons increases, the standard deviation approaches that predicted from a Gaussian approximation to the posterior likelihood function, $\sigma_{\text{Gauss}} = {}^1\!/_2 n^{-1/2}$, except as ϕ approaches 0° or 90°. Near the angles where one of the channels becomes zero, the variance for a given number of photons decreases. This appears to contradict the results in Figure 8. The difference between the two results lies in the different systems that each describes. In Figure 8, distinguishing dipoles separated by a particular angle becomes impossible if they are angularly equidistant from 0° or 90° because of the symmetry of the detection. By contrast, there is no such difficulty when trying to estimate the angle of a single dipole to specified resolution.
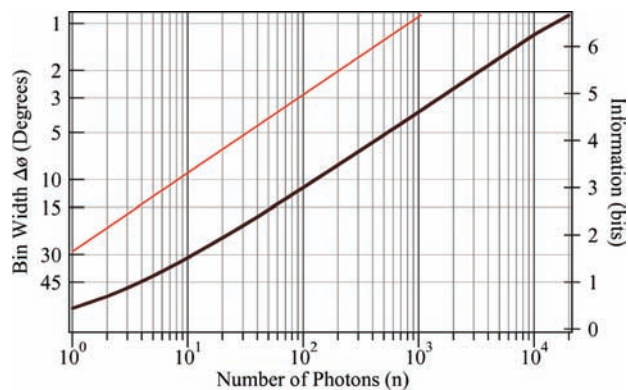
Figure 10 shows the increase in information and improvement of resolution as the number of photons increases. The difference in effective bin width between the information theory result and the Bayesian calculation of the variance is a result of the level of significance implied by localization to a bin of width $\Delta\phi$ and that implied by the standard deviation. The information theory analysis provides a way of determining what the smallest useful bin width for a histogram will be, on the basis of the number of photons being observed.

D. Measurement of FRET Distance from Two-Channel Data. One of the most common applications of two-channel–two-color single-molecule measurements is to determine the efficiency of FRET between a donor and an acceptor dye.[26,40,45] Changes in the energy transfer efficiency $\Phi_X = I_A/(I_A + I_D)$ can be related to structural fluctuations in enzymes,[46–48] folding proteins[49–52] or peptides,[27,53] or nucleic acids.[40,45,54]

The leakage parameters used for the two-color problem can be adapted to FRET. If we initially neglect details such as spectral cross-talk between donor and acceptor detection channels, then the zero-order leakage parameter, $\epsilon°$, for use in eq 6 is

$$\epsilon° = \frac{I_A}{I_A + I_D} = \Phi_X = \frac{1}{1 + (r/r_0)^6} \qquad (9)$$

where I_A is the acceptor intensity, I_D is the donor intensity, r is the donor–acceptor distance, r_0 is the Förster radius, and Φ_X is the energy transfer efficiency.

Experimental details can be included with additional Bernoulli processes in serial with the one from FRET. Spectral overlap can be included the same as it was for the two-color problem in section 3.2, except now there is the additional Bernoulli process associated with FRET. The leakage parameter for the acceptor channel in this case is defined by

$$1 - \epsilon_A = \Phi_X(1 - \delta) + (1 - \Phi_X)\,\epsilon \qquad (10)$$

Finite signal-to-background is included the same way it was in section II.C by using eq 10 in eq 7

$$\epsilon_A = \frac{\gamma}{1 + \gamma}\,[\epsilon(1 - \Phi_X) + (1 - \delta)\Phi_X] + \frac{1}{1 + \gamma}\frac{1}{1 + \beta} \qquad (11)$$

Since the acceptor absorbs at a longer wavelength than does the donor, the acceptor will, in general, also be excited by the laser, albeit less efficiently. If there is nonnegligible direct
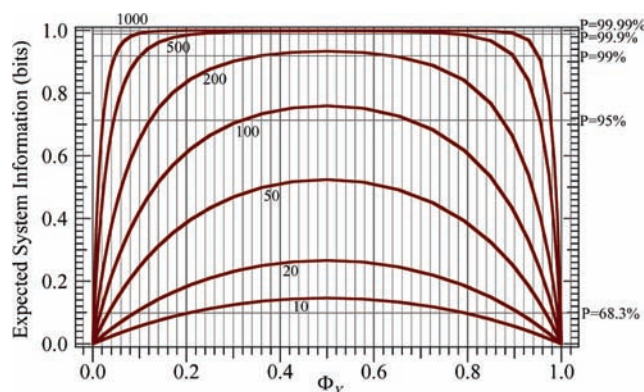
Single-Molecule Design and Interpretation

J. Phys. Chem. A, Vol. 110, No. 31, 2006 **9751**

Figure 11. The average information delivered, when measuring a 10% relative distance change, as a function of the energy transfer efficiency between the donor and the acceptor is plotted for different numbers of photons as labeled.
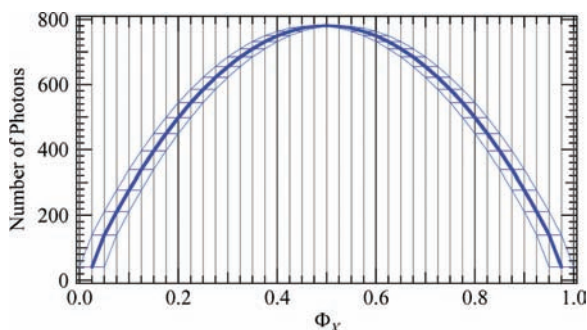

Figure 12. The number of photons required (thick line) to determine Φ_X to within 0.05 (indicated by the thin lines).

excitation of the acceptor with an acceptor/donor excitation ratio of ξ, then the leakage parameter is

$$\epsilon_\ddagger = \frac{\epsilon\,(1 - \Phi_X) + (1 - \delta)(\xi + \Phi_X)}{1 + \xi}\frac{\gamma}{1 + \gamma} + \frac{1}{1 + \beta}\frac{1}{1 + \gamma} \quad (12)$$

Many experiments resolve states simply by the changes in energy transfer efficiency. Figure 11 shows the system information expected for different numbers of photons when attempting to distinguish between two states that differ in distance by 10% of their mean distance. The intersections between the horizontal confidence levels and the information curves show the useful range of Φ_X for the given number of photons. One often can tailor the system to fluctuate in this region with judicious choice of donor and acceptor dyes to provide a Förster radius in the center of the fluctuation range. The illumination conditions must be adjusted to allow collection of enough photons before the system fluctuates. For diffusive experiments, the residence time of the system in the observation volume is often the limiting factor for the number of photons observable. However, there is always a broad distribution of total burst amplitudes, and information theory can set the burst amplitude threshold for analysis purposes for a given desired distance change resolution.

A common way of representing single-molecule data is to calculate a histogram to determine the distribution of energy transfer efficiency; the desired resolution is effectively the bin width of the calculated histogram. Uniform bin widths are the most common choice, and this has an important effect on the uncertainty associated with assigning a particular segment of an immobilized trajectory or a particular fluorescence burst to a unique bin. Figure 12 shows the number of photons required to assign an observation with a particular Φ_X to a bin of width
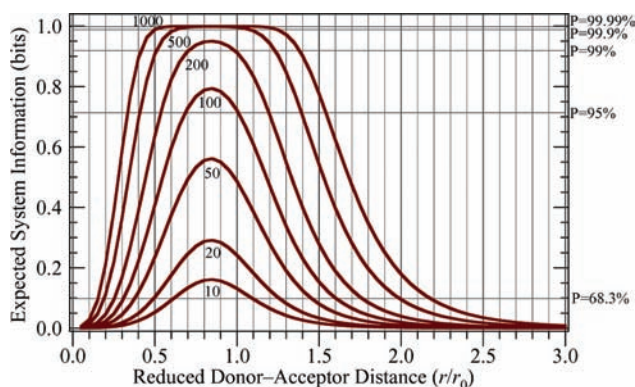

Figure 13. The average information delivered, when measuring a distance change of 10% of r_0, as a function of the reduced distance between the donor and the acceptor is plotted for $n = \{10, 20, 50, 100, 200, 500, 1000\}$ as labeled in the figure.

5% at 95% confidence. Note that there is nearly an order-of-magnitude difference between the photons required to assign the bins at ~0 or ~1 and those near ~0.5. At first, this seems to be in contradiction with the results shown in Figure 11. This comparison illustrates the effects of the highly nonlinear relationship between the variables Φ_X and r/r_0. In the case of Figure 11, there is a 10% change in distance that is to be detected, whereas in Figure 12, it is a 5% absolute difference in Φ_X that is to be detected. A 10% change in distance near the Förster radius corresponds to a 15% change in Φ_X. The first two bins of the 20-bin histogram implied in Figure 12 would correspond to a change of r/r_0 from 1.84 to 1.52. The details of the system dictate the most appropriate way to design the experiment and treat the resulting data. The flexibility of information theory allows fine-tuning of experiments to allow them to evaluate specific hypotheses.

When an estimate of the Förster radius is used, the changes in energy transfer can be interpreted in terms of a physical distance change. Figure 13 shows the information content of a photon stream coming from a molecule that can exist in two distance-states that are separated by 10% of the Förster radius. The intersection of the horizontal likelihood lines and the information curves illustrate the useful range of distances that can be distinguished in a two-state system. Selection of a filtering procedure for immobilized trajectory data and thresholds for diffusive fluorescence burst data is contingent on the resolution that is required given the limits imposed by information theory.

The exact value of the Förster radius depends on the choice of donor and acceptor dyes and their local environment once attached to the molecule. Choice of dyes for a FRET experiment is subject to many constraints such as commercial availability, proper chemical reactivity for conjugation, photostability, and desired Förster radius. Figure 14 shows the expected information from a given number of photons versus Förster radius for a system undergoing a relative distance change of 10% centered at some average distance r. The information is maximized when the Förster radius is identical to the mean distance, r, as expected. However, it is not often the case that the exact distances can be known ahead of time, nor is it usually possible to tune the Förster radius to exactly the desired distance. Figure 14 shows that the exact value of r_0 becomes less important as the number of photons collected increases. In this case, information theory can guide the decision between a more convenient donor–acceptor pair and one that is optimized for a particular system.
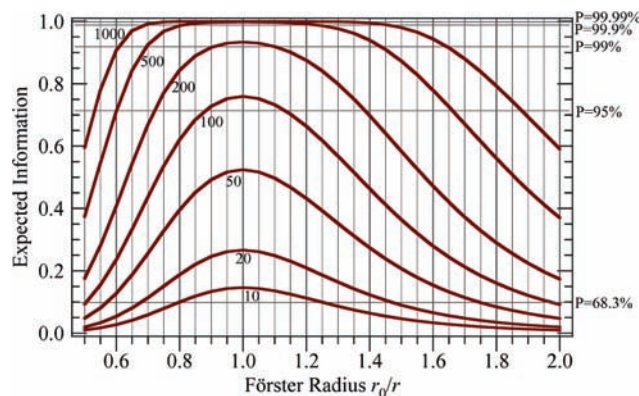
Figure 14. Expected information for a system undergoing a 10% relative distance fluctuation about a center position of *r* vs the relative Förster radius at various numbers of photons as labeled in the figure.
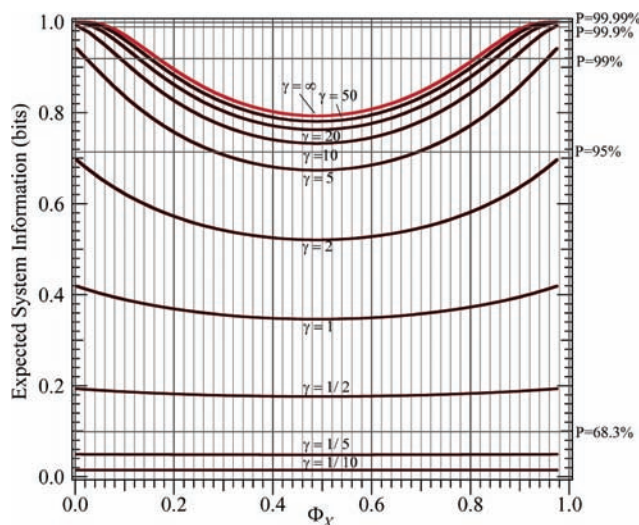

Figure 16. The effective resolution, Δr of a FRET measurement vs the distance between the donor and the acceptor. Δr represents the minimum distance from r/r_0 required to distinguish the state in question for a given number of photons as labeled in the figure.

Figure 15. The degradation due to finite signal-to-background of the information from 1000 photons in a measurement of Φ_X to within 0.05. The top line, in red, is the background-free limit of information.

Degradation of the information present in a FRET photon stream can occur because of background photons. Figure 15 shows the effect of increasing levels of uniform background ($\beta = {}^1\!/_2$) on the information present in 1000 photons when determing the energy transfer efficiency to within 0.05. The information advantage of states near $\Phi_X = 0$, 1 disappears with increasing levels of background. As in the two-color case, we see that there are only modest improvements in information content for signal-to-background ratios greater than ~10.

The two-state information theory formulation allows us to determine the uncertainty in distance that will result, on average, from a fixed number of photons. This uncertainty will depend on the distance as shown in Figure 16. This is expected from the strong distance dependence of the mutual information in Figure 13.

The resolution of a FRET measurement depends on the center point of the measurement as well as the number of photons observed. This is illustrated in Figure 16. The distance dependence of the uncertainty would suggest that for any linear filtering operation the errors will be distance-dependent. A filtering operation performed by photon number would give errors as predicted in Figure 16 according to the effective number of photons included in the kernel of the filter. Likewise, a time-domain filter would also show varying error according to the kernel-weighted number of photons that go into calculating the distance at a given point in the trajectory.
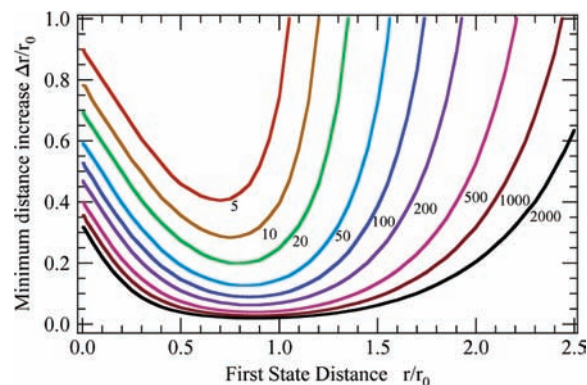
Single-molecule trajectories commonly show transitions in and out of nonemitting states. These photophysical and photochemical processes can potentially perturb the results derived from the measurement of a single-molecule photon stream. One of the most common photophysical processes observed is intersystem crossing that results in "triplet blinking".[19] During a blinking event, the only photons that arrive are due to the background. No system information will arrive until the dye reverts to an emissive state.

The presence of two dyes that are coupled due to FRET further complicates the analysis in the presence of blinking. A full treatment of these issues requires abandoning the assumption that the state of the molecule is stationary during the measurement and including the information present from the relative intensities. The transient blinking of the donor will turn off its fluorescence, eliminate the energy transfer ($\Phi_X = 0$), and cause large changes in the count rate. This will provide a source of information in addition to that arising from partitioning of photons between detectors. Transient blinking of the acceptor will eliminate its fluorescence and the energy transfer, but will typically change the count rate only by as much as the acceptor is directly excited.

The information theory formalism can treat these effects by including the information present in the interphoton times (sequential exponential processes) along with the information due to the distribution of photons between detectors (sequential and multiple Bernoulli processes) as has been the focus of this paper. By using only the information due to the detector distribution, it is possible to estimate whether a FRET measurement will be able to distinguish transient blinking events of the acceptor or the donor even if the information from the explicit interphoton times is not included. This can then be compared to the information gained by including the intensity information. This gives an effective way to estimate which regions of the FRET efficiency or distance will be corrupted due to ambiguity induced by the presence of blinking in the donor and the acceptor as well as determining what information is most important for identifying transient nonemissive states of the donor and acceptor.

One must explicitly include both transient blinking events as new states in the information theoretical analysis. The leakage parameter for a FRET system with a dark donor is

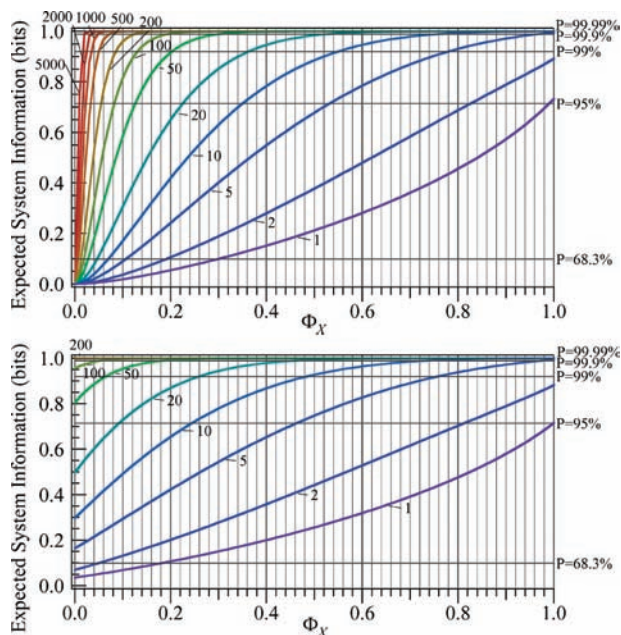$$\frac{\beta}{(1 + \beta)(1 + \gamma)} + \frac{\gamma\epsilon}{1 + \gamma} \tag{13}$$

Figure 17. The top panel shows the expected system information for distinguishing dark acceptors from a state at a given FRET efficiency using a signal-to-background ratio of 10:1 with background evenly split between donor and acceptor channels. There is no direct acceptor excitation, and the information only includes the binomial information of the n photons (as labeled) being divided between the two detectors. The bottom panel shows the same system with 20% direct acceptor excitation.

Figure 18. The top panel shows the expected system information for distinguishing dark donors from a state at a given FRET efficiency using a signal-to-background ratio of 10:1 with background evenly split between donor and acceptor channels. There is no direct acceptor excitation, and the information only includes the binomial information of the n photons being divided between the two detectors. The bottom panel shows the same system with 10% direct acceptor excitation and includes the information from the interphoton times.

The leakage parameter for a FRET system with a nonabsorptive dark acceptor is

$$\frac{\beta + \gamma + \beta\gamma}{(1+\beta)(1+\gamma)} - \frac{\gamma\delta}{1+\gamma} \tag{14}$$

Similar to the previous analyses, the ability to distinguish dark or damaged states will depend on the nature of the undamaged states and the experimental design. If there is no direct excitation of the acceptor, then dark acceptor states are indistinguishable from $\Phi_X = 0$ states. This is illustrated in the top panel of Figure 17. If a small amount of direct excitation of the acceptor is present, then the dark acceptor states are much more readily distinguishable from states with small values of Φ_X. The result is that far less of the histogram will be corrupted when dark acceptor states are significant as long as a small amount of direct acceptor excitation is permitted and taken into account. The intensity makes little contribution to the information that will allow separation of dark acceptor states.

For dark donor states in the absence of direct acceptor excitation, the only signal is due to the background. In this case, the ratio of background signals in the donor and acceptor channels determines the region of the Φ_X histogram that could be corrupted by donor blinking. The top panel of Figure 18 illustrates this for the case of symmetric background $\gamma = 10$, $\beta = 0.5$. Direct excitation of the acceptor does not substantially improve the ability to distinguish dark acceptor states and results in a shift of the corrupted region to higher Φ_X values. In fact, direct acceptor excitation slightly reduces the information in the count rate. The information for distinguishing dark donors in the interphoton times is much higher than that of the channel ratios.

The bottom panel of Figure 18 shows the effect of including the intensity information in the information theoretical analysis. It is apparent that inclusion of this second source of information
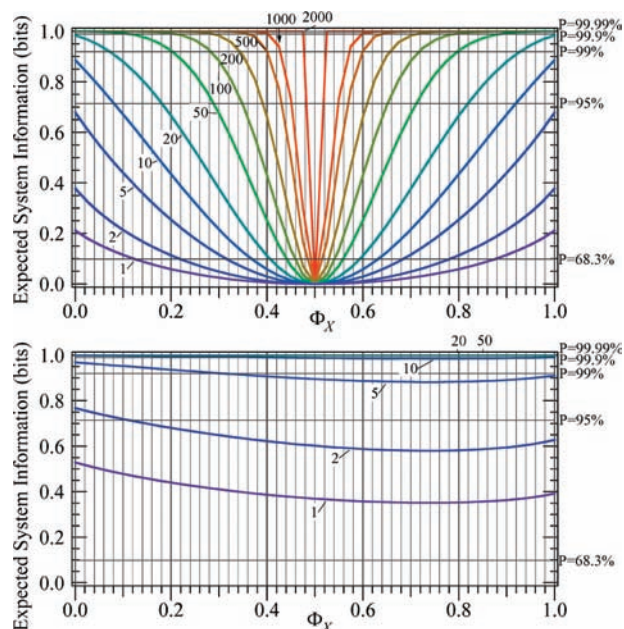
has resulted in a substantial improvement in the experiment's ability to distinguish dark donor states from states in the corrupted region of the histogram. Note that, with a combination of the information from direct acceptor excitation and the interphoton times, the number of photons required to distinguish a damaged donor−acceptor system is substantially fewer than that required to specify a bin in the distribution function.

IV. Conclusions

Both the rate of delivery and total amount of information is limited in single-molecule experiments. These limits are fundamental and unavoidable because the sequence of observations of a single molecule are each sampled stochastically from a distribution. Single-molecule trajectories typically end upon photodestruction of the dye. When interpreting bulk measurements, we are often shielded from this basic fact by the sheer number of replicate single-molecule experiments that occur synchronously when we excite a sample. The random sampling means that it becomes difficult to analyze single-molecule data "by eye", so sophisticated data processing algorithms must be devised.[28] Information theory provides the limit of our ability to make inferences from the data independent of the specific algorithm used to process the data. No data processing technique can increase the amount of information present in the data.

When single-molecule data contains features that are easily discernible "by eye", it becomes tempting to ignore those features that are not obvious. Information theory tells us if it is possible to learn anything about the system from more subtle data features or through more sophisticated data analysis. Information theoretical analysis is critical, because the amount of information recorded in the data set usually greatly exceeds the amount of mutual information between the data and the system. It is this mutual information that allows us to learn about the system from the measurement. If the measurement does not deliver enough mutual information in the time before the system

either fluctuates to another state or is photobleached, then determination of the state is not, in general, possible. However, information theory also shows that sometimes inference can be made with confidence and can tell us what fraction of the time inference will be possible.

The flexibility of the information theory approach allows the experimentalist to optimize an experiment to evaluate a specific hypothesis and formulate it in different ways. It also illustrates the importance of prior information through control experiments. Since the number of photons required for a reliable measurement vary greatly depending on the details of the states to be distinguished, it is vital to have some idea what those states might be so that the experiment can be properly designed and not give inconclusive results.

We have demonstrated the utility of information theory in the analysis of single-molecule fluorescence experiments. We analyzed the limitations that single-molecule spectroscopy puts on the photon stream arising from a single-molecule system. We quantified the amount of information that is typically stored in a single-molecule measurement and how this compares to the amount of information present in the photon stream. For the case of a spectrally shifting dye, we quantitatively determined the consequences of spectral resolution, cross-talk between channels, and background on the ability to distinguish the two colors. We applied information theory to the case of polarized single-molecule spectroscopy and treated both discrete and continuous state spaces. We analyzed single-molecule FRET measurements using information from detector ratios and total intensity. We discussed the consequences of limited information on the selection of FRET histogram resolution, donor−acceptor dye-pair choice, and discrimination of dark donor or acceptor states.

For the purpose of our analysis, we considered the state of the system to be stationary. Many systems will not be stationary. The fluctuations of the system serve to increase the entropy of the system. However, it is usually the case that the state will be stationary for some period of time, on average. In this case, the information delivered to the observer in that finite amount of time must exceed that amount needed for the desired level of accuracy for the state or parameter determination. This provides a very useful principle for experimental design and data analysis. An in-depth analysis of single-molecule information as dynamic variable will be the topic of a forthcoming paper.

Appendix A

In this Appendix, we define our information theory notation and derive the equations that we have used to treat single-molecule experiments.

Shannon defined the information functional ($\mathscr{I}$) as

$$\mathscr{I}(O_k) = -\log_2 \mathscr{P}(O_k) \quad \text{(A1)}$$

with $\mathscr{P}(O_k)$ being the probability of observing an event O_k that has several possible random outcomes $\mathbb{O} = \{O_1, O_2, ..., O_n\}$ with probabilities $\{\mathscr{P}(O_1), \mathscr{P}(O_2), ..., \mathscr{P}(O_n)\}$ that sum to unity ($\sum_{k=1}^{n} \mathscr{P}(O_k) = 1$). The logarithm base 2 gives the information

in bits. Information quantifies changes in uncertainty; the more unlikely an outcome is, the more information it transmits. For example, knowing that a particular English word contains the letter "q" conveys much more information− 5.8 bits more− about what the word might be than knowing that the word contains the letter "e". This information is based upon the fact that ∼57 times more English words contain "e" than contain "q".[55] From the point of view of the scientific method, the more unlikely a hypothesis is, the more evidence is required to accept that the hypothesis is plausible.

Shannon entropy ($\mathscr{H}$) is defined as the expectation value of information

$$\mathscr{H}(\mathbb{O}) = \langle \mathscr{I}(O_k) \rangle_k = -\sum_{k=1}^{n} \mathscr{P}(O_k) \log_2 \mathscr{P}(O_k) \quad \text{(A2)}$$

Because $\lim_{p\to 0}(-p \log_2 p) = 0$, events with zero probability are omitted from the sum. Similarly, when $n = \infty$ the summation converges rapidly for normalizable $\mathscr{P}(O_k)$ and physically realizable events. The definition of Shannon entropy is derivable as a unique function based upon two requirements.[56,57] The first requirement is that entropy must reach its maximum value, $\mathscr{H}^*(\mathbb{O}) = \log_2 n$, when $\mathscr{P}(O_1) = \mathscr{P}(O_2) = \cdots = \mathscr{P}(O_n) = 1/n$, that is, when we have no knowledge as to the state of the system. The second requirement is that, for the case of two (or more) observables, the entropy must be maximized when knowledge of one observable imparts no knowledge about the other observable, that is, when the two observables are independent. The entropy reaches its minimum value of zero when there is complete certainty in the outcome $\mathscr{P}(S_m) = 1$, $\mathscr{P}(S_{j\neq m}) = 0$. The entropy of the system measures the average amount of information needed to make a certain determination of the outcome.

The goal of a single-molecule measurement is often to determine some unknown property or state of the system. This property is most often not directly observable, and the investigator's objective is to determine this hidden state on the basis of some experimental information. The uncertainty in the hidden state should be reduced by the observation; this can be quantified by Bayesian statistics. From a Bayesian perspective, if the state of the molecule is unknown and can be any of a set of states $\mathbb{S} = \{S_1, S_2, ..., S_i, ..., S_n\}$, each with respective probability $\{\mathscr{P}(S_1), \mathscr{P}(S_2), ..., \mathscr{P}(S_i), ..., \mathscr{P}(S_n)\}$, then the change in likelihood of a particular state of the system after an observation O_j can be expressed by the Bayes theorem

$$\mathscr{P}(S_i|O_j) = \mathscr{P}(S_i) \frac{\mathscr{P}(O_j|S_i)}{\mathscr{P}(O_j)} = \frac{\mathscr{P}(S_i)\mathscr{P}(O_j|S_i)}{\sum_k \mathscr{P}(S_k)\mathscr{P}(O_j|S_k)} \quad \text{(A3)}$$

where $\mathscr{P}(S_i|O_j)$ is the conditional probability that the real state is S_i given the condition that O_j has been observed. $\mathscr{P}(S_i)$ is the probability of S_i being the real state prior to the observation. $\mathscr{P}(O_j|S_i)$ is the likelihood of O_j occurring if it were known that the real state is S_i. $\mathscr{P}(O_j)$ is the marginal probability of O_j, that is, the likelihood of O_j independent of the state of the system.

To quantify the degree to which the experimental photon stream is communicating information about the hidden state of the system, we evaluate the information and entropy conditionally. Conditional information

$$\mathscr{I}(S_i|O_j) = -\log_2[\mathscr{P}(S_i|O_j)] = -\log_2\left[\frac{\mathscr{P}(S_i, O_j)}{\mathscr{P}(O_j)}\right] \quad \text{(A4)}$$

Single-Molecule Design and Interpretation

J. Phys. Chem. A, Vol. 110, No. 31, 2006 **9755**

is the information of S_i, the hidden state of the single-molecule system, given that O_j, the observation of one or more photons, has occurred. $\mathscr{P}(O_j, S_i)$ is the likelihood that both O_j and S_i occur.

The combination of eqs A2 and A4 gives the *conditional entropy*

$$\mathscr{H}(\mathbb{S}|\mathbb{O}) = \sum_i \sum_j \mathscr{P}(S_i, O_j)\mathscr{T}(S_i|O_j) =$$
$$-\sum_i \sum_j \mathscr{P}(S_i, O_j) \log_2[\mathscr{P}(S_i|O_j)] \quad (A5)$$

which is the information expected to be still undetermined in the system $\mathbb{S}$ (the set of all relevant hidden states of the system) once an observation has been taken from $\mathbb{O}$, the set of all possible outcomes of the experiment.

The mutual information

$$\mathscr{T}(\mathbb{S}, \mathbb{O}) = \sum_i \sum_j \mathscr{P}(S_j, O_k)\mathscr{T}(S_j, O_k) =$$
$$\sum_i \sum_j \mathscr{P}(S_i, O_j) \log_2 \left[\frac{\mathscr{P}(S_i, O_j)}{\mathscr{P}(S_i)\mathscr{P}(O_j)}\right] \quad (A6)$$

is the amount of information that the measurement $\mathbb{O}$ conveys, on average, about the system $\mathbb{S}$. It is equivalent to the decrease in system entropy that occurs as a result of the measurement

$$\mathscr{T}(\mathbb{S}, \mathbb{O}) = \mathscr{H}(\mathbb{S}) - \mathscr{H}(\mathbb{S}|\mathbb{O}) \quad (A7)$$

Typically, state assignments will require observation of multiple photons. The interphoton times of a Poisson emitter will be exponentially distributed. Multiple sequential interphoton times from a Poisson emitter will be Γ-distributed. For most of the examples in this paper, the overall emission rate was constant and should not provide any information for state inference.

Other observables that can be simultaneously measured, such as wavelength and polarization, are determined by multiple detection channels and are treatable with the categorical distribution, which is the generalization of the Bernoulli distribution to multiple, mutually exclusive outcomes. Multiple events drawn from a categorical distribution give multinomial statistics

$$\mathscr{P}(O_{\vec{l}}| S_j, O_n) = \mathrm{Mul}(\vec{l}; \vec{p}_j) = \frac{n}{l_1, \cdots, l_m} \prod_{i=1}^m p_{i,j}^{l_i} \quad (A8)$$

with $\vec{l}$ and $\vec{p}_j$ being vectors containing the number of times l_i each mutually exclusive event i occurs out of n total events and the probabilities $p_{i,j}$ of each event given that the state of the system is j.

Often, there are multiple observables recorded jointly in an experiment that are independent for a given state, e.g., the excited-state lifetime and interphoton time may be simultaneously recorded. They are essentially independent, but not mutually exclusive. The conditional self-information of the state S_j given n observables $O_{k_1}, \cdots, O_{k_n}$ is

$$\mathscr{T}(S_j|O_{k_1}, \cdots, O_{k_n}) = -\log_2 \mathscr{P}(S_j|O_{k_1}, \cdots, O_{k_n}) =$$
$$-\log_2 \left[\frac{\mathscr{P}(S_j, O_{k_1}, \cdots, O_{k_n})}{\mathscr{P}(O_{k_1}, \cdots, O_{k_n})}\right] \quad (A9)$$

The conditional entropy between the set of all states $\mathbb{S}$ and the set of all combinations of n observables $\mathbb{O}_{k_1}, \cdots, \mathbb{O}_{k_n}$ is

$$\mathscr{H}(\mathbb{S}|\mathbb{O}_1, \cdots, \mathbb{O}_n) = -\sum_j \sum_{k_1} \cdots \sum_{k_n} \mathscr{P}(S_j, O_{k_1}, \cdots, O_{k_n}) \times$$
$$\log_2 \left[\frac{\mathscr{P}(O_{k_1}, \cdots, O_{k_n}|S_j)\,\mathscr{P}(S_j)}{\mathscr{P}(O_{k_1}, \cdots, O_{k_n})}\right] \quad (A10)$$

with

$$\mathscr{H}(\mathbb{S}|\mathbb{O}_1, \cdots, \mathbb{O}_n) = \mathscr{H}(\mathbb{O}_1, \cdots, \mathbb{O}_n|\mathbb{S}) + \mathscr{H}(\mathbb{S}) - \mathscr{H}(\mathbb{O}_1, \cdots, \mathbb{O}_n) \quad (A11)$$

The information and entropy of multiple instances of groups of jointly measured conditionally independent observables can be treated using eq A8, for each observable, in eqs A9 and A10, respectively.

In a spectrally resolved experiment, the amount of information delivered per photon will be wavelength-dependent: λ is $\mathscr{T}(\mathbb{S}, \mathbb{O}_\lambda) = \mathscr{H}(\mathbb{S}) - \mathscr{H}(\mathbb{S}|\mathbb{O}_\lambda)$. For a two-state system with equal a priori probabilities for each state (i.e., $\mathscr{P}(S_0) = \mathscr{P}(S_1) = (1/2)$), eq 1 gives the entropy $\mathscr{H}(\mathbb{S}) = 1$ bit. The required conditional entropy is

$$\mathscr{H}(\mathbb{S}|\mathbb{O}_\lambda) = -\sum_{i=0}^{1} \sum_\lambda \mathscr{P}(S_i, O_\lambda) \log_2 \left[\frac{\mathscr{P}(S_i, O_\lambda)}{\mathscr{P}(O_\lambda)}\right] \quad (A12)$$

with $\mathscr{P}(O_\lambda) = \sum_{k=0}^{1} \mathscr{P}(O_\lambda|S_k)\mathscr{P}(S_k)$, the state-weighted sum of the two spectra. Evaluating the sum over just the i states in eq A12 gives the wavelength-dependent information shown in Figure 3.

The two-color problem reduces the spectral resolution to two channels. The channel observed for a particular photon is a Bernoulli random event. The parameter describing the Bernoulli random event is state-dependent. The larger the change, the more information per photon. If n photons are observed, l of which are in a given channel and $n - l$ of which are the other channel, then this is a special case of eq A8 for n observations of two mutually exclusive observables

$$\mathscr{P}(O_l| S_j, O_n) = \begin{cases} \binom{n}{l}(1 - \epsilon)^l \epsilon^{n-l} : & S_0 \\ \binom{n}{l}(1 - \delta)^{n-l}\delta^l : & S_1 \end{cases} \quad (A13)$$

The conditional entropy for such a binomial observation is given by

$$\mathscr{H}(\mathbb{S}|\mathbb{O}) = -\sum_{j=0}^{1} \sum_{l=0}^{n} \mathscr{P}(S_j, O_l|O_n) \log_2 \frac{\mathscr{P}(S_j, O_l|O_n)}{\mathscr{P}(O_l|O_n)} \quad (A14)$$

where

$$\mathscr{P}(S_j, O_l|O_n) = \mathscr{P}(O_l| S_j, O_n)\mathscr{P}(S_j)$$

$$\mathscr{P}(O_l|O_n) = \sum_{m=0}^{1} \mathscr{P}(S_m, O_l|O_n)$$

Equation A14 averaged over l gives the plot in Figure 6.

Polarization is determined from division of the observed photons into two channels with a polarizer. A traditional Bayesian analysis is a useful starting point for information theory analysis of the resolution of a polarized experiment to determine the dye azimuthal angle. The posterior likelihood of ϕ given observation of n photons, l of which were in a particular channel,

is obtained from

$$\mathscr{P}(\phi|l, n) = \mathscr{P}(l|\phi, n)\frac{\mathscr{P}(\phi|n)}{\mathscr{P}(l|n)} \qquad (A15)$$

with the uniform prior distribution

$$\mathscr{P}(\phi|n) = 2/\pi$$

and the binomial probability for polarized detection

$$\mathscr{P}(l|\phi, n) = \binom{n}{l} \cos^2(\phi)^l \sin^2(\phi)^{n-l}$$

marginalized over the angle ϕ

$$\mathscr{P}(l|n) = \int_0^{\pi/2} \mathscr{P}(l|\phi, n)\mathscr{P}(\phi|n)\, d\phi =$$
$$\binom{n}{l} \mathbf{B}\left(l + \frac{1}{2}, n - l + \frac{1}{2}\right)\pi^{-1} \quad (A16)$$

to obtain the posterior distribution

$$\mathscr{P}(\phi|l, n) = \frac{2 \cos^2(\phi)^l \sin^2(\phi)^{n-l}}{\mathbf{B}\left(l + \dfrac{1}{2}, n - l + \dfrac{1}{2}\right)} \qquad (A17)$$

where $\mathbf{B}(l + 1/2, n - l + 1/2)$ is the complete β function. The most likely angle, derived from the maximum of the posterior distribution (eq A17), is

$$\phi^* = \pm\cos^{-1}(\pm\sqrt{l/n}) \qquad (A18)$$

Using the Bayesian posterior likelihood, A17, we can evaluate the resolution of the emission dipole determination as a function of photon number. The information acquired when the angle has been localized to the range $\phi \pm \Delta\phi/2$ is

$$\mathscr{T}(\mathbb{S}_\phi|\phi \pm \Delta\phi/2) = \log_2 \Delta\phi \qquad (A19)$$

This information must be acquired from the mutual information between the measurement and the angle. Equation A7 and eq 4 for the range $[0, \pi/2]$ gives

$$\mathscr{T}(\mathbb{S}_\phi, \mathbb{O}_{l,n}) = \log_2(\pi/2) - \mathscr{H}(\mathbb{S}_\phi|\mathbb{O}_{l,n}) \qquad (A20)$$

with

$$\mathscr{H}(\mathbb{S}_\phi|\mathbb{O}_{l,n}) = -\int_0^{\pi/2} \sum_{l=0}^n \frac{2}{\pi}\binom{n}{l} \cos^2(\phi)^l \sin^2(\phi)^{n-l} \times$$
$$\log_2\left[\frac{2 \cos^2(\phi)^l \sin^2(\phi)^{n-l}}{\mathbf{B}\left(l + \dfrac{1}{2}, n - l + \dfrac{1}{2}\right)}\right] d\phi \quad (A21)$$

Information from the state-dependent intensity comes from the interphoton times, which is a Γ distribution for n photons from a Poisson emitter

$$\mathscr{P}(t|\tau_j, n) = [\Gamma(n)\tau_j]^{-1}(t/\tau_j)^{n-1}\, e^{-t/\tau_j} \qquad (A22)$$

The interphoton times and the channel are conditionally independent and nonmutually exclusive random variables. Using eqs A10, A13, and A22 gives the conditional entropy for both observables

$$\int_0^\infty \sum_{j=0}^1 \sum_{l=0}^n \mathscr{P}(S_j, O_{l,n})\mathscr{P}(t|\tau_j, n) \times$$
$$\log_2\left[\frac{\mathscr{P}(S_j, O_{l,n})\,\mathscr{P}(t|\tau_j, n)}{\displaystyle\sum_{m=0}^1 \mathscr{P}(S_j, O_{l,n})\mathscr{P}(t|\tau_j, n)}\right] dt \quad (A23)$$

References and Notes

(1) Geva, E.; Reilly, P. D.; Skinner, J. L. *Acc. Chem. Res.* **1996**, *29*, 579–584.

(2) Deschenes, L. A.; Vanden Bout, D. A. *Science (Washington, DC, U. S.)* **2001**, *292*, 255–258.

(3) Moerner, W. E. *Springer Ser. Chem. Phys.* **2001**, *67*, 32–61.

(4) Barbara, P. F.; Gesquiere, A. J.; Park, S.-J.; Lee, Y. J. *Acc. Chem. Res.* **2005**, *38*, 602–610.

(5) Ha, T. *Biochemistry* **2004**, *43*, 4055–4063.

(6) Rigler, R.; Edman, L.; Foldes-Papp, Z.; Wennmalm, S. *Springer Ser. Chem. Phys.* **2001**, *67*, 177–194.

(7) Xie, X. S.; Lu, H. P. *J. Biol. Chem.* **1999**, *274*, 15967–15970.

(8) Ha, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 287–292.

(9) Schuler, B. *ChemPhysChem* **2005**, *6*, 1206–1220.

(10) Haran, G. *J. Phys.: Condens. Matter* **2003**, *15*, R1291–R1317.

(11) Zhuang, X.; Rief, M. *Curr. Opin. Struct. Biol.* **2003**, *13*, 88–97.

(12) Lakadamyali, M.; Rust, M. J.; Zhuang, X. *Microbes Infect.* **2004**, *6*, 929–936.

(13) Xie, X. S.; Trautman, J. K. *Annu. Rev. Phys. Chem.* **1998**, *49*, 441–480.

(14) Moerner, W. E. *J. Phys. Chem. B* **2002**, *106*, 910–927.

(15) Nie, S.; Zare, R. N. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 567–596.

(16) Weiss, S. *Nat. Struct. Biol.* **2000**, *7*, 724–729.

(17) Ambrose, W. P.; Goodwin, P. M.; Jett, J. H.; Van Orden, A.; Werner, J. H.; Keller, R. A. *Chem. Rev. (Washington, DC, U. S.)* **1999**, *99*, 2929–2956.

(18) Kou, S. C.; Xie, X. S. *Phys. Rev. Lett.* **2004**, *93*, 180603/1–180603/4.

(19) Eggeling, C.; Widengren, J.; Brand, L.; Schaffer, J.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. A* **2006**, *110*, 2979–2995.

(20) Neuweiler, H.; Schulz, A.; Boehmer, M.; Enderlein, J.; Sauer, M. *J. Am. Chem. Soc.* **2003**, *125*, 5324–5330.

(21) Moerner, W. E.; Fromm, D. P. *Rev. Sci. Instrum.* **2003**, *74*, 3597–3619.

(22) Margittai, M.; Widengren, J.; Schweinberger, E.; Schroeder, G. F.; Felekyan, S.; Haustein, E.; Koenig, M.; Fasshauer, D.; Grubmueller, H.; Jahn, R.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15516–15521.

(23) Slaughter, B. D.; Allen, M. W.; Unruh, J. R.; Urbauer, R. J. B.; Johnson, C. K. *J. Phys. Chem. B* **2004**, *108*, 10388–10397.

(24) Watkins, L. P.; Yang, H. *J. Phys. Chem. B* **2005**, *109*, 617–628.

(25) Talaga, D. S.; Jia, Y.; Bopp, M. A.; Sytnik, A.; DeGrado, W. A.; Cogdell, R. J.; Hochstrasser, R. M. *Springer Ser. Chem. Phys.* **2001**, *67*, 313–325.

(26) Jia, Y.; Talaga, D. S.; Lau, W. L.; Lu, H. S. M.; DeGrado, W. F.; Hochstrasser, R. M. *Chem. Phys.* **1999**, *247*, 69–83.

(27) Talaga, D. S.; Lau, W. L.; Roder, H.; Tang, J.; Jia, Y.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13021–13026.

(28) Andrec, M.; Levy, R. M.; Talaga, D. S. *J. Phys. Chem. A* **2003**, *107*, 7454–7464.

(29) Enderlein, J.; Robbins, D. L.; Ambrose, W. P.; Goodwin, P. M.; Keller, R. A. *Bioimaging* **1997**, *5* (3), 88–98.

(30) Haran, G. *Chem. Phys.* **2004**, *307*, 137–145.

(31) Yang, H.; Xie, X. S. *J. Chem. Phys.* **2002**, *117*, 10965–10979.

(32) Watkins, L. P.; Yang, H. *Biophys. J.* **2004**, *86*, 4015–4029.

(33) Cramér, H. *Mathematical Methods of Statistics;* Princeton University Press: Princeton, NJ, 1946.

(34) Rao, C. R. *Proc. Cambridge Philos. Soc.* **1949**, *45*, 213?218.

(35) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620?630.

(36) Jaynes, E. T. *Probability Theory: The Logic of Science;* Cambridge University Press: Cambridge, 2003.

(37) Jaynes, E. T. *Annu. Rev. Phys. Chem.* **1980**, *31*, 579–601.

(38) Jones, D. S. *Elementary Information Theory;* Oxford University Press: Oxford, 1979.

(39) Jones, G.; Jackson, W.; Halpern, A. *Chem. Phys. Lett.* **1980**, *72*, 391–395.

Single-Molecule Design and Interpretation

J. Phys. Chem. A, Vol. 110, No. 31, 2006 **9757**

(40) Zhuang, X.; Bartley, L. E.; Babcock, H. P.; Russell, R.; Ha, T.; Herschlag, D.; Chu, S. *Science (Washington, DC, U. S.)* **2000**, *288*, 2048−2051.

(41) Margeat, E.; Kapanidis, A. N.; Tinnefeld, P.; Wang, Y.; Mukhopadhyay, J.; Ebright, R. H.; Weiss, S. *Biophys. J.* **2006**, *90*, 1419−1431.

(42) Bopp, M. A.; Jia, Y.; Haran, G.; Morlino, E. A.; Hochstrasser, R. M. *Appl. Phys. Lett.* **1998**, *73*, 7−9.

(43) Forkey, J. N.; Quinlan, M. E.; Goldman, Y. E. *Biophys. J.* **2005**, *89*, 1261−1271.

(44) Osborne, M. A. *J. Phys. Chem. B* **2005**, *109*, 18153−18161.

(45) Deniz, A. A.; Dahan, M.; Grunwell, J. R.; Ha, T.; Faulhaber, A. E.; Chemla, D. S.; Weiss, S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3670−3675.

(46) Ha, T.; Ting, A. Y.; Liang, J.; Caldwell, W. B.; Deniz, A. A.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 893−898.

(47) Zhuang, X.; Kim, H.; Pereira, M. J. B.; Babcock, H. P.; Walter, N. G.; Chu, S. *Science (Washington, DC, U. S.)* **2002**, *296*, 1473−1476.

(48) van Oijen, A. M.; Blainey, P. C.; Crampton, D. J.; Richardson, C. C.; Ellenberger, T.; Xie, X. S. *Science (Washington, DC, U. S.)* **2003**, *301*, 1235−1239.

(49) Lipman, E. A.; Schuler, B.; Bakajin, O.; Eaton, W. A. *Science (Washington, DC, U. S.)* **2003**, *301*, 1233−1235.

(50) Rhoades, E.; Gussakovsky, E.; Haran, G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3197−3202.

(51) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5179−5184.

(52) Kuzmenkina, E. V.; Heyes, C. D.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15471−15476.

(53) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2754−2759.

(54) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938−18943.

(55) Letter analysis of words in the concise Oxford dictionary (9th edition, 1995); http://askoxford.com/asktheexperts/faq/aboutwords/frequency (accessed June 2006).

(56) Shore, J.; Johnson, R. *IEEE Trans. Inf. Theory* **1980**, *26* (1), 26−37.

(57) Gull, S.; Skilling, J. *Proc. IEE* **1984**, *131-F*, 646−659.