

# Photon Arrival-Time Interval Distribution (PAID): A Novel Tool for Analyzing Molecular Interactions

Ted A. Laurence,<sup>\*,†,‡,§,¶</sup> Achillefs N. Kapanidis,<sup>‡</sup> Xiangxu Kong,<sup>‡</sup> Daniel S. Chemla,<sup>†</sup> and Shimon Weiss<sup>\*,‡,§</sup>

Department of Physics, University of California, Berkeley, California 94720-7300,

Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095-1569, and

Department of Physiology, David Geffen School of Medicine at the University of California, Los Angeles, California 90095-1751

Received: August 21, 2003; In Final Form: December 5, 2003

Photon arrival-time interval distribution (PAID) analysis is a new method for monitoring macromolecular interactions. PAID uses fluorescence fluctuations to extract simultaneously coincidence, brightness in multiple channels, diffusion time, and concentration of fluorescently labeled molecules diffusing in a confocal detection volume. PAID is based on recording arrival times of photons detected on one or more detection channels and plotting two-dimensional histograms of photon pairs, where one axis is the time interval between each pair of photons 1 and 2 (in general not consecutive or detected in the same channel) and the second axis is the number of other photons detected (not necessarily detected in the same channels as photons 1 and 2) in the time interval between detection of photons 1 and 2. PAID is related to fluorescence correlation spectroscopy (FCS) by a collapse of the PAID histogram onto the time interval axis. PAID extends auto- and cross-correlation FCS by measuring the brightness of fluorescent species. Studies of fluorescently labeled DNA demonstrate the capabilities of PAID to analyze interactions: PAID detected that DNA carrying two copies of the fluorophore Cy3B is twice as bright as DNA carrying only one copy (simulating dimer vs monomer). PAID also distinguished well between DNA carrying only one fluorophore (Cy3 or Cy5) and DNA carrying both Cy3 and Cy5 (simulating complex vs free components). Solutions containing mixtures of these DNA fragments were successfully analyzed. Studies of *Escherichia coli* RNA-polymerase–DNA interactions demonstrate the ability of PAID to analyze the complex mixtures expected in studies of macromolecular interactions. The statistical accuracy of PAID matches the accuracy of other fluorescence fluctuation methods while providing additional information.

## 1. Introduction

In vitro and in vivo analysis of macromolecular interactions (such as protein–protein, protein–nucleic acid, and protein–ligand interactions) are crucial for understanding fundamental biological mechanisms. Fluorescence fluctuation spectroscopy (FFS), which includes fluorescence correlation spectroscopy (FCS) and related methods, is used for such analysis.<sup>1–7</sup> FFS monitors fluorescence fluctuations due to flow or diffusion of fluorescent species across a femtoliter confocal detection volume (defined by a tightly focused laser excitation profile and a detection pinhole). At concentrations  $C < 1$  nM, the average molecular occupancy of the detection volume (“occupancy”) is smaller than one (“low-occupancy” regime), allowing the detection of photon bursts from single molecules.<sup>8,9</sup> For  $1$  nM  $< C < 100$  nM (“intermediate-occupancy” regime), fluorescence fluctuations are still sensitive to the addition or subtraction of one molecule to or from the detection volume. At  $C > 100$  nM

(“high-occupancy” regime), many molecules occupy the detection volume and the fluctuations are averaged out. For interactions where each partner is fluorescently labeled, this limits the applicability of FFS to interactions with picomolar to nanomolar dissociation constants (unlabeled partners may be present at higher concentrations). Recent methods reduce the detection volume, allowing FFS analysis at micromolar concentrations of fluorophores.<sup>10</sup>

For low- or intermediate-occupancy systems, molecular properties translate into features of the fluorescence signal (Figure 1). Three important characteristics of photon bursts detected as molecules diffuse in and out of the detection volume are (Figure 1A, right panel) the following: (i) duration (*red arrows*), which is proportional to the diffusion time of the species across the detection volume, (ii) brightness (*blue arrows*), which is the product of the laser intensity, the extinction coefficient of the fluorophore, the quantum yield of the fluorophore, the detection efficiency, and the number of identical fluorophores per species (“stoichiometry”), and (iii) the time interval between bursts of a species (*green arrows*), which is inversely proportional to the concentration of the species. Single detection channel (“single-channel”) experiments allow analysis of interactions through measurement of diffusion time and brightness (Figure 1A). Diffusion time can be used to detect interactions since complexes diffuse more slowly than free species; however, large changes in the hydrodynamic volume

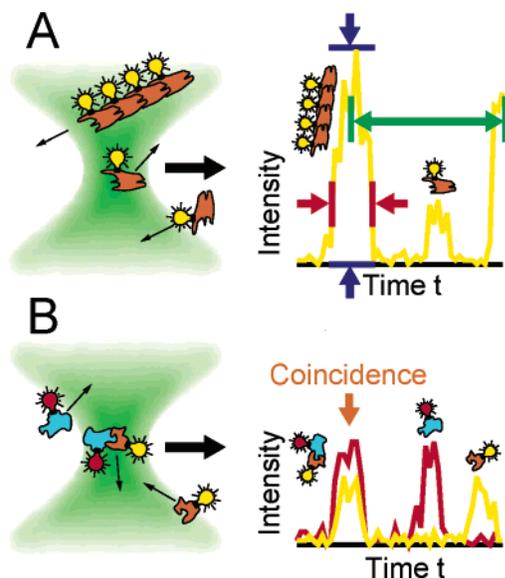
\* Authors to whom correspondence should be addressed. E-mail: laurence2@llnl.gov (T.A.L.); sweiss@chem.ucla.edu (S.W.). Fax: (310) 267-4672 (S.W.).

<sup>†</sup> University of California, Berkeley.

<sup>‡</sup> Department of Chemistry and Biochemistry, University of California, Los Angeles.

<sup>¶</sup> Present Address: Physical Biosciences Institute, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550.

<sup>§</sup> Department of Physiology, David Geffen School of Medicine at the University of California, Los Angeles.



**Figure 1.** Detection of macromolecular interactions using fluorescence fluctuation signals. (A) Left panel: two monomers and a tetramer of a fluorescently labeled species traverse the laser excitation profile (in green). Right panel: corresponding fluorescence bursts along with their major properties, width (red arrows, related to the size of species), height or intensity (blue arrows, related to the number of fluorophores on species), and time interval between bursts from identical species (green arrows, related to the occupancy of species). (B) Left panel: two different molecules (one labeled with a red fluorophore and the other with a yellow fluorophore) interact to form a complex. Right panel: free species emit only in one color; complexes emit in both colors. Coincident detection of red and yellow fluorescence (orange arrow) indicates a complex.

(a function of molecular size and shape) are required to measure binding using diffusion constants (doubling the hydrodynamic volume for spheres leads to only a 26% increase in diffusion time). Brightness can also be used, since a dimeric species has twice the brightness of a free species (ignoring possible quenching). Brightness is a more sensitive reporter of interactions than diffusion, partly because it is shape-independent. However, brightness distributions are widened as a result of varying diffusion paths through the detection volume.

Several FFS methods have been used for the analysis of interactions. FCS<sup>1</sup> analyzes fluorescence fluctuations using correlation functions, reporting on the time scale and amplitude of molecular processes but not on distributions in brightness. Moment analysis of fluorescence-intensity distribution (MAFID) and higher-order correlation amplitudes use moments of photon-counting histograms to monitor occupancy and brightness.<sup>3,4,11</sup> Photon-counting histogram (PCH)<sup>6</sup> and fluorescence intensity distribution analysis (FIDA)<sup>7</sup> fit photon-counting histograms directly (originally developed in ref 5 as fluorescence distribution spectroscopy), allowing more accurate extraction of brightness and occupancy and better identification of species with different brightness.<sup>12</sup> For a single species, FCS and PCH/FIDA performed on the same data set extracts both brightness and diffusion time.<sup>13</sup> Fluorescence intensity multiple distribution analysis (FIMDA) extends FIDA to simultaneously use multiple time bin widths to obtain diffusion time as well as brightness.<sup>14</sup>

In dual detection channel (“dual-channel”) experiments, molecule 1 is labeled with one fluorophore (Figure 1B, *yellow bulbs*) and molecule 2 is labeled with a second fluorophore with distinct emission (Figure 1B, *red bulbs*); a complex of the two molecules carries both labels, resulting in simultaneous photon bursts on both channels (Figure 1B, *orange arrow*). Dual-channel experiments are more sensitive reporters of binding than

single-channel ones<sup>2,15–17</sup> because coincident bursts appear only as a result of an interaction. To detect simultaneously free species and complex (dimer), a single-channel experiment must differentiate species that differ by a factor of 2 in brightness using intensity distributions which are widened as a result of varying paths through the detection volume. In contrast, dual-channel experiments need only detect differences in the path-independent ratio between the intensity of two channels.<sup>16</sup>

There are several existing methods for analyzing dual-channel experiments. Fluorescence cross-correlation spectroscopy (FCCS) monitors interactions between molecules labeled with two fluorophores with distinct emission using the cross-correlation.<sup>2,18</sup> Multiparameter fluorescence detection (MFD) is a low-occupancy, single-molecule burst analysis method that measures fluorescence lifetime, ratiometric expressions,<sup>17,19,20</sup> and brightness.<sup>21</sup> Single-molecule fluorescence ratiometric methods monitor fluorescence resonance energy transfer (FRET) (if fluorophores are spaced by 2–8 nm), fluorescence anisotropy, and spectral fluctuations.<sup>16,22</sup> Two-dimensional fluorescence intensity distribution analysis (2D-FIDA) extends FIDA to the joint photon-counting histogram for two channels, extracting occupancy and brightness in both channels but not diffusion times or temporal dynamics.<sup>15</sup>

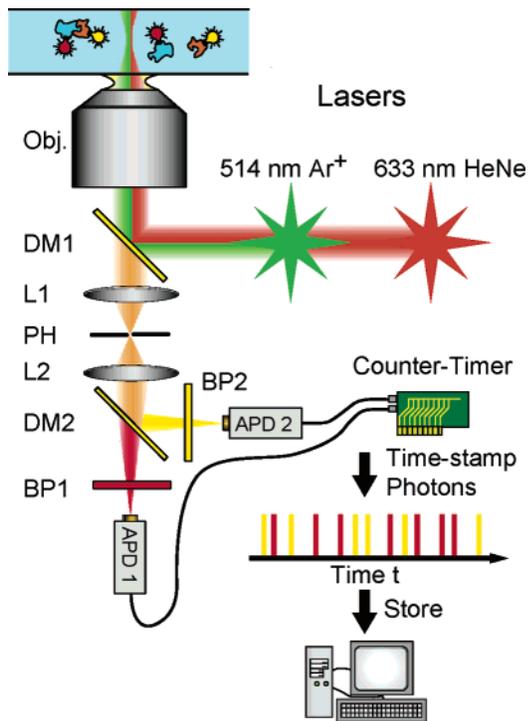
Here, we introduce photon arrival-time interval distribution (PAID), a new *multidimensional* data-analysis and data-representation method that simultaneously measures occupancy, diffusion (and other temporal fluctuations), and brightness of several species on *multiple detection channels*. In contrast to previous methods, PAID emphasizes photon-rich time intervals when fluorescent molecules are present in the detection volume, since it is based on observing time intervals between photons (not necessarily consecutive) rather than counting photons on evenly spaced intervals; this feature allows intuitive visual interpretation of the data. PAID also provides “backward compatibility”, since the collapse of the PAID histogram onto the time axis provides the familiar FCS correlation function. We characterized the ability of single- and dual-channel PAID to analyze distinct species using both simulations and solution-based experiments. Analysis using fluorescently labeled DNA model systems and a protein–DNA interaction showed that PAID is capable of evaluating both stoichiometry and affinity of interactions.

## 2. Theory

**2.1. Development and Description of PAID.** PAID analyzes streams of photon-arrival times rather than sequences of photon counts. We obtain photon-arrival times using a dual-channel microscope (Figure 2) with avalanche photodiode detectors (APDs); we represent the resulting stream of detected photons as a sum of Dirac delta functions:

$$I_A(t) = \sum_{i=1}^{N_A} \delta(t - t_{Ai}) \quad (1)$$

Here,  $t_{Ai}$  is the arrival time of the  $i$ th photon from APD-channel  $A$  and  $N_A$  is the number of photons detected in channel  $A$ . The number of photons arriving between times  $t_{\min}$  and  $t_{\max}$  is  $\mathbf{n} = \int_{t_{\min}}^{t_{\max}} dt I_A(t)$ . The arrival time  $t_{Ai}$  of each photon is recorded as an integer  $\mathbf{t}_{Ai}$  with  $\mathbf{t}_{Ai} = [t_{Ai}/\Delta t]$ , (brackets indicate the greatest integer function, e.g.,  $[3.14] = 3$ ). The main source of uncertainty in specifying arrival times is the resolution of the digital clock used for timing photons ( $\Delta t = 12.5$  ns for a National Instruments 6602 counting board), which is shorter than the  $>1$   $\mu$ s time scale studied here. For an experiment of



**Figure 2.** Instrumentation for two-color cross-correlation measurements. Two laser sources (514 and 633 nm) are introduced into an inverted microscope, reflected by a dichroic mirror (DM1), and focused by a 1.3 NA oil-immersion objective into a sample cell, assembled using two glass coverslips separated by a 1 mm thick silicone gasket. Fluorescence from the sample is collected through the objective, transmitted through DM1, and focused onto a pinhole (PH) by the microscope tube lens (L1). After a second lens (L2), the fluorescence is split by a dichroic mirror (DM2) into red and yellow channels. Red and yellow fluorescence pass through band-pass filters (BP1 and BP2, respectively) and are detected using APD1 and APD2, respectively. The APDs emit an electronic pulse for every photon detected, and the pulses are time-stamped using a counter-timer board and stored in a PC (the timed photons are represented in the figure as bars where the color of the bar represents whether the red or yellow channel detected the photon).

duration  $T$ ,  $t_{Ai}$  will be in the range  $0, 1, \dots, T = [T/\Delta t]$ , converting eq 1 to

$$I_A(t) = \sum_{i=1}^{N_A} \delta(t, t_{Ai}) \quad (2)$$

where  $\delta$  is the Kronecker delta:  $\delta(t, t_{Ai}) = 1$  if  $t = t_{Ai}$ , and  $\delta(t, t_{Ai}) = 0$  if  $t \neq t_{Ai}$ .

In FCS, the correlation function measures the relationship between a photon stream at one time  $I_S(t)$  and a second photon stream at a later time  $I_T(t + \tau)$  as a function of the time interval  $\tau$ ,

$$C_{ST}(\tau) = \langle I_S(t) I_T(t + \tau) \rangle / \langle I_S(t) \rangle \langle I_T(t) \rangle \quad (3)$$

If  $S = T$ , eq 3 defines an autocorrelation; if  $S \neq T$ , eq 3 defines a cross-correlation. Assuming a stationary stochastic process, the ensemble averages can be evaluated as time averages:  $\langle \dots \rangle \rightarrow \lim_{T \rightarrow \infty} 1/T \sum_{t=0}^T (\dots)$ . Substituting eq 2 into eq 3, we obtain an estimator for the correlation (the hat denotes an estimator),

$$\hat{C}_{ST}(\tau) = T \sum_{i=1}^{N_S} \sum_{j=1}^{N_T} \delta(\tau, t_{Tj} - t_{Si}) / N_S N_T \quad (4)$$

We see that the correlation is estimated by a histogram of *time intervals* between photon arrivals: each pair of photons  $t_{Si}$  and  $t_{Tj}$  is added to a histogram at the bin corresponding to the interval  $\tau = t_{Tj} - t_{Si}$ . We can specify  $\tau \geq 0$  without loss of generality, since swapping the roles of  $I_S$  and  $I_T$  is equivalent to changing the sign of  $\tau$ . Photons from  $I_T$  stop the intervals, so we define  $I_S$  and  $I_T$  to be the “start” and “stop” photon streams respectively; all quantities associated with  $I_S$  and  $I_T$  are labeled with the subscripts  $S$  and  $T$ , respectively.

PAID adds a photon-counting dimension  $\mathbf{n}$  to the correlation function that includes brightness information missing in the correlation function. Previous methods<sup>5-7,14,15</sup> count photons on evenly spaced time intervals. In contrast, PAID directly extends the correlation in eq 4 by counting photons on the time intervals between “start” photon times  $t_{Si}$  and “stop” photon times  $t_{Tj}$ . This way of choosing the time intervals adapts the spacing of the counted time intervals to the photon stream, focusing the histogram on “photon-rich” times. More importantly, by varying  $S$  and  $T$ , as well as the channel on which the photons are counted, PAID can combine brightness information from more than one channel with the temporal information in the correlation. Choosing different channels  $S$  and  $T$  emphasizes species that emit in both  $S$  and  $T$  (see sections 4.1 and 4.4). This allows occupancy, diffusion time, and brightness in more than one channel to be extracted simultaneously for multiple species.

For a single-channel experiment ( $S = T$ ), the number of photons counted between photons  $i$  and  $j$  is  $\mathbf{n} = j - i - 1$ . The PAID histogram classifies the photon pairs in eq 4 by the time interval  $\tau = t_j - t_i$  between start and stop photons *and* by the number of photons  $\mathbf{n} = j - i - 1$  counted between them,

$$\hat{C}_{SSS}(\tau, \mathbf{n}) = T \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \delta(\tau, t_{Sj} - t_{Si}) \delta(\mathbf{n}, j - i - 1) / N_S^2 \quad (5)$$

By rewriting  $\sum_{j=1}^{N_S} \delta(\tau, t_{Sj} - t_{Si}) \delta(\mathbf{n}, j - i - 1) = \delta(\tau, t_{S, i+\mathbf{n}+1} - t_{Si})$  in eq 5, we see that  $\hat{C}_{SSS}(\tau, \mathbf{n}) = \hat{P}_{\mathbf{n}+1}(\tau) / N_S T$ , where  $\hat{P}_{\mathbf{n}+1}(\tau) = \sum_{i=1}^{N_S} \delta(\tau, t_{S, i+\mathbf{n}+1} - t_{Si}) / N_S$  is the histogram of waiting times to the  $(\mathbf{n} + 1)$ th photon,  $t_{S, i+\mathbf{n}+1} - t_{Si}$ , for  $\mathbf{n} = 0, 1, \dots$  (Waiting-time histograms for  $\mathbf{n} = 0$  have been used as approximations for the correlation at short times,<sup>23</sup> and histograms for  $\mathbf{n} > 0$  have been used in criteria for sifting for bursts<sup>21</sup>.) By summing eq 5 over  $\mathbf{n}$  we obtain the autocorrelation of eq 4,

$$\hat{C}_{SS}(\tau) = \sum_{\mathbf{n}=0}^{\infty} \hat{C}_{SSS}(\tau, \mathbf{n}) \quad (6)$$

restating the result that the sum of the waiting-time distributions gives the autocorrelation function.<sup>23</sup>

For a Poisson process with constant rate  $k = N_S/T$ , the waiting-time distributions are Gamma distributions ( $P_{\mathbf{n}+1}(\tau) = k^{\mathbf{n}+1} \tau^{\mathbf{n}} \exp(-k\tau) / \mathbf{n}!$ ), so the PAID function is  $C_{SSS}(\tau, \mathbf{n}) = k^{\mathbf{n}+1} \tau^{\mathbf{n}} \exp(-k\tau) / \mathbf{n}!$ . Using eq 6 (true for the function as well as the estimator), we see that the correlation function is constant,  $C_{SS}(\tau, \mathbf{n}) = 1$ , as expected for the uncorrelated Poisson process.

We now generalize the PAID histogram in eq 5 to more than one channel by defining three roles for the photons (the simple relationship  $\mathbf{n} = j - i - 1$  no longer applies). We use the start and stop channel definitions from eq 4, but we also count the number of photons  $\mathbf{n}_M$  that arrive between times  $t_{Si}$  and  $t_{Tj}$  in an additional photon stream  $I_M(t)$  (the “monitor” photon stream), with photon arrival times  $t_{Mk}$ :

$$n_M(\mathbf{t}_{Si}, \mathbf{t}_{Tj}) \equiv \sum_{t=\mathbf{t}_{Si}+1}^{\mathbf{t}_{Tj}-1} I_M(t) \quad (7)$$

$I_M(\mathbf{t})$  is formally distinct from  $I_S$  and  $I_T$  but can be set to be identical to one or both. When the photon count dimension is added to eq 3, the general PAID function for more than one channel is

$$C_{STM}(\boldsymbol{\tau}, \mathbf{n}) = \langle I_S(\mathbf{t}) I_T(\mathbf{t}) \delta[\mathbf{n}, \mathbf{n}_M(\mathbf{t}, \mathbf{t} + \boldsymbol{\tau})] \rangle / \langle I_S(\mathbf{t}) \rangle \langle I_T(\mathbf{t}) \rangle \quad (8)$$

This function can be graphed in two dimensions (a discrete axis  $\mathbf{n}$  and a continuous axis  $\boldsymbol{\tau}$ ), where the additional  $\mathbf{n}$  axis slices the correlation into strips with different numbers of counted monitor photons along the vertical axis. Rather than a photon count probability distribution (the photon-counting histogram of PCH/FIDA), the PAID function acts as a photon-count distribution of correlations. For an experiment of duration  $\mathbf{T}$ , eq 8 can be estimated by a digital double summation histogram,

$$\hat{C}_{STM}(\boldsymbol{\tau}, \mathbf{n}) = \mathbf{T} \sum_{i=1}^{N_S} \sum_{j=1}^{N_T} \delta(\boldsymbol{\tau}, \mathbf{t}_{Tj} - \mathbf{t}_{Si}) \delta[\mathbf{n}, \mathbf{n}_M(\mathbf{t}_{Si}, \mathbf{t}_{Tj})] / N_S N_T \quad (9)$$

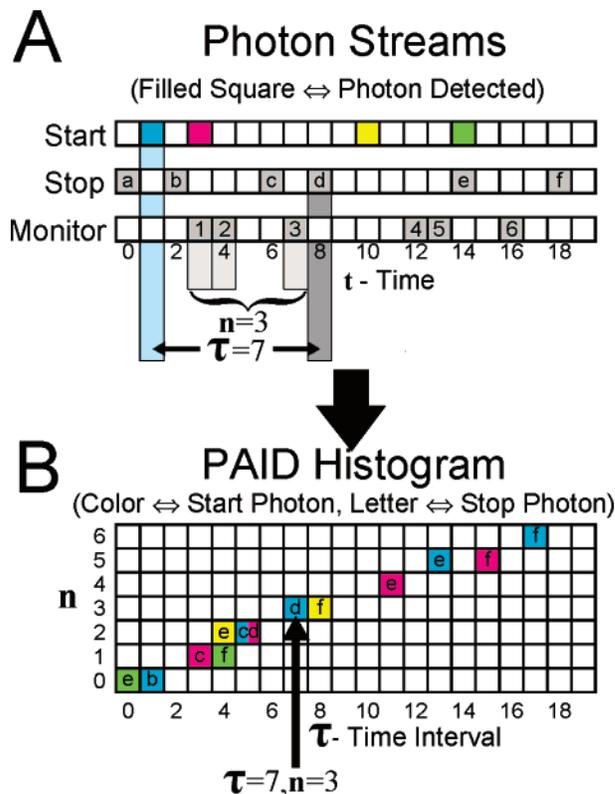
For each pair of photons  $\mathbf{t}_{Si}$  and  $\mathbf{t}_{Tj}$ , an event is placed in a 2D histogram, where one axis is the time interval  $\boldsymbol{\tau} = \mathbf{t}_{Tj} - \mathbf{t}_{Si}$ , and the other axis is the number of monitor photons counted  $\mathbf{n} = \mathbf{n}_M(\mathbf{t}_{Si}, \mathbf{t}_{Tj})$ . Figure 3 shows how a PAID histogram is formed from the start, stop, and monitor photon streams using eq 9. To capture a wide dynamic range, log or quasi-log axes are used for the histograms (Supporting Information sections 1S and 2S). By summing over  $\mathbf{n}$  and comparing to eq 4, we obtain

$$\hat{C}_{ST}(\boldsymbol{\tau}) = \sum_{\mathbf{n}=0}^{\infty} \hat{C}_{STM}(\boldsymbol{\tau}, \mathbf{n}) \quad (10)$$

showing that the cross-correlation histogram is the collapse of  $\hat{C}_{STM}(\boldsymbol{\tau}, \mathbf{n})$  onto the time interval  $\boldsymbol{\tau}$  axis.

Equation 9 is extended to include higher-order temporal correlations, or more monitor photon streams, by adding factors of the form  $\delta(\boldsymbol{\tau}_2, \mathbf{t}_{T2j} - \mathbf{t}_{S2i})$  for temporal correlations and factors of the form  $\delta(\mathbf{n}_2, \mathbf{n}_{M2}(\mathbf{t}_{S2i}, \mathbf{t}_{T2j}))$  for additional monitor photon streams ( $S_2$ ,  $T_2$ , and  $M_2$  for additional formal photon streams;  $\mathbf{n}_2$  and  $\boldsymbol{\tau}_2$  for additional photon-counting and time interval variables). Higher-order temporal correlations are useful for monitoring non-Markovian dynamics,<sup>24</sup> and multiple monitor photon count axes in a single histogram take better advantage of dual-channel measurements (see Supporting Information section 7.2S).

**2.2. Application of PAID to Interacting Fluorescent Species Diffusing in Solution.** When fluorescent molecules diffusing in solution are excited and detected using a single-molecule epi-fluorescence confocal microscope (Figure 2), the fluorescence is split into different channels (denoted  $A$ ) that detect spectral regions matching the emission spectra of the fluorophores.  $\text{Exc}(\bar{x})$  is the laser excitation profile, and  $\text{CEF}(\bar{x})$  is the collection efficiency function of the collection optics.<sup>25</sup> The detectivity,  $\phi(\bar{x}) \equiv \text{Exc}(\bar{x})\text{CEF}(\bar{x})$ , is used to calculate the effective detection volume,  $V_{\text{eff}} = (\int dV \phi(\bar{x}))^2 / \int dV \phi^2(\bar{x})$ . For our experiments, we use a numerically calculated detectivity (section 3.5). For simulations, we use a Gaussian detectivity,  $\phi(\bar{x}) = \exp[-2(x^2 + y^2)/\omega^2 - 2z^2/l^2]$ , where  $\omega$  is the  $1/e^2$  width in the  $x$  and  $y$  directions and  $l$  is length in the  $z$  direction ( $V_{\text{eff}} = \pi^{3/2}\omega^2 l$ ).



**Figure 3.** Generation of a PAID histogram. (A) Photon streams from three channels (start, stop, and monitor channels). The start, stop, and monitor channels are distinct in this example, which is not necessary for the formation of PAID histograms. In fact, for a single-channel experiment, they are all necessarily identical; for a dual-channel experiment, at least two are identical. Time axis is shown in discrete clock units. A filled square indicates photon detection at that specific time position. Start photons are identified with different colors, stop photons with different letters, and monitor photons with different numbers. Each photon in the start channel is paired with each stop-channel photon occurring later in time. (B) PAID histogram. The  $x$ -axis is the time interval between the start and stop photons; the  $y$ -axis is the number of monitor photons counted in the time interval between the start and stop photons. Example photon pair: the blue start photon is paired with stop photon “d”. The time interval between these two photons is 7 clock units, and there are three monitor photons (numbers 1, 2, and 3) between them. The corresponding entry into the histogram in part B is coded with the blue background and the letter “d”.

For  $F$  freely diffusing, fluorescent species in solution, we specify the species by an index  $\alpha = 1 \dots F$ ; we associate  $\alpha = 0$  with a constant background exhibiting Poisson statistics. The primary parameters of a diffusing species  $\alpha$  extracted from fluorescence fluctuation measurements are the following: (1)  $c_\alpha$ , the occupancy or average number of molecules in  $V_{\text{eff}}$ ; the concentration  $[\alpha]$  is obtained by dividing  $[\alpha] = c_\alpha/V_{\text{eff}}$ . (2)  $\tau_\alpha^D$ , the diffusion time of a molecule across  $V_{\text{eff}}$ , or the time at which the autocorrelation amplitude decays by a factor of 2 (excluding other fluctuations); for a Gaussian detection volume with  $l > \omega$ , the “lateral diffusion time”,  $\tau_\alpha^D = \omega^2/4D_\alpha$ , is used, where  $D_\alpha$  is the diffusion constant. (3)  $q_{\alpha A}$ , the brightness or count rate per molecule averaged over  $V_{\text{eff}}$  in channel  $A$  (the average count rate for species  $\alpha$  in channel  $A$  is  $k_{\alpha A} = c_\alpha q_{\alpha A}$ ). For background,  $k_{0A}$  is the count rate in channel  $A$ . PAID currently extracts  $c_\alpha$ ,  $\tau_\alpha^D$ ,  $q_{\alpha A}$ , and  $k_{0A}$  (other properties not yet extracted using PAID include rotational diffusion,<sup>26</sup> intersystem crossing,<sup>27</sup> and photobleaching<sup>28,29</sup>).

PAID is used to study macromolecular interactions by extracting occupancy, diffusion time, and brightness for interacting species. Since occupancies are directly proportional to the

concentration of species, they can be used to assess the affinity of an interaction; for example, for a simple interaction with equilibrium  $A + B \rightleftharpoons AB$ , varying  $[A]$  and  $[B]$  and measuring occupancies of  $A$ ,  $B$ , and  $AB$  using PAID allows the calculation of the dissociation constant<sup>30</sup>  $K_d = [A][B]/[AB]$ . At nanomolar concentrations generally used for FFS studies, association or dissociation events are rare within the diffusion time scale. Hence, we only observe  $A$ ,  $B$ , and  $AB$  (but not any association or dissociation events) and can treat the diffusing species as “static” species. An upper bound of  $k_{\text{encounter}} < 10 \text{ s}^{-1}$  for the association rate is found using diffusion-limited encounter rates between equal-sized macromolecules  $A$  and  $B$ <sup>30</sup> present at 1 nM concentrations. This is greater than 100-fold slower than typical diffusion times. Since  $K_d = k_d/k_a$ , where  $k_a$  is the association rate constant and  $k_d$  is the dissociation rate,  $k_d < 10 \text{ s}^{-1}$  to have a significant population of  $AB$ .

**2.3. PAID Function Model.** We now briefly describe how we model the PAID function of eq 8 for data fitting (described in more detail in Supporting Information section 3S). A model for the single-channel PAID function of a constant background exhibiting Poisson statistics was given in section 2.1. A possible extension of that model is a renewal process, where successive time intervals  $t_i - t_{i-1}$  and  $t_{i+1} - t_i$  are independent but have identical distributions,  $P(\tau)$ . For fluorescent species diffusing in solution, however, count rates depend more on the diffusion of molecules through the confocal detection volume than on the time since the previous photon; a renewal process would be an inadequate model for this situation.

We model the photon emission statistics as a Poisson process with a stochastic, fluctuating rate  $k_A(t)$  for each channel  $A$ , appropriate if the photons antibunching<sup>31</sup> is neglected. Additionally, we ignore intersystem crossing to triplet states and photobleaching. For the rate  $k_A(t)$ , the probability for the number of photons counted up to time  $t$  follows the series of differential equations  $\partial P_A(t, \mathbf{n}) / \partial t = k_A(t)P_A(t, \mathbf{n} - 1) - k_A(t)P_A(t, \mathbf{n})$ . Solving, we obtain the Poisson distribution,<sup>32,33</sup>

$$P_A(t, \mathbf{n} | t_0) = \exp[-K_A(t_0, t)] \frac{[K_A(t_0, t)]^{\mathbf{n}}}{\mathbf{n}!} \equiv \text{Poi}(K_A(t_0, t), \mathbf{n}) \quad (11)$$

where we define  $K_A(t_0, t) = \int_{t_0}^t dt' k_A(t')$  as the cumulative intensity. Equation 11 only accounts for the “shot noise” intrinsic to the counting process; varying  $k_A(t)$  widens this distribution further.

We now apply eq 11 to the PAID function in eq 8, where the ensemble averages are performed first to account for shot noise, then for the varying rates  $\langle \dots \rangle = \langle \dots \rangle_{\text{shot}}|_{\text{rate}}$ .  $k(t)$  and its integral  $K_A(t_0, t)$  are functions of continuous time, and will not be written using discrete time variables (we will use  $t = t\Delta t$  and  $\tau = \tau\Delta t$ ). Using the rates  $k_S(t)$  and  $k_T(t)$ , and  $K_M(t_0, t) = \int_{t_0}^t dt' k_M(t')$ , eq 8 is converted to

$$C_{STM}(\boldsymbol{\tau}, \mathbf{n}) = \langle k_S(0)k_T(\tau) \text{Poi}[K_M(\Delta t, \tau), \mathbf{n}] \rangle_{\text{rate}} / \bar{k}_S \bar{k}_T \quad (12)$$

where  $\bar{k}_S = \langle k_S(t) \rangle_{\text{rate}}$  and  $\bar{k}_T = \langle k_T(t) \rangle_{\text{rate}}$  are average rates. Stationarity is assumed, so  $t$  is set to 0.  $K_M(\Delta t, \tau)$  integrates  $k_M(t)$  from continuous times  $\Delta t$  to  $\tau$ , converting from the expression in eq 7.

Each rate  $k_A(t)$  is split into contributions from background and each of  $\mathcal{N}_\alpha$  independently diffusing and emitting molecules from each species  $\alpha = 1 \dots F$  in the sample volume  $\mathcal{V}_{\text{sample}}$  (concentration is  $[\alpha] = \mathcal{N}_\alpha / \mathcal{V}_{\text{sample}} = c_\alpha / V_{\text{eff}}$ );  $k_A(t) = \sum_{\alpha=0}^F \sum_{i=1}^{\mathcal{N}_\alpha} k_{\alpha i} A(t) \equiv \sum_{(\alpha, i)} k_{\alpha i} A(t)$ .  $\alpha = 0$  denotes the constant background. Using the rates  $k_{\alpha i A}(t)$ , eq 12 simplifies to a sum of convolutions of expressions for single molecules,

$$C_{STM}(\boldsymbol{\tau}, \mathbf{n}) = \int_0^\infty dK_M \text{Poi}(K_M, \mathbf{n}) \left\{ \sum_\alpha \mathcal{N}_\alpha f_{\alpha STM}^{(1,1)}(K_M | \tau) \right\} \prod_\gamma [f_{\gamma STM}^{(0,0)}(K_M | \tau)]^{\lambda_\gamma} + \sum_{\alpha, \beta} \mathcal{N}_\alpha \mathcal{N}_\beta f_{\alpha STM}^{(1,0)}(K_M | \tau) f_{\beta STM}^{(0,1)}(K_M | \tau) \prod_\gamma [f_{\gamma STM}^{(0,0)}(K_M | \tau)]^{\lambda_\gamma} / \bar{k}_S \bar{k}_T \quad (13)$$

The asterisk (\*) inside the brackets indicates repeated convolutions (not multiplications). The single-molecule expressions,  $f_{\alpha STM}^{(p,q)}(K_M | \tau) = \int_0^\infty dk_S \int_0^\infty dk_T P_{\alpha STM}(k_S, k_T, K_M | \tau) k_S^p k_T^q$ , are integrals over  $P_{\alpha STM}(k_S, k_T, K_M | \tau)$ , which is the joint probability distribution for a molecule of species  $\alpha$  to have count rate  $k_S$  in  $S$  at time lag 0, count rate  $k_T$  in  $T$  at time lag  $\tau$ , and integrated intensity  $K_M$  between time lags  $\Delta t$  and  $\tau$ . Inside the braces in eq 13, the first sum contains the correlated contribution,  $f_{\alpha STM}^{(1,1)}$ , accounting for the emission of each molecule at time lags 0 and  $\tau$ ; the second sum contains the uncorrelated contribution,  $f_{\alpha STM}^{(1,0)} * f_{\beta STM}^{(0,1)}$ , accounting for the emission of one molecule at time lag 0 and another at  $\tau$ . The repeated convolutions  $\prod_\gamma [f_{\gamma STM}^{(0,0)}]^{\lambda_\gamma}$  account for the effects of the uncorrelated molecules on the distribution in  $K_M$ . The integration over  $K_M$  against the  $\text{Poi}(K_M, \mathbf{n})$  converts the expression from cumulative intensity space to photon count space, accounting for shot noise.

The expressions  $f_{\alpha STM}^{(p,q)}(K_M | \tau)$  are approximated by Monte Carlo simulation of possible diffusion paths through  $V_{\text{eff}}$ .<sup>34</sup>  $V_{\text{eff}}$  may be set as an analytical Gaussian detectivity, as a numerical approximation, or as an experimentally measured detectivity. A scaling law is used to model changes in brightness or diffusion time; if we change variables,  $K_M^0 = K_M / q_{\alpha M} \tau_\alpha^D$  and  $\tau^0 = \tau / \tau_\alpha^D$ , we find  $f_{\alpha STM}^{(p,q)}(K_M | \tau) dK_M = q_{\alpha S}^p q_{\alpha T}^q f_{\alpha STM}^{(p,q)}(K_M^0 | \tau^0) dK_M^0$  (see Supporting Information section 3S). Once the function  $f_{\alpha STM}^{(p,q)}(K_M^0 | \tau^0)$  (where  $q_{\alpha S}^0 = q_{\alpha T}^0 = q_{\alpha M}^0 = \tau_\alpha^{D0} = 1$ ) is known, the function for any other parameter values can be obtained by scaling the variables according to the definitions of  $\tau^0$  and  $K_M^0$ .

The convolutions in eq 13 consume most of the time needed to calculate the PAID function model. Because of the wide temporal and dynamic ranges over which fluorescence fluctuations occur, log axes are desirable. For the most efficient calculation, the convolution method used must work in a log domain (standard fast Fourier transform (FFT) methods use linearly spaced data). We developed a novel method that combines the use of the FFT with a quasi-log scale, making the model calculation practical (Supporting Information section 3.2S). On a 1.2 GHz Pentium 3 based PC, the simultaneous fits of the 8 possible dual-channel PAID histograms with 14 free parameters shown in Table 6 took 3 min per iteration (with  $\sim 10$  iterations per fit), whereas the fits for the single-channel PAID histograms in Table 1 took 10 s per iteration (with  $\sim 10$  iterations per fit). Narrowing the range of time intervals used or decreasing the number of fitted parameters reduces fitting time substantially, without compromising the ability of PAID to monitor interactions. For example, by reducing the time range of interest from a range of 1  $\mu\text{s}$  to 1 s to a range of 10  $\mu\text{s}$  to 10 ms, reducing the time bin spacing from 10 bins per decade to 5 bins per decade, simultaneously fitting only half of the 8 possible two-channel PAID histograms, and reducing the resolution in the convolution algorithm by a factor of 2, the fits from Table 6 take 10 s per iteration (with  $\sim 10$  iterations per fit). The resulting fitted parameters typically differ by  $< 10\%$ .

### 3. Materials and Methods

**3.1. Preparation of DNA.** We used fluorescently labeled DNA fragments as model systems for exploring the capabilities

of PAID. The fluorophores used were Cy3B ( $\lambda_{\text{ex}} \sim 560$  nm,  $\lambda_{\text{em}} \sim 580$  nm), Cy3 ( $\lambda_{\text{ex}} \sim 550$  nm,  $\lambda_{\text{em}} \sim 570$  nm), and Cy5 ( $\lambda_{\text{ex}} \sim 650$  nm,  $\lambda_{\text{em}} \sim 670$  nm). Six DNA fragments were synthesized (the nomenclature used in the superscript report the fluorophore, the DNA position, and the DNA strand where the fluorophore was introduced, i.e., T is the top strand and B is the bottom strand): (1) DNA<sup>Cy3B,1T</sup>, (2) DNA<sup>Cy3B,1T/Cy3B,65B</sup>, (3) DNA<sup>Cy5,1T</sup>, (4) DNA<sup>Cy3,65B</sup>, (5) DNA<sup>Cy5,1T/Cy3,65B</sup>, and (6) DNA<sup>Cy5,65B</sup>. DNA fragments were prepared using standard polymerase chain reaction (PCR) protocols<sup>35</sup> with one or two 5'-labeled DNA primers, followed by purification using non-denaturing gel electrophoresis. The sequence of the top strand was 5'-AGGCTTTACACTTTATGCTTCCGGCTCGTATAA-TGTGTGGAATTGTGAGAGCGGATAACAATTTTC-3'. We prepared mixtures of DNA<sup>Cy3B,1T</sup> and DNA<sup>Cy3B,1T/Cy3B,65B</sup> for single-channel applications and mixtures of DNA<sup>Cy5,1T</sup>, DNA<sup>Cy3,65B</sup>, and DNA<sup>Cy5,1T/Cy3,65B</sup> for dual-channel applications. In all cases, the intramolecular separation between the two fluorophores is large (65 bp (base pairs),  $\sim 240$  Å) to preclude FRET. Data were acquired for 5 min using 30 pM to 1 nM DNA in 20 mM HEPES–NaOH (pH 7), 50 mM NaCl, 5% glycerol, and 1 mM mercaptoethylamine (Fluka, Milwaukee, WI). We added 0.1% BSA (Panvera, Madison, WI) to the DNA<sup>Cy3B,1T/Cy3B,65B</sup> and DNA<sup>Cy3B,1T</sup> samples to reduce adsorption onto surfaces. The concentration of DNA<sup>Cy5,1T</sup> was determined using UV–vis spectrophotometry and fluorescence spectrophotometry for calibration. Higher-concentration (10–25 nM) samples were prepared for all DNA fragments, and occupancies and diffusion times were extracted using FCS as a basis for dilutions.

**3.2. Preparation of RNA-Polymerase (RNAP) and RNAP–DNA Complex.** *Escherichia coli* RNAP core was purchased from Epicentre (Milwaukee, WI), and RNAP  $\sigma$  subunit ( $\sigma^{\text{Cys596}}$ ) was purified and labeled at amino acid Cys596 using tetramethylrhodamine (TMR) as described.<sup>36</sup> RNAP holoenzyme and RNAP–DNA complexes were formed essentially as described,<sup>36</sup> using DNA<sup>Cy5,65B</sup> for the formation of the complex; the complex was diluted to 1 nM nominal concentration for the PAID measurements. The large distance ( $\gg 100$  Å) between TMR and Cy5 in the RNAP–DNA open complex precludes FRET between the fluorophores.<sup>37</sup> After formation of the RNAP–DNA complex, the sample was loaded in nondenaturing 5% polyacrylamide gels and was electrophoresed at 10 V/cm for 1 h; the resulting gels were imaged using an x–y fluorescence imager (Molecular imager FX, Biorad, Hercules, CA) equipped with 532 nm and 633 nm excitation lasers (for excitation of “yellow” and “red” fluorophores respectively) and 585BP60 and 640LP emission filters (for detecting “yellow” and “red” channel emissions, respectively).

**3.3. Confocal Fluorescence Microscopy.** The instrumentation used is similar to that used in refs 16 and 22 (Figure 2). For single-channel experiments, the 532 nm line from a solid-state pumped Nd:YAG laser (GCL-100-S, Crystalaser, Reno, NV; 100 kW/cm<sup>2</sup>; excites Cy3 and Cy3B) was introduced using fiber optics and reflected by an excitation dichroic mirror (DM1; 400–535–635 TBDR, Omega Optical, Brattleboro, VT). For the dual-channel experiments, two laser beams are introduced: 514 nm Ar<sup>+</sup> (543-A-A02, Melles-Griot, Carlsbad, CA; 200 kW/cm<sup>2</sup>; excites Cy3 and TMR, and to a much lesser degree, Cy5) and 633 nm HeNe (05-LHP-171, Melles-Griot, Carlsbad, CA; 66 kW/cm<sup>2</sup>; excites Cy5), and a different excitation dichroic mirror (DM1; 390–510–630 TBDR, Omega Optical, Brattleboro, VT) was used. For both experiments, the laser excitation is focused 20 mm inside the solution by a 100  $\times$  1.3 NA Zeiss Neofluar

oil-immersion objective. Fluorescence is focused by the microscope tube lens (L1) on a 100  $\mu\text{m}$  pinhole (PH) and split with a dichroic mirror (DM2; 630 DMLP, Omega Optical) into two detection channels: the Cy5 channel (filtered using a 650LP filter (BP1), Omega Optical) and the TMR/Cy3/Cy3B channel (filtered using a 580DF60 filter (BP2), Omega Optical). Silicon avalanche photodiodes (APD1 and APD2, SPCM-AQR-14, Perkin Elmer, Vaudreuil, QB, Canada) detect fluorescence photons; a counter-timer board (PCI-6602, National Instruments, Austin, TX) times photon-associated electronic pulses.

**3.4. Simulation of Translational Diffusion of Molecules and Photon Emission and Detection.** Simulations for translational diffusion of molecules, photon emission, and detection were performed similarly to ref 38. A Gaussian detectivity with  $\omega = 0.35$   $\mu\text{m}$ ,  $l = 1.75$   $\mu\text{m}$  is placed at the center of a 3D simulation box with size  $V_{\text{box}} = 3.5 \times 3.5 \times 17.5$   $\mu\text{m}^3$ , assuming periodic boundary conditions (a molecule that leaves  $V_{\text{box}}$  reappears at the opposite side with the same lateral position); a fixed number of molecules is placed inside the box. Diffusion is simulated by a series of steps with  $\Delta t = 1$   $\mu\text{s}$ , during which the excitation rate is kept constant (the diffusion time  $\tau^D \geq 100$   $\mu\text{s}$ ). At each step, the distance changes for  $x$ ,  $y$ ,  $z$  are determined by Gaussian random numbers, with mean  $\mu = 0$  and standard deviation  $\sigma = (2D\Delta t)^{1/2}$ , where  $D$  is the diffusion constant. Until time  $\Delta t$  is passed, waiting times between a series of photon detection events are generated using exponential random numbers with a constant rate  $\lambda$  that depends on the excitation rate at the molecule's position; the fluorescence lifetime is neglected, as are other photophysical effects. A uniform random number determines which channel detects the photon.

**3.5. Calculation of Detection Volume and Diffusion Parameters.** A numerically approximated, non-Gaussian volume is used for PAID analysis to account for our experimental conditions. The laser excitation profile,  $\text{Exc}(\vec{x})$ , for an oil-immersion objective focused 20  $\mu\text{m}$  inside the aqueous solution was calculated using Monte Carlo integration of plane-wave contributions using expressions from ref 39, simplified to scalar diffraction, and accounting for the water-glass dielectric interface.<sup>40</sup> The effect of the pinhole,  $\text{CEF}(\vec{x})$ , in the detection path (100  $\mu\text{m}$  pinhole) was calculated using geometric optics,<sup>25</sup> accounting for the water-glass dielectric surface (Supporting Information section 4S). The peaks of the  $\text{Exc}(\vec{x})$  and  $\text{CEF}(\vec{x})$  were translated with respect to each other to ensure good overlap in the  $z$  direction (to simulate adjustment of the pinhole or the APDs). The detectivity  $\phi(\vec{x}) \equiv \text{Exc}(\vec{x})\text{CEF}(\vec{x})$  is used for calculating the kernels described in section 2.3. The calculated effective detection volume is 3.2  $\mu\text{m}^3$  (a concentration of 1 nM corresponds to an occupancy of 1.9). The diffusion time of a molecule of species  $\alpha$  with a diffusion constant  $D_\alpha$  was calculated by simulating many paths through the detection volume:  $\tau^D = (2.8 \times 10^{-10} \text{ cm}^2)/D_\alpha$ .

Since the persistence length of dsDNA is  $\sim 150$  bp or 500 Å,<sup>41</sup> we treat the DNA as rodlike polymers with length  $L \approx 240$  Å and diameter  $b \approx 20$  Å, with averaged translational diffusion constant<sup>42</sup>

$$D = \ln(L/b)k_B T / 3\pi\eta L \quad (14)$$

where  $\eta$  is the dynamic viscosity of the solution ( $\eta = 1.16$  mPa·s for aqueous solution with 5% v/v glycerol<sup>43</sup>). For  $T = 25$  °C,  $D = 3.9 \times 10^{-7}$  cm<sup>2</sup>/s, so  $\tau^D = 710$   $\mu\text{s}$  in the calculated detection volume. We model the RNAP holoenzyme (dimensions 120  $\times$  140  $\times$  150 Å<sup>3</sup>)<sup>37</sup> as a sphere with radius in the range 60–75 Å, where

$$D = k_B T / 6\pi\eta R \quad (15)$$

In this case, the expected diffusion time is 0.9–1.1 ms.

**3.6. Fitting Routine.** We use nonlinear least-squares fitting to extract the parameters from the PAID histograms, calculating the necessary partial derivatives numerically. We use 10 independent instances of the histogram to calculate estimates of the statistical errors of each bin to be used as weights in the fitting routine.<sup>38</sup> If a bin is nonzero in fewer than 10 instances, it is excluded from the fit. Without this restriction, we found that only a few points dominate the value of  $\chi^2$ . The 10 independent instances were individually fitted, extracting the means and standard errors of the model parameters. The standard error of the mean with 10 instances is  $\sqrt{N} \sim 3$  ( $N = 10$  is the sample size) times smaller than the sample standard deviation for each parameter. The same procedure is used for FIMDA, FCS, and FIDA; each model used was specialized to diffusion within a Gaussian detection volume (excluding triplet-state fluctuations, etc.). FIMDA histograms were formed using the same bins as the PAID histograms (Supporting Information section 1S), and the spacing between counted intervals was set to 10  $\mu$ s for all time bin widths.

For experimental data, 300 s of data were split into 10 sections of 30 s each; PAID histograms were calculated for each section, and the standard deviation of each bin was calculated to estimate the error of each bin; each section was fit separately, and the mean and error of the mean of each fitted parameter was calculated. Because PAID does not yet account for triplet-state related fluctuations, the range of fitting for the experiments is restricted to time intervals greater than 10  $\mu$ s. The procedure was modified for the low-occupancy dual-channel experiments since there were too few bursts from DNA<sup>Cy5,1T/Cy3,65B</sup> and DNA<sup>Cy5,1T</sup>, resulting in large errors for the extracted diffusion times. To obtain better statistics, the histogram for all 300 s was fitted, and subsequent bootstrap sampling was used to obtain error estimates for the extracted parameters.<sup>44</sup> To obtain one bootstrap sample, the data were split into 10 sections of 30 s each; 10 of these sections were randomly selected with replacement (the same section can be selected multiple times, or omitted). The PAID histograms for the selected sections were then averaged. We obtained 10 bootstrap samples in this manner and fit each of the resulting histograms; the standard deviation of the 10 values extracted for each parameter is quoted as error bars for low-occupancy data.

## 4. Results

**4.1. Graphical Representation of Single-Channel Experiments Using PAID.** The PAID histogram is ideal for simultaneous measurements of diffusion time, brightness, and occupancy. We now describe how the parameters of a diffusing species (occupancy,  $c_1$ ; brightness per molecule in channel ( $A = 1$ ),  $q_{11}$ ; diffusion time,  $\tau_1^D$ ; and background rate,  $k_{01}$ ) affect the single-channel PAID histogram. Figure 4A shows the PAID histogram for a 30 s simulation with  $c_1 = 0.1$ ,  $\tau_1^D = 100 \mu$ s,  $q_{11} = 50$  kHz, and  $k_{01} = 0$  kHz. At time interval  $\tau = 0$ , a photon from one of the molecules is detected; the position probability distribution for the molecule that emitted the photon matches the detectivity  $\phi(\vec{x})$ . The autocorrelation amplitude is  $C_{SS}(\tau = 0) = (c_1 + 1)/c_1$ ,<sup>1</sup> meaning there are on average  $c_1 + 1 = 1.1$  molecules of species  $\alpha = 1$  inside  $V_{\text{eff}}$  at  $\tau = 0$ , and producing a ridge of high correlation density at small  $\tau$ , (Figure 4A, *white and red contours*). The molecule exits the detection volume with the diffusion time scale  $\tau_1^D$  (the position probability distribu-

tion flattens); hence, the ridge of high correlation density decays with a time scale  $\tau \sim \tau_1^D$  (as with autocorrelation in FCS).

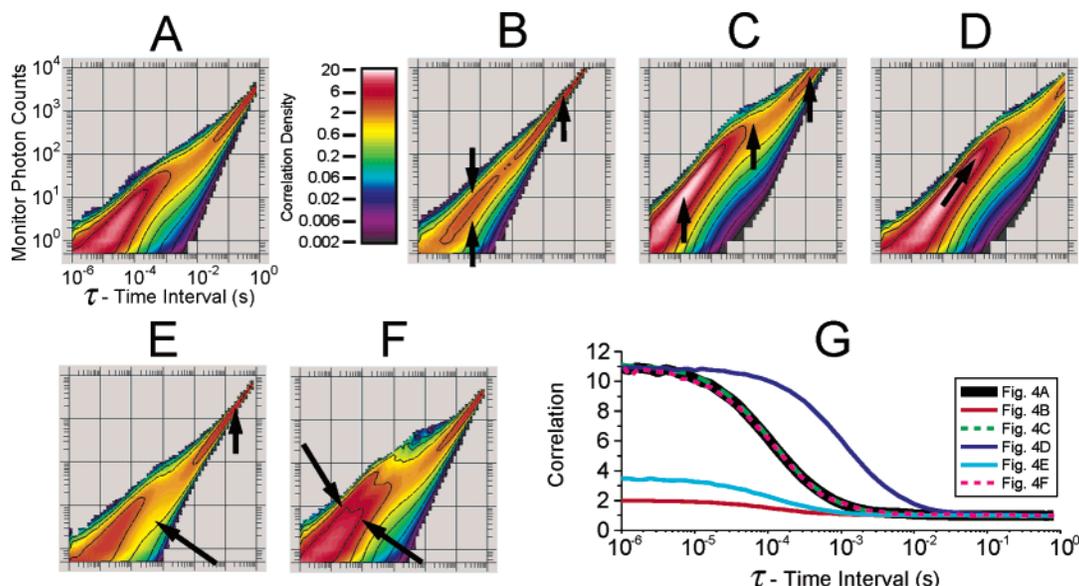
While the molecule that emitted the photon is present (time interval  $\tau \ll \tau_1^D$ ), the average count rate in channel  $A$  is  $(1 + c_1)q_{11}$  so that the peak of high correlation density follows a trajectory in the histogram of the form  $\mathbf{n} \sim [(1 + c_1)q_{11}]\tau$  (in the log plot in Figure 4A, the slope of the red ridge is 1 whereas the vertical offset is  $\log[(1 + c_1)q_{11}]$ ). After a time interval  $\tau \gg \tau_1^D$ , the molecule has diffused out of  $V_{\text{eff}}$  and the average count rate returns to the value for a random time,  $c_1q_{11}$  so that the correlation density peak follows a trajectory of the form  $\mathbf{n} \sim (c_1q_{11})\tau$  (in the log plot in Figure 4A, the slope of the second, orange ridge in the *upper-right corner* is 1 whereas the vertical offset is  $\log(c_1q_{11})$ ).

Individual parameters influence the PAID histograms in specific ways (Figure 4B–E). When occupancy increases 10-fold (Figure 4B), the histogram narrows along the monitor photon count axis (*opposing arrows*), the correlation amplitude decreases 10-fold, and the long time count rate increases 10-fold (*arrow* on the upper-right). When the brightness increases 10-fold (Figure 4C), the histogram shifts up in the log–log plot (as indicated by the *arrows*) and narrows along the monitor photon axis. When the diffusion time increases 10-fold (Figure 4D), the red and white contours are extended 10-fold along the time interval and the monitor photon count axes (*arrow*). When a constant background with rate  $k_{01} = 5$  kHz is added (Figure 4E), the correlation density decreases (since many start photons are now uncorrelated background photons), the background gives rise to an additional shallow slope (*arrow*) of the histogram to the right of the main correlation peak, and the total count rate is doubled (seen at long time intervals  $\tau$ ). In Figure 4F, a second brighter species with  $q_{21} = 4q_{11} \times$ ,  $c_2 = 0.006$ ,  $\tau_2^D = 100 \mu$ s, and background  $k_{01} = 0.765$  kHz is added (parameters chosen so that the correlation curves corresponding to Figure 4A and Figure 4F overlap, demonstrating the limitations of FCS).

The advantages of PAID over FCS are demonstrated by comparing Figure 4G to the PAID histograms in Figure 4A–F. In FCS, the occupancy is extracted from the correlation amplitude, which is  $C_{SS}(\tau = 0) = 1 + 1/c_1$  for a single species with no background ( $C_{SS}(\tau = 0) = 11.0$  for the black curve). One FCS limitation is that both *occupancy increases* (Figure 4B) and *background increases* (Figure 4E) decrease the correlation amplitude (cf. the *red curve* ( $c'_1 = 10 \times c_1$ ) and the *cyan curve* (with  $k_{01} = 5$  kHz) in Figure 4G). Another FCS limitation is that changes in molecular brightness (Figure 4C) do not affect the correlation curve (cf. the *black curve* and *dashed green curve* ( $q'_{11} = 10 \times q_{11}$ ) in Figure 4G), preventing the ability to distinguish between components with different brightness (cf. the overlap of the *magenta line* and the *black line* in Figure 4G to the difference between Figure 4A and Figure 4F).

The 2D representation shown in Figure 4, where changes in brightness and diffusion time can be identified by changes in specific features, is made possible by the definition used for the PAID histogram. These features are not emphasized by histograms based on counting photons on evenly spaced time intervals (cf. Figure 4A and Supporting Information Figure 4SA).

**4.2. Simulations: Quantitative Analysis of Stoichiometry Using Single-Channel PAID.** To demonstrate the ability of PAID to extract simultaneously occupancy, diffusion time, and brightness, and to compare it to FCS, FIDA/PCH, and FIMDA, we generated single-channel ( $S = T = M$ ), single-component simulations in a Gaussian detection volume (section 3.4). The



**Figure 4.** Effects of changes in the characteristics of diffusing molecules on the single-channel PAID histogram. The first histogram (A) was formed from a single-channel simulation, with a single diffusing species (start = stop = monitor channel). The x-axis is the time interval between the start and stop photons; the y-axis is the number of monitor photon counts between them. The occupancy is  $c_1 = 0.1$ , the diffusion time is  $\tau_1^D = 100 \mu\text{s}$ , the brightness is  $q_{11} = 50 \text{ kHz}$ , and the background count rate is  $q_{01} = 0 \text{ kHz}$ . Panels B–F illustrate how changes in the parameters of the diffusing species, the background count rate, and the sample composition affect the PAID histogram. Arrows indicate changes in features (see text for details). (B) The occupancy is increased 10-fold,  $c'_1 = 10c_1$ . (C) The brightness per molecule is increased 10-fold,  $q'_{11} = 10q_{11}$ . (D) The diffusion time of the molecule is increased 10-fold,  $\tau_1^{D'} = 10\tau_1^D$ . (E) A background component is added,  $k'_{01} = 5 \text{ kHz}$ . (F) A species with 4 times the brightness is introduced,  $q_{21} = 200 \text{ kHz}$ , along with a background  $k'_{01} = 0.8 \text{ kHz}$  (values chosen so that autocorrelations corresponding to parts A and F overlap). (G) Autocorrelations, corresponding to the collapse of the PAID histograms in parts A–F onto the time interval axis, are shown for comparison with FCS.

**TABLE 1: Parameters Extracted Using PAID, FIMDA, FCS, and FIDA Fits for Single-Channel, One-Component Simulations in a Gaussian Detection Volume<sup>a</sup>**

parameters	simulation:	averaged fits for 10 simulations (30 s each): low occupancy			
		PAID	FIMDA	FCS	FIDA
$\chi^2$		$1.2 \pm 0.1$	$4.7 \pm 0.2$	$1.3 \pm 0.1$	$0.6 \pm 0.1$
$k_{01}$ (kHz)	0.0	$0.01 \pm 0.01$	$0.01 \pm 0.01$	N/A <sup>b</sup>	$0.02 \pm 0.01$
$c_1$ (mol)	0.1	$0.100 \pm 0.001$	$0.100 \pm 0.001$	$0.099 \pm 0.001$	$0.105 \pm 0.001$
$\tau_1^D$ ( $\mu\text{s}$ )	100.0	$98 \pm 1$	$99 \pm 1$	$100 \pm 1$	N/A <sup>b</sup>
$q_{11}$ (kHz)	50.0	$48.9 \pm 0.3$	$49.7 \pm 0.3$	N/A <sup>b</sup>	$47.3 \pm 0.3$

<sup>a</sup> Fixed values are in italics, fitted parameters are quoted with error estimates. <sup>b</sup> Not available.

low-occupancy series consists of 10 simulations:  $c_1 = 0.1$ ,  $\tau_1^D = 100 \mu\text{s}$ ,  $q_{11} = 50 \text{ kHz}$ , and  $k_{01} = 0 \text{ kHz}$  (Table 1). In the intermediate-occupancy series,  $c_1 = 1.0$  (Table 1S). PAID performed well in both series, with an average error for each parameter of  $<2\%$ . For low-occupancy simulations, FIMDA extracts values with accuracy similar to PAID, but with a worse fit ( $\chi^2 \sim 4$ ; see Supporting Information section 5S). For the intermediate-occupancy simulations, the fits are good ( $\chi^2 \sim 1$ ) and the extracted parameters are close to the simulation values (except for a 5% downward bias in the diffusion time). The accuracy of FCS-extracted parameters was similar to PAID and FIMDA. The errors of the FIDA-extracted parameters were similar to the errors found using the other methods. There are biases in  $c_1$  and  $q_{11}$  due to diffusion during the time bin (time bin width was  $20 \mu\text{s}$ ,  $1/5$  the shortest diffusion time). These biases can be estimated using a correction factor  $\Gamma_{\text{diff}}$  for occupancy  $c_{\text{app}} = c/\Gamma_{\text{diff}}$  and brightness  $q_{\text{app}} = q\Gamma_{\text{diff}}$ .<sup>14</sup> For the Gaussian detectivity used, we obtain  $c_{\text{app}} = 0.107$  and  $q_{\text{app}} = 47 \text{ kHz}$ , matching the values in Tables 1 and 1S.

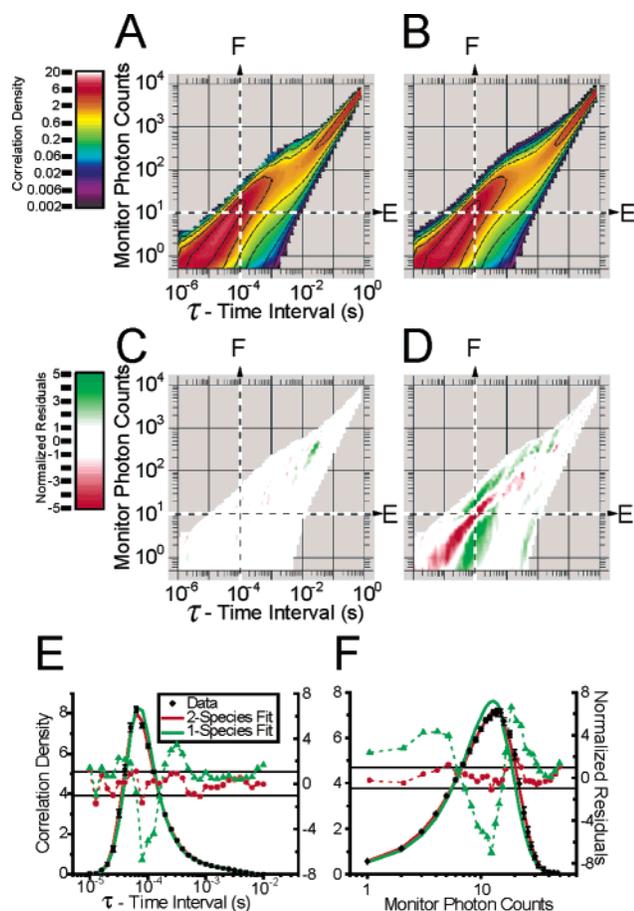
We also performed low- ( $c_1 = c_2 = 0.05$ ; Table 2) and intermediate-occupancy ( $c_1 = c_2 = 0.5$ ; Table 2S) simulations to test the ability of PAID to differentiate two components. Figure 5 shows a representative PAID fit for a low-occupancy simulation from Table 2; a one-species fit is shown to be a poor

fit of the data. PAID performed well: low-occupancy simulations yielded low errors (1% to 2%); intermediate-occupancy simulations yielded somewhat higher errors (1% to 8%), as well as biases up to 10% for the diffusion time and occupancy of the lower brightness species. The errors and biases of the FIMDA-extracted parameters were in general similar to those extracted using PAID. However, for low-occupancy simulations, the occupancy biases were larger ( $\sim 20\%$ ). For intermediate-occupancy simulations, occupancy biases were reduced, but the background count rate values were off by 25%. Differences between PAID and FIMDA at low occupancy are primarily related to modeling, rather than to the histograms used (Supporting Information section 5S). FCS fits with fixed brightness did not converge for either high- or low-occupancy simulations. If both occupancy and brightness are fixed, diffusion times can be extracted, but with poorer accuracy than with PAID or FIMDA; since FIDA/PCH can (at best) fix the occupancies and brightnesses to their correct values, this demonstrates that simultaneous fitting of FCS and FIDA/PCH cannot match the performance of PAID or FIMDA. The errors of the parameters extracted from the two-species simulations were larger for FIDA than for PAID or FIMDA.

**4.3. Experiments: Quantitative Analysis Using Single-Channel PAID.** Measurements were performed on DNA

**TABLE 2: Parameters Extracted Using PAID, FIMDA, FCS, and FIDA Fits for Single-Channel, Two-Component Simulations in a Gaussian Detection Volume**

parameters	simulation:	averaged fits for 10 simulations (30 s each): low occupancy			
		PAID	FIMDA	FCS	FIDA
$\chi^2$		$0.77 \pm 0.04$	$1.1 \pm 0.3$	$1.3 \pm 0.2$	$0.5 \pm 0.1$
$k_{01}$ (kHz)	2.0	$2.00 \pm 0.01$	$1.90 \pm 0.01$	2.0	$1.99 \pm 0.03$
$c_1$ (mol)	0.05	$0.048 \pm 0.001$	$0.064 \pm 0.001$	0.05	$0.059 \pm 0.002$
$\tau_1^D$ ( $\mu$ s)	100.0	$100 \pm 2$	$117 \pm 4$	$105 \pm 7$	N/A <sup>a</sup>
$q_{11}$ (kHz)	50.0	$50 \pm 1$	$51 \pm 2$	50.0	$48 \pm 2$
$c_2$ (mol)	0.05	$0.052 \pm 0.001$	$0.042 \pm 0.002$	0.05	$0.048 \pm 0.003$
$\tau_2^D$ ( $\mu$ s)	150.0	$145 \pm 2$	$146 \pm 3$	$142 \pm 3$	N/A <sup>a</sup>
$q_{21}$ (kHz)	100.0	$98 \pm 1$	$106 \pm 1$	100.0	$100 \pm 2$

<sup>a</sup> Not Available.

**Figure 5.** Representative PAID fit from Table 2. (A) The PAID histogram for the simulation (low occupancy). (B) The fitted PAID histogram for the 2-species fit. (C) A 2D map of residuals for the fitted histogram in part B. (D) A 2D map of residuals for the 1-species fit. (E) Horizontal slices of PAID histograms. Slices of the simulation are black; slices of fits are red and green; residuals are dashed. (F) Vertical slices.

fragments to test the ability of single-channel PAID to differentiate species in solution. Three samples were tested in the low-occupancy regime: 0.1 nM DNA<sup>Cy3B,1T</sup> only; 0.1 nM DNA<sup>Cy3B,1T/Cy3B,65B</sup> only; and a mixture of 0.05 nM DNA<sup>Cy3B,1T</sup> and 0.05 nM DNA<sup>Cy3B,1T/Cy3B,65B</sup> (Tables 3 and 4). Similarly, three samples were tested with 10 times higher concentration (in the intermediate-occupancy regime; Tables 3S and 4S). For the fits of the one-species samples (Tables 3 and 3S), all parameters were fitted; the model fits well ( $\chi^2$  in the range of 0.8–1.5). The important feature of the analysis is that the DNA<sup>Cy3B,1T/Cy3B,65B</sup> is  $\sim 2.1$  times as bright as the DNA<sup>Cy3B,1T</sup>, demonstrating the ability of PAID to evaluate stoichiometry (See Supporting Information section 6.2S on initial experiments

performed with Cy3 rather than Cy3B); similar stoichiometry results are obtained using FCS in combination with the mean count rate (not shown). On the basis of the calculated detection volume (section 3.5), the occupancies of 1 nM and 0.1 nM samples are expected to be, respectively, 1.9 and 0.19. The occupancies extracted are within 30% of these values (the errors are attributed to pipetting errors and losses on surfaces). The measured diffusion times (620–760  $\mu$ s) match the calculated diffusion time (710  $\mu$ s). The measured background rates for three of the samples (0.2–0.4 kHz) are consistent with buffer-only measurements (0.24 kHz). The exception is the intermediate-occupancy sample of DNA<sup>Cy3B,1T/Cy3B,65B</sup>, however, the background makes up  $< 1\%$  of the signal.

Tables 4 and 4S list the results for two-component fits of the one- and two-species samples. For the “fixed ratio fit”, the brightness of one species is fixed to be twice the brightness of the other; the total brightness can vary and the diffusion times of the two components are kept equal. For the “restricted fit”, only occupancies are allowed to vary; the brightness values for the two components are set to be 1 and 2 times the brightness from the single-labeled species in Tables 3 and 3S, the diffusion times are taken from the same species, and the background is taken from separate, buffer-only measurements.

For the single-species samples, only one component is fitted with a significant occupancy for most of the fits (ratio of occupancies  $> 10:1$ ); the only exception is the “fixed ratio fit” of the intermediate-occupancy sample of DNA<sup>Cy3B,1T/Cy3B,65B</sup> (the ratio is 4:1; however, the restricted fit has a ratio of 12:1). For the mixture samples, significant occupancies are fitted for both components for all of the fits. By dividing the single-species occupancies (Tables 3 and 3S) by 2, we obtain the expected occupancies for the mixture; the expected occupancies of DNA<sup>Cy3B,1T</sup> and DNA<sup>Cy3B,1T/Cy3B,65B</sup> are, respectively, 0.10 and 0.07 for the low-occupancy samples and 1.1 and 1.2 for the intermediate-occupancy samples. The occupancies extracted from the low-occupancy mixture match the expected occupancies within 10% (except the lower brightness species is fit with a 30% higher occupancy with the restricted fit). The occupancies extracted from the intermediate-occupancy mixture are biased toward the lower brightness species (identified with DNA<sup>Cy3B,1T</sup>; 30% and 80% higher in the fixed ratio fit and restricted fit, respectively), decreasing the amount detected in the higher brightness species (identified with DNA<sup>Cy3B,1T/Cy3B,65B</sup>; 20% lower and 50% lower in the fixed ratio fit and the restricted fit, respectively). These biases are likely due to inadequate modeling of the detection volume; spurious components or biases are generally the result of imperfect overlap of model and data.<sup>45</sup> Nevertheless, significant occupancies for two components were extracted in the mixtures, but not in the single-species samples, demonstrating the capabilities of PAID to detect heterogeneity in single-channel experiments. For the fixed ratio fits, the

**TABLE 3: Parameters Extracted Using Single-Component PAID Fits for Single-Channel, Single-Species Experiments<sup>a</sup>**

parameters	averaged fits for 10 measurements (30 s each): low occupancy	
	DNA <sup>Cy3B,1T</sup>	DNA <sup>Cy3B,1T/Cy3B,65B</sup>
$\chi^2$	0.8 ± 0.1	1.2 ± 0.1
$k_{01}$ (kHz)	0.39 ± 0.03	0.37 ± 0.04
$c_1$ (mol)	0.19 ± 0.01	0.14 ± 0.01
$\tau_1^D$ ( $\mu$ s)	640 ± 20	760 ± 30
$q_{11}$ (kHz)	9.9 ± 0.2	21.5 ± 0.7

<sup>a</sup> For the buffer-only measurements,  $k_{01} = 0.24 \pm 0.01$  kHz.

**TABLE 4: Parameters Extracted Using Two-Component PAID Fits for Single-Channel, One- and Two-Species Experiments**

parameters	averaged fits for 10 measurements (30 s each): low occupancy					
	DNA <sup>Cy3B,1T</sup>	DNA <sup>Cy3B,1</sup>	DNA <sup>Cy3B,1T/Cy3B,65B</sup>	DNA <sup>Cy3B,1T/Cy3B,65B</sup>	DNA <sup>Cy3B,1T</sup>	DNA <sup>Cy3B,1T</sup>
	fixed ratio fit <sup>a</sup>	restricted fit <sup>b</sup>	fixed ratio fit	restricted fit	fixed ratio fit	restricted fit
$\chi^2$	0.8 ± 0.1	1.9 ± 0.3	1.2 ± 0.1	2.3 ± 0.3	1.1 ± 0.1	1.9 ± 0.3
$k_{01}$ (kHz)	0.36 ± 0.02	0.24	0.35 ± 0.03	0.24	0.40 ± 0.02	0.24
$c_1$ (mol)	0.19 ± 0.01	0.21 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.10 ± 0.02	0.13 ± 0.01
$\tau_1^D$ ( $\mu$ s)	680 ± 10	640	760 ± 20	640	760 ± 20	640
$q_{11}$ (kHz)	9.7 ± 0.2	9.9	10.8 ± 0.3	9.9	9.6 ± 0.4	9.9
$c_2$ (mol)	0.01 ± 0.01	0.001 ± 0.001	0.14 ± 0.01	0.16 ± 0.01	0.08 ± 0.01	0.07 ± 0.01
$\tau_2^D$ ( $\mu$ s)	680 ± 10	640	760 ± 20	640	760 ± 20	640
$q_{21}$ (kHz)	4.8 ± 0.1	19.9	21.6 ± 0.5	19.9	19.3 ± 0.8	19.9

<sup>a</sup> The ratio between the brightness of both components, fixed to a factor of 2, and the diffusion times of the two components are linked. <sup>b</sup> The brightness values are 1 and 2 times the values extracted from the DNA<sup>Cy3B,1T</sup> samples, and the diffusion times are fixed. The background rates were extracted from separate experiments.

diffusion times extracted (630–760  $\mu$ s) match the calculated value (710  $\mu$ s). The brightness values are consistent with those extracted from the single-component fits (all within 15%).

**4.4. Graphical Representation of Dual-Channel Experiments Using PAID.** In dual-channel experiments, the ability of PAID to extract brightness for each species in multiple channels is extremely important. As illustrated in Figure 1B, in a dual-color fluorescence binding assay, one molecule A<sup>y</sup> is labeled with a fluorophore of one color (for example “yellow”, denoted *y*), while the second molecule B<sup>r</sup> is labeled with a fluorophore of a second color (for example “red”, denoted *r*). Three species are present in solution: free A<sup>y</sup>, free B<sup>r</sup>, and complexes A<sup>y</sup>B<sup>r</sup>. The fluorophores *y* and *r* have corresponding “yellow” and “red” detection channels: *Y* and *R*, respectively. Because of the vibronic tail of organic fluorophores toward the red end of their emission spectra, there is a small contribution from *y* into the channel *R*; the contribution from *r* into channel *Y* is typically negligible.

For each dual-channel data set (from experiment or simulation), we performed a global fit of the series of all possible dual-channel PAID histograms, extracting brightness (in both channels), diffusion time, and occupancy for the species A<sup>y</sup>B<sup>r</sup>, A<sup>y</sup>, and B<sup>r</sup>. We write the channel assignments as three letter codes, *STM*, specifying the start, stop, and monitor channels (eg., the *RYR* PAID histogram uses *S* = *R*, *T* = *Y*, and *M* = *R*). For dual-channel experiments, there are 8 unique channel assignments for the PAID histogram: *RRR*, *RRY*, *RYR*, *RYY*, *YRR*, *YRY*, *YYR*, and *YYY*. The assignments of *S* and *T* (first two letters) select species of interest, which emit photons in both *S* and *T*. The choice of *M* determines which fluorophore will be analyzed in terms of brightness. Figure 6 shows how each PAID histogram emphasizes particular species.

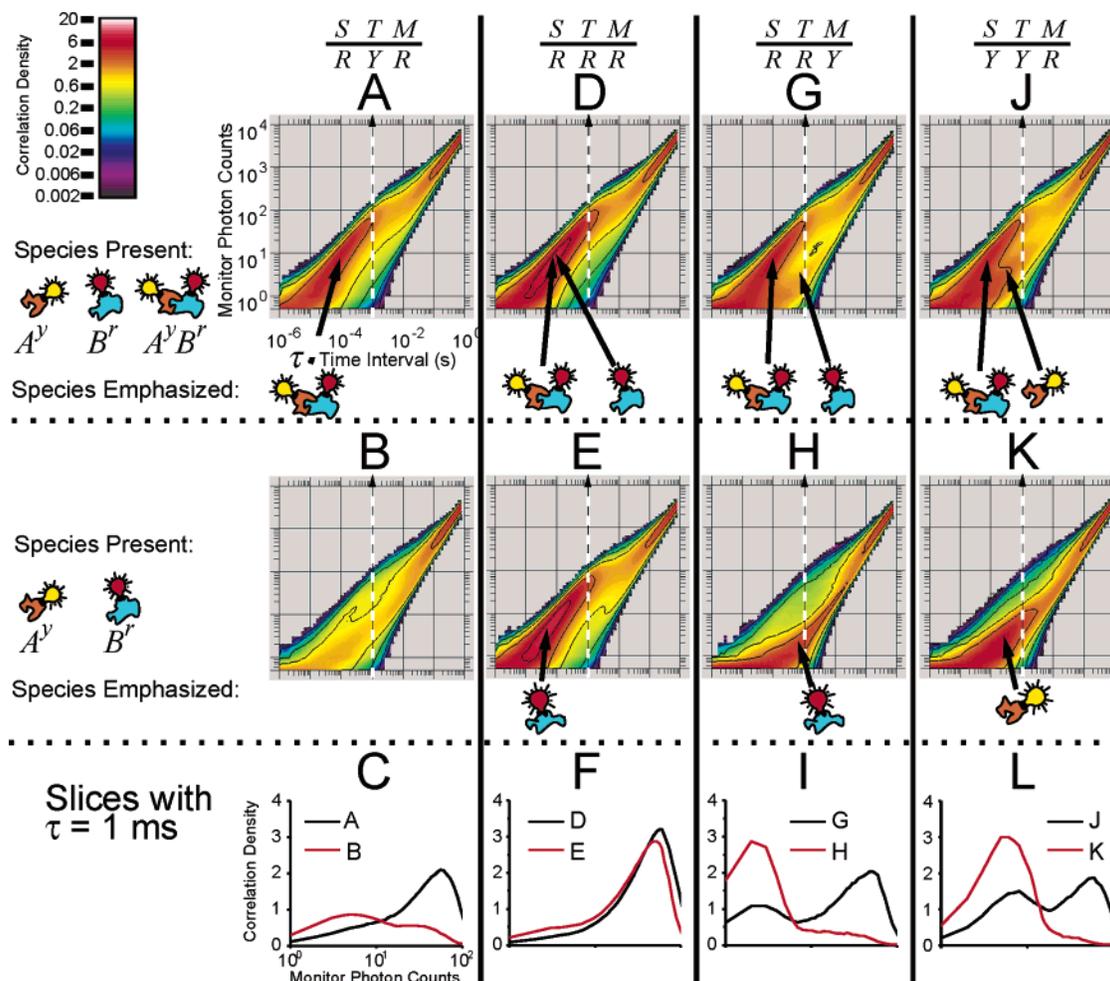
Taken together, the *RYR*, *RYY*, *YRR*, and *YRY* histograms can determine the properties of A<sup>y</sup>B<sup>r</sup>. The histograms with *S* = *R* and *T* = *Y* emphasize time regions when A<sup>y</sup>B<sup>r</sup> is present, since A<sup>y</sup>B<sup>r</sup> emits photons in both *R* and *Y* whereas A<sup>y</sup> and B<sup>r</sup> emit only in one channel. Specifically, the *RYR* histogram monitors the brightness of *r* (Figure 6A, *RYR* with A<sup>y</sup>B<sup>r</sup> present; Figure

6B, *RYR* with no A<sup>y</sup>B<sup>r</sup>). The correlation density peak in Figure 6A corresponding to A<sup>y</sup>B<sup>r</sup> (shown by the *arrow*) is absent in Figure 6B (the small correlation peak in Figure 6B corresponds to the contribution of *y* into *R*). The *RYY*, *YRR*, and *YRY* histograms (not shown) are similar to the *RYR* histogram.

The histograms with *S* = *R* and *T* = *R* emphasize time regions where B<sup>r</sup> and A<sup>y</sup>B<sup>r</sup> are present, since both emit photons in *R*. Specifically, for the *RRR* histogram, the correlation peaks from B<sup>r</sup> and A<sup>y</sup>B<sup>r</sup> overlap since both have a similar brightness in *R* (Figure 6D–F). There is only one correlation peak visible, where B<sup>r</sup> and A<sup>y</sup>B<sup>r</sup> both contribute (Figure 6D, *arrows*), so the *RRR* histogram cannot distinguish well between B<sup>r</sup> and A<sup>y</sup>B<sup>r</sup>. However, for the *RRY* histogram (Figure 6G–I), the correlation peaks from B<sup>r</sup> and A<sup>y</sup>B<sup>r</sup> are well separated. A<sup>y</sup>B<sup>r</sup> emits in *Y* (resulting in a correlation peak with high monitor photon count; *left arrow*), whereas B<sup>r</sup> does not emit in *Y* (resulting in a correlation peak with low monitor photon counts; *right arrow*) (Figure 6, parts G and I). When A<sup>y</sup>B<sup>r</sup> is absent, the corresponding correlation peak is noticeably absent, leaving only the correlation peak resulting from B<sup>r</sup> (Figure 6, parts H and I). The *YYY* and *YYR* PAID histograms emphasize A<sup>y</sup> and A<sup>y</sup>B<sup>r</sup> in a similar manner (Figure 6J–L).

In Supporting Information section 7.1S, all 8 dual-channel PAID histograms were simultaneously fitted for the low-occupancy simulations shown in Figure 6 and for intermediate-occupancy simulations as well. The accuracy of the extracted parameters was within 7% for diffusion times and within 2% for the remaining parameters.

**4.5. Experiments: Quantitative Analysis Using Dual-Channel PAID.** We performed measurements on fluorescently labeled DNA fragments to test the ability of dual-channel PAID to detect multiple species in solution and measure their occupancies, diffusion times, and brightnesses in both channels. The fluorophores used in these experiments were Cy3 (as the “yellow” fluorophore *y*) and Cy5 (as the “red” fluorophore *r*). On the basis of the occupancies measured at higher concentration and the dilutions used, the occupancies for the low-occupancy samples were expected to be  $0.09 \pm 0.01$  for



**Figure 6.** Dual-channel PAID histograms for simulations containing the species expected in an interaction study. Symbols  $r$  and  $y$  indicate the red and yellow fluorophores, respectively;  $R$  and  $Y$  indicate the red and yellow detector channels, respectively. Parts A, D, G, and J are PAID histograms  $R Y R$ ,  $R R R$ ,  $R R Y$ , and  $Y Y R$  of a mixture containing  $A^y$  ( $c_1 = 0.05$ ,  $\tau_1^D = 300 \mu\text{s}$ ,  $q_{1R} = 5 \text{ kHz}$ , and  $q_{1Y} = 45 \text{ kHz}$ ),  $B^r$  ( $c_2 = 0.05$ ,  $\tau_2^D = 300 \mu\text{s}$ ,  $q_{2R} = 50 \text{ kHz}$ , and  $q_{2Y} = 0 \text{ kHz}$ ), and  $A^y B^r$  ( $c_3 = 0.05$ ,  $\tau_3^D = 400 \mu\text{s}$ ,  $q_{3R} = 55 \text{ kHz}$ , and  $q_{3Y} = 45 \text{ kHz}$ ). The background in each channel was  $k_{0R} = k_{0Y} = 2 \text{ kHz}$ . In parts B, E, H, and K, the mixture does not contain the complex ( $A^y B^r$ ) present in parts A, D, G, and J. The effect of the absence of  $A^y B^r$  is seen by comparing part A with part B, part D with part E, part G with part H, and part J with part K. Differences are pointed to by arrows and discussed in the text. The cartoons of the free molecules and complexes indicate which species contribute to any large correlation density peak. Vertical slices from the PAID histograms at time interval  $\tau = 1 \text{ ms}$  are shown in parts C, F, I, and L, showing the difference between the histograms with and without  $A^y B^r$ .

**TABLE 5: Parameters Extracted Using PAID Fits for Two-Channel, Single-Species, Low-Occupancy Experiments**

parameters	fit for 1 measurement (300 s): low occupancy <sup>a</sup>		
	DNA <sup>Cy5,1T</sup>	DNA <sup>Cy3,65B</sup>	DNA <sup>Cy5,1T/Cy3,65B</sup>
$\chi^2$	$9.3 \pm 0.2$	$5.5 \pm 0.1$	$1.9 \pm 0.3$
$k_{0R}$ (kHz)	$0.82 \pm 0.01$	$0.76 \pm 0.01$	$0.68 \pm 0.01$
$k_{0Y}$ (kHz)	$1.30 \pm 0.01$	$1.14 \pm 0.03$	$0.95 \pm 0.01$
DNA fragment 1	DNA <sup>Cy5,1T</sup>	DNA <sup>Cy3,65B</sup>	DNA <sup>Cy5dark,1T/Cy3,65B</sup>
$c_1$ (mol)	$0.022 \pm 0.001$	$0.073 \pm 0.003$	$0.023 \pm 0.003$
$\tau_1^D$ ( $\mu\text{s}$ )	$430 \pm 10$	$570 \pm 10$	$670 \pm 60$
$q_{1R}$ (kHz)	$12.0 \pm 0.1$	$0.94 \pm 0.03$	$0.8 \pm 0.1$
$q_{1Y}$ (kHz)	$0.02 \pm 0.01$	$10.0 \pm 0.3$	$9.2 \pm 0.8$
DNA fragment 2	none	none	DNA <sup>Cy5,1T/Cy3,65B</sup>
$c_2$ (mol)			$0.012 \pm 0.001$
$\tau_2^D$ ( $\mu\text{s}$ )			$650 \pm 10$
$q_{2R}$ (kHz)			$6.8 \pm 0.3$
$q_{2Y}$ (kHz)			$6.7 \pm 0.3$

<sup>a</sup> All 300 s of low-occupancy data fit at once to improve statistics.

DNA<sup>Cy5,1T</sup>,  $0.12 \pm 0.01$  for DNA<sup>Cy3,65B</sup>, and  $0.08 \pm 0.01$  for DNA<sup>Cy5,1T/Cy3,65B</sup>. The occupancy for DNA<sup>Cy5,1T/Cy3,65B</sup> was found by analyzing the Cy3 fluorescence, so the occupancy quoted also includes DNA<sup>Cy5dark,1T/Cy3,65B</sup> (with nonfluorescent

Cy5, as seen in previous studies<sup>22,46</sup>). The fragments were prepared and analyzed as free components (Table 5) and in the following mixtures (Table 6): DNA<sup>Cy5,1T</sup>/DNA<sup>Cy3,65B</sup> (to simulate noninteracting species) and DNA<sup>Cy5,1T/Cy3,65B</sup>/DNA<sup>Cy5,1T</sup>/

**TABLE 6: Parameters Extracted Using PAID Fits for Two-Channel, Multiple-Species, Low-Occupancy Experiments**

parameters	fit for 1 measurement (300 s): low occupancy			
	DNA <sup>Cy3,65B</sup> , DNA <sup>Cy5,1T</sup> unrestricted 2 compo- nent fit	DNA <sup>Cy3,65B</sup> , DNA <sup>Cy5,1T</sup> restricted 3 compo- nent fit	DNA <sup>Cy3,65B</sup> , DNA <sup>Cy5,1T</sup> , DNA <sup>Cy5,1T/Cy3,65B</sup> unrestricted 3 component fit	DNA <sup>Cy3,65B</sup> , DNA <sup>Cy5,1T</sup> , DNA <sup>Cy5,1T/Cy3,65B</sup> restricted 3 component fit
$\chi^2$	2.8 ± 0.2	3.5 ± 0.1	2.3 ± 0.4	3.9 ± 0.1
$k_{0R}$ (kHz)	0.87 ± 0.01	0.88 ± 0.01	0.80 ± 0.01	0.78 ± 0.01
$k_{0Y}$ (kHz)	1.04 ± 0.01	1.04 ± 0.01	1.12 ± 0.01	1.08 ± 0.01
DNA fragment 1	DNA <sup>Cy5,1T</sup>	DNA <sup>Cy5,1T</sup>	DNA <sup>Cy5,1T</sup>	DNA <sup>Cy5,1T</sup>
$c_1$ (mol)	0.018 ± 0.001	0.016 ± 0.001	0.011 ± 0.001	0.016 ± 0.001
$\tau_1^D$ ( $\mu$ s)	430 ± 10	430	360 ± 10	420
$q_{1R}$ (kHz)	11.0 ± 0.2	12.0	15.4 ± 0.4	12.5
$q_{1Y}$ (kHz)	0.01 ± 0.02	0.02	0.02 ± 0.03	0.0
DNA fragment 2	DNA <sup>Cy3,65B</sup>	DNA <sup>Cy3,65B</sup>	DNA <sup>Cy3,65B</sup> <sup>a</sup>	DNA <sup>Cy3,65B</sup> <sup>a</sup>
$c_2$ (mol)	0.068 ± 0.001	0.067 ± 0.001	0.077 ± 0.003	0.088 ± 0.001
$\tau_2^D$ ( $\mu$ s)	530 ± 10	570	570 ± 10	570
$q_{2R}$ (kHz)	0.95 ± 0.02	0.94	0.89 ± 0.02	0.94
$q_{2Y}$ (kHz)	10.2 ± 0.1	10.0	11.2 ± 0.3	10.0
DNA fragment 3	none	DNA <sup>Cy5,1T/Cy3,65B</sup>	DNA <sup>Cy5,1T/Cy3,65B</sup>	DNA <sup>Cy5,1T/Cy3,65B</sup>
$c_3$ (mol)		0.0001 ± 0.0001	0.015 ± 0.002	0.013 ± 0.001
$\tau_3^D$ ( $\mu$ s)		650	680 ± 40	650
$q_{3R}$ (kHz)		6.8	7.3 ± 0.7	6.8
$q_{3Y}$ (kHz)		6.7	6.8 ± 0.3	6.7

<sup>a</sup> Includes contributions from DNA<sup>Cy5dark,1T/Cy3,65B</sup>.

DNA<sup>Cy3,65B</sup> (to simulate interacting species). The consistency between the values extracted from fits of data from single-species samples and values extracted from multiple-species fits of mixtures demonstrates the ability of PAID to analyze mixtures of species. The same samples were also prepared with 10-fold higher occupancies (intermediate-occupancy samples; Supporting Information section 8S).

Figure 7 shows how to visually detect the presence of the double-labeled DNA<sup>Cy5,1T/Cy3,65B</sup> using PAID histograms. Two low-occupancy data sets are shown: DNA<sup>Cy5,1T/Cy3,65B</sup>/DNA<sup>Cy5,1T</sup>/DNA<sup>Cy3,65B</sup> (Figure 7, parts A, D, G, and J) and DNA<sup>Cy5,1T</sup>/DNA<sup>Cy3,65B</sup> (Figure 7, parts B, E, H, and K). By comparing the two sets of histograms, we can identify DNA<sup>Cy5,1T/Cy3,65B</sup>. Refer to section 4.4 and Figure 6 for a description of which species are emphasized by each histogram (DNA<sup>Cy3,65B</sup> corresponds to A<sup>x</sup>, DNA<sup>Cy5,1T</sup> to B<sup>x</sup>, and DNA<sup>Cy5,1T/Cy3,65B</sup> to A<sup>x</sup>B<sup>x</sup>). In Figure 7, because the occupancy of DNA<sup>Cy3,65B</sup> is higher than that of DNA<sup>Cy5,1T/Cy3,65B</sup>, the peaks corresponding to DNA<sup>Cy5,1T/Cy3,65B</sup> are less pronounced than those in Figure 6.

In Tables 5 and 6, we obtained fits with  $10 > \chi^2 > 1$  (ideally  $\chi^2 \sim 1$ ). Although the quality of the fits is sufficient for accurate and consistent extraction of parameters, the deviations indicated by the higher  $\chi^2$  sometimes give rise to biases (see below). Figure 8 shows a fit of the RYR histogram (same as Figure 7A, emphasizing DNA<sup>Cy5,1T/Cy3,65B</sup>) for the mixture in Table 6; this is one of the eight simultaneously fitted histograms ( $\chi^2 = 2.3$ ). Residuals show no large scale patterns (see Figure 8C and fit 1 in Figure 8, parts E and F).

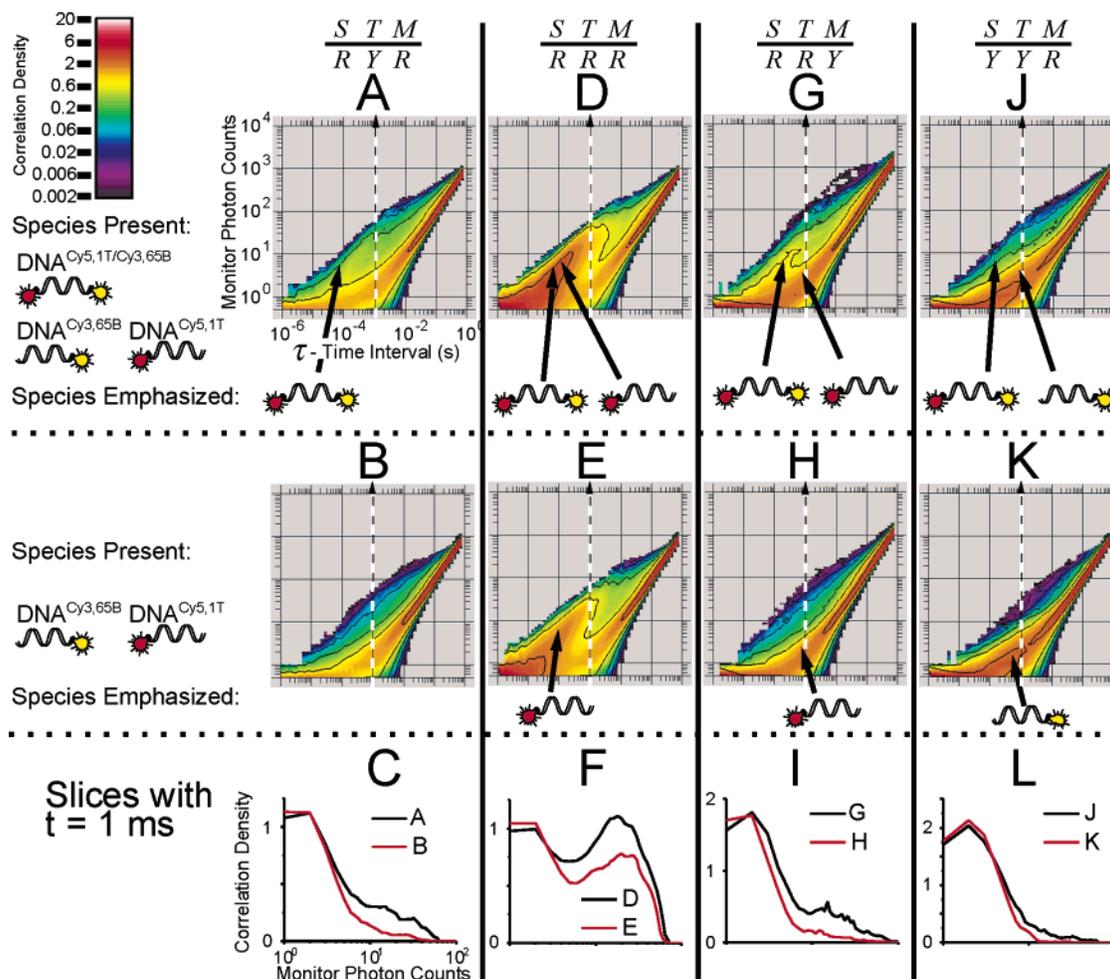
The deviations between model and fit can be attributed to two general sources. First, many photophysical properties are not yet modeled in PAID, such as triplet-state fluctuations and fluorescence saturation,<sup>27</sup> photobleaching,<sup>28,29</sup> and photoinduced isomerization.<sup>47</sup> Although time intervals less than 10  $\mu$ s are excluded from the fit to minimize photophysical effects, these properties still affect fitted values. Second, deviations of the experimental detection volume from the modeled volume may also play a role in the increased  $\chi^2$ .<sup>45</sup> For Tables 5 and 6, we used a non-Gaussian detectivity (section 3.5); when we used a

Gaussian detectivity, significantly higher  $\chi^2$  values were found. For example, fitting the three-species mixture data from Table 6 using the Gaussian volume resulted in fits with  $\chi^2 = 13$  (see residuals for “fit 2” in Figure 8D–F); the non-Gaussian volume yielded  $\chi^2 = 2.3$ .

Table 5 shows fitted values for the samples with one species of DNA fragments; all parameters were allowed to vary. To fit the data for DNA<sup>Cy5,1T/Cy3,65B</sup>, two components were necessary, one with Cy3 only and another with Cy3 and Cy5. This is due to nonfluorescent Cy5; single-species fits and inspection of time traces both indicate the presence of a species emitting only in the Cy3 channel.<sup>22,46</sup> A species with nonemitting Cy3 was not found. For the single-species data (Table 5), the extracted occupancies were 35% to 70% lower than expected from higher-concentration FCS experiments (note: for DNA<sup>Cy5,1T/Cy3,65B</sup>, the occupancies for DNA<sup>Cy5,1T/Cy3,65B</sup> and DNA<sup>Cy5dark,1T/Cy3,65B</sup> are added). The difference between extracted and estimated occupancies can be attributed to loss of DNA on surfaces during handling (see Supporting Information section 8S for intermediate-occupancy samples). Diffusion times extracted for DNA<sup>Cy5,1T/Cy3,65B</sup> and DNA<sup>Cy3,65B</sup> are similar to the theoretical ones (550–700  $\mu$ s vs 710  $\mu$ s); diffusion times for DNA<sup>Cy5,1T</sup> are shorter ( $\sim 400$   $\mu$ s), largely due to photoinduced isomerization,<sup>47</sup> although photobleaching of Cy5 within the detection volume and triplet-state fluctuations may play a role. Extracted brightness values are similar to those found in single-channel experiments.

Data from intermediate-occupancy samples were analyzed using FCS over the same range of time intervals, modeling only diffusion (Supporting Information section 8.3S). Variations in diffusion time were correlated variations found using PAID, indicating that the photophysical properties that affect FCS have similar effects on PAID.

We performed two fits for mixtures of DNA fragments simulating noninteracting and interacting species (Table 6). The first fit assumes the correct number of species but allows all parameters to freely vary; the second fit uses the single-species parameters already extracted to restrict the parameters for the



**Figure 7.** Dual-channel PAID histograms for samples containing species of labeled DNA fragments corresponding to those expected in an interaction study. The histograms chosen are the same as in Figure 6. Parts A, D, G, and J are PAID histograms for  $R Y R$ ,  $R R R$ ,  $R R Y$ , and  $Y Y R$  of a  $\text{DNA}^{\text{Cy5,1T}}$ ,  $\text{DNA}^{\text{Cy3,65B}}$ , and  $\text{DNA}^{\text{Cy5,1T/Cy3,65B}}$  mixture. Parts B, E, H, and K are PAID histograms of a  $\text{DNA}^{\text{Cy5,1T}}$  and  $\text{DNA}^{\text{Cy3,65B}}$  mixture. Differences in the histograms (due to the absence of  $\text{DNA}^{\text{Cy5,1T/Cy3,65B}}$ ) are pointed to by arrows. Parts C, F, I, and L compare vertical slices of the histograms at time interval  $\tau = 1$  ms. Because of the inactive component of Cy5, the occupancy of the dual-labeled species is significantly lower than the other species, lowering the peaks corresponding to the dual-labeled species (cf. Figure 6). See Figure 5S in the Supporting Information to see correlations corresponding to the collapse of these PAID histograms.

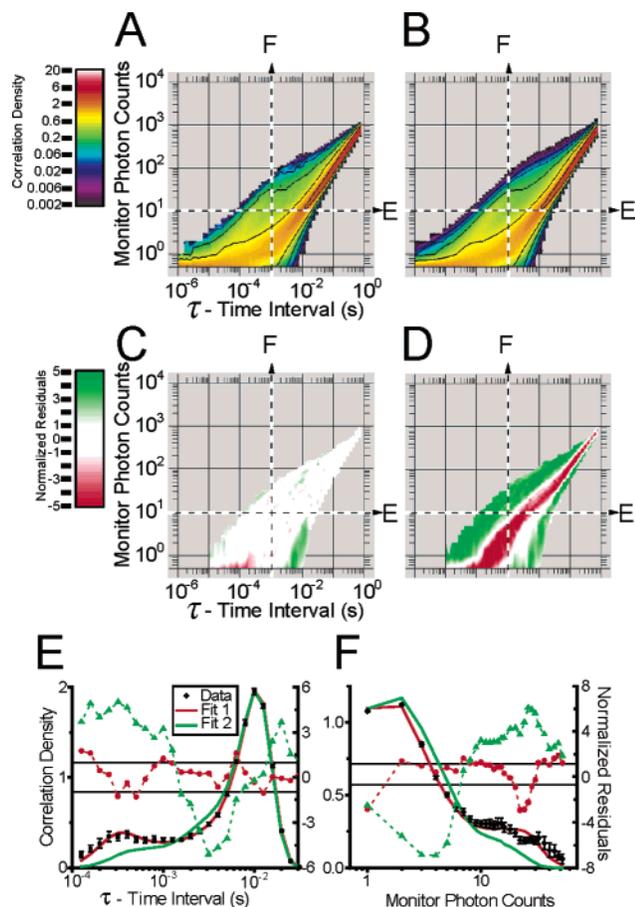
free components, except for occupancy. These fits show that a sample with two species can be distinguished from a sample with three species. PAID performed extremely well; the occupancies extracted from the mixtures using restricted fits were consistent with the occupancies extracted from the single-species samples (within 15%, except for  $\text{DNA}^{\text{Cy5,1T}}$ , which is 27% lower).

For the unrestricted fits, the occupancies extracted for  $\text{DNA}^{\text{Cy3,65B}}$  and  $\text{DNA}^{\text{Cy5,1T/Cy3,65B}}$  are consistent with the values obtained using the restricted fits (typically within 10%). However, the occupancy extracted for  $\text{DNA}^{\text{Cy5,1T}}$  in the three-species sample is smaller (30%), with compensating increases in the brightness extracted in  $R$ . The diffusion times and brightness values extracted are consistent with those found for the single-species samples (within 10%, except for the brightness of  $\text{DNA}^{\text{Cy5,1T}}$ , the three-species mixture, which is about 20% higher).

**4.6. Analysis of RNA-Polymerase–DNA Interactions.** We used PAID to study the formation of RNA-polymerase complexes with DNA, showing that PAID can resolve and quantitate species with differences in brightness and/or diffusion time. DNA was labeled with Cy5 (as the “red” fluorophore  $r$ ) to yield  $\text{DNA}^{\text{Cy5,65B}}$ ; RNAP was labeled with TMR (as the “yellow” fluorophore  $y$ ) on the  $\sigma$  subunit to yield  $R\sigma^{\text{TMR}}$ . In the  $R\sigma^{\text{TMR}}$ –

$\text{DNA}^{\text{Cy5}}$  complex, the distance between TMR and Cy5 is  $\gg 100$  Å, excluding interactions between the fluorophores (Murakami et al.).<sup>37</sup> We examined free  $\text{DNA}^{\text{Cy5}}$ , free  $R\sigma^{\text{TMR}}$ , and an interaction mixture that contained the  $R\sigma^{\text{TMR}}$ – $\text{DNA}^{\text{Cy5,65B}}$  complex along with free species (“RNAP + DNA” sample; Table 7). PAID analysis of  $\text{DNA}^{\text{Cy5,65B}}$  recovered brightness and diffusion times similar to the ones of the previous DNA fragments (8.9 kHz vs 9.4–10.2 kHz and 440  $\mu\text{s}$  vs  $\sim 400$   $\mu\text{s}$ , respectively). Analysis of  $R\sigma^{\text{TMR}}$  recovered brightness of  $\sim 10$  kHz, while the diffusion time of  $\sim 710$   $\mu\text{s}$  was larger than that of  $\text{DNA}^{\text{Cy5}}$  but smaller than the theoretical value ( $\sim 1$  ms; section 3.5). This may be due to the reasons mentioned in previous sections, or to imperfect modeling of the diffusion constant of the protein.

The RNAP + DNA sample shows the presence of four major species that are also observed after resolution of RNAP + DNA samples on polyacrylamide gels (Figure 9). For the RNAP–DNA interaction, distinct species can be identified on the basis of their gel mobility and their  $Y:R$  emission ratio (note that electrophoretic mobility depends on size and charge and is not simply related to free translational diffusion measured by PAID). The major species are characterized as (i) free DNA, (ii)  $\sigma$ -associated species, (iii) RNAP–DNA complex, and (iv) aggregates (a heterogeneous species with a large variety of sizes



**Figure 8.** PAID histogram with  $STM = RYR$  and fit (Table 6). (A) The PAID histogram for the data. (B) The fit to the PAID histogram using non-Gaussian detectivity (fit 1; 1 of 8 histograms fitted simultaneously). (C) A 2D map of the residuals for fit 1. (D) A 2D map of the residuals for the fit using Gaussian detectivity (fit 2). (E) Horizontal slices of the PAID histograms. Slices of the simulation are black; slices of the fits are red and green; residuals are dashed. (F) Vertical slices.

ranging from smaller oligomers to large aggregates). (This will be discussed in further detail in another manuscript, which is in preparation.) When a three-component PAID fit was performed on the RNAP + DNA sample (not shown), a single species with only *Y*-channel emission was extracted, with brightness corresponding to  $\sim 2$  copies of TMR per species. This result was inconsistent with the biochemical system (Figure 9), with visual inspection of the PAID histograms, and with fluorescence intensity time traces showing few exceptionally bright bursts in the *Y* channel (such bursts were absent in the buffer or  $\text{DNA}^{\text{Cy5,65B}}$  and were extremely rare in the  $\text{Ro}^{\text{TMR}}$  sample). Therefore, a four-component PAID fit was performed on the RNAP + DNA data, recovering four species with diffusion times and *Y*:*R* brightness ratios that correspond well to gel-resolved species of RNAP + DNA samples (Figure 9). This fit yielded additional information inaccessible to ensemble fluorescence imaging or other FFS methods.

Species 1 has properties consistent with those of free  $\text{DNA}^{\text{Cy5,65B}}$  (emission only in the *R* channel, corresponding to a single copy of Cy5); we identify this species as “free DNA”. Species 2 shows emission only in the *Y* channel (corresponding to a single copy of TMR) and diffusion time smaller than the RNAP–DNA complex or holoenzyme but similar to DNA (compared with the 600–700  $\mu\text{s}$  found with Cy3 in previous sections); we identify this species as a combination of a “ $\sigma$ -associated species” generated during the preparation of RNAP–

DNA complex (section 3.2) and of RNAP–DNA complex with nonfluorescent Cy5 (analogous to  $\text{DNA}^{\text{Cy5dark,IT/Cy3,65B}}$  in section 4.5). Species 3 shows comparable emission at both *Y* and *R* channels (corresponding to a single copy of TMR and a single copy of Cy5) and diffusion time (790  $\mu\text{s}$ ) slightly larger than the free holoenzyme (710  $\mu\text{s}$ ) and larger than free DNA (440–480  $\mu\text{s}$  with Cy5 attached, 550–700  $\mu\text{s}$  with Cy3); we identify this species as “RNAP–DNA complex”. Species 4 shows emission only at the *Y* channel, with high brightness (corresponding to a mean of  $\sim 3$  copies of TMR per species), and has the longest average diffusion time (850  $\mu\text{s}$ ) of all species; we identify this species as “aggregates”. Species 4 is a heterogeneous species, likely containing aggregates with varying numbers of subunits; different sections of data contained bursts of highly variable height and duration (most of the bright bursts from species 4 are of  $\sim 1$  ms duration; however, we have observed “yellow-labeled” species that generated bursts longer than 1 s with a peak height of 60 kHz). This variability is manifested in the PAID fits by the larger uncertainties found for the brightness and diffusion time of species 4. PAID also recovered the occupancy of the various species; however, since PAID measurements are performed in a different concentration regime, matrix, and buffer than the gel mobility assays, we do not necessarily expect quantitative agreement of PAID results with the gel results, and therefore, this comparison was not performed.

The qualitative agreement between PAID and gel imaging in terms of molecular size (measured as PAID-based diffusion time or gel-based electrophoretic mobility) and relative brightness (measured as the ratio of PAID-based molecular brightness or gel-based fluorescence intensity ratio) of the major species validates the use of PAID for analysis of complex mixtures. Such a complicated mixture will present a great challenge for other popular FFS methods, such as FCS, FCCS, PCH/FIDA, 2D-FIDA, or FIMDA. These methods are not able to simultaneously monitor the diffusion time and brightness on both channels, preventing the reliable detection and characterization of the species obtained using PAID (2D-FIDA would be able to separate and identify the species, but would lack the information on diffusion time). PAID reliably separates the species by extracting the relative brightness in the two channels; the diffusion time assists in the separation but also proves very useful in species identification. For example, the diffusion time provides additional evidence for the discrimination between the aggregates and the lower brightness,  $\sigma$ -associated species that emit in the same channel.

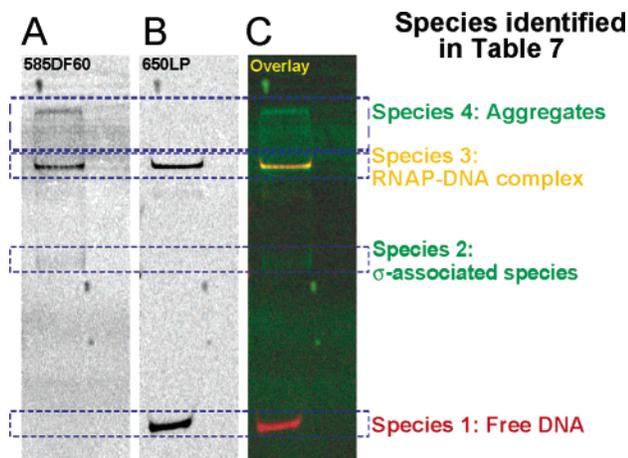
## 5. Discussion

Because of their ability to differentiate distinct species and extract their individual properties, FFS methods are valuable tools in the analysis of macromolecular interactions. We have introduced PAID, a multidimensional method applicable to multiple channels, and have demonstrated robust, simultaneous extraction of occupancy, diffusion time, and brightness in single- and dual-channel formats for multiple species. Such simultaneous extraction of occupancy, diffusion time, and brightness, which improves identification and characterization, has previously been performed in only one channel.<sup>14</sup> The PAID histogram provides a visual representation that emphasizes the interesting features of the species such as occupancy, brightness, and diffusion time. The modeling technique used for PAID accurately models the possible diffusion paths through any known detectivity, allowing the direct application of diffraction calculations or experimentally determined detectivity. The model

**TABLE 7: Parameters Extracted Using PAID Fits for the RNAP–DNA Interaction**

parameters	averaged fits for 10 measurements (30 s each): low occupancy		
	DNA	RNAP	RNAP + DNA sample
$\chi^2$	$1.3 \pm 0.1$	$0.8 \pm 0.1$	$0.7 \pm 0.1$
$k_{0R}$ (kHz)	$1.0 \pm 0.1$	$0.9 \pm 0.1$	$1.2 \pm 0.1$
$k_{0Y}$ (kHz)		$1.4 \pm 0.1$	$1.3 \pm 0.1$
species 1	DNA <sup>Cy5,1T</sup>	R $\sigma$ <sup>TMR</sup>	free DNA
$c_1$ (mol)	$0.23 \pm 0.01$	$0.18 \pm 0.01$	$0.11 \pm 0.01$
$\tau_1^D$ ( $\mu$ s)	$440 \pm 10$	$710 \pm 30$	$480 \pm 20$
$q_{1R}$ (kHz)	$8.9 \pm 0.1$	$1.4 \pm 0.1$	$12.4 \pm 0.7$
$q_{1Y}$ (kHz)		$12.4 \pm 1.5$	$0.4 \pm 0.2$
species 2			$\sigma$ -associated species <sup>a</sup>
$c_2$ (mol)			$0.30 \pm 0.02$
$\tau_2^D$ ( $\mu$ s)			$590 \pm 50$
$q_{2R}$ (kHz)			$0.9 \pm 0.1$
$q_{2Y}$ (kHz)			$14.0 \pm 0.5$
species 3			RNAP–DNA complex
$c_3$ (mol)			$0.11 \pm 0.01$
$\tau_3^D$ ( $\mu$ s)			$790 \pm 30$
$q_{3R}$ (kHz)			$10.0 \pm 0.1$
$q_{3Y}$ (kHz)			$9.3 \pm 0.2$
species 4			aggregates
$c_4$ (mol)			$0.05 \pm 0.01$
$\tau_4^D$ ( $\mu$ s)			$850 \pm 70$
$q_{4R}$ (kHz)			$3.3 \pm 0.3$
$q_{4Y}$ (kHz)			$35 \pm 3$

<sup>a</sup> Includes contributions from RNAP–DNA complex with nonfluorescent Cy5 attached to DNA.



**Figure 9.** Fluorescence-based gel images showing the major species present in the RNAP + DNA samples. (A) Species with TMR-based signals (532 excitation, 585BP60 emission); species identified as aggregates (found in the wells of the gel), RNAP–DNA complex, and  $\sigma$ -associated species are visible. (B) Species with Cy5-based signals (633 excitation, 650LP emission); species identified as RNAP–DNA complex and DNA are visible. (C) overlay of images A (green color) and B (red color); coincidence of the red and green signal is shown as orange. The only species showing coincidence is the RNAP–DNA complex.

works naturally with logarithmic axes, allowing PAID to be used over a wide dynamic range. With simulations, PAID has been shown to match the statistical accuracy of existing methods while providing additional information.

In single-channel applications, PAID monitors occupancy, diffusion time, and brightness of several species as functions of experimental conditions or time, allowing monomeric species to be distinguished from oligomeric species (Figure 1A). With our experiments on DNA model systems labeled with Cy3B, we obtained a basic result needed to demonstrate that FFS can extract stoichiometry by using brightness to count fluorophores: that a pure species with two fluorophores is measured

to be twice as bright as a pure species with one fluorophore. Previous studies suggested that this is possible, but since no systematic attempts to control stoichiometry or purity were undertaken, their results cannot be considered as rigorous tests of the ability of FFS to extract stoichiometry.<sup>7,12</sup> In contrast, the careful design, construction, and purification of the DNA model systems used in the present work, combined with PAID analysis, establishes this capability. The experiments also demonstrated that PAID is capable of detecting heterogeneity in brightness, while extracting diffusion times.

FIMDA is the only previous method that has capabilities similar to single-channel PAID. PAID and FIMDA show similar accuracy for most of the simulations, but FIMDA shows higher biases for some parameters extracted in low-occupancy conditions (the more important experimental comparison has not been performed). On the other hand, FIMDA already accounts for triplet-state fluctuations, whereas these remain to be incorporated into the PAID model. It is possible to use a simultaneous fit of FCS and FIDA/PCH to extract information about brightness and diffusion time simultaneously.<sup>13</sup> However, our fits show that, even if FIDA/PCH fixes the occupancy and brightness of two species to their correct values, FCS is not able to extract the values for the diffusion times as reliably as PAID or FIMDA.

Correlated increases in brightness and diffusion time can help verify molecular interactions, providing more evidence than if brightness and diffusion time are determined individually. With its capabilities, single-channel PAID can be applied to areas such as assembly and stoichiometry of membrane proteins (receptors, ion channels)<sup>48</sup> and amyloid plaque formation.<sup>49,50</sup> FCS has been applied to the characterization of amyloid  $\beta$ -peptide polymerization,<sup>51,52</sup> where the presence of aggregates was detected as large changes in diffusion times. Some arguments were presented about the height of fluorescence bursts indicating aggregation, but this was not quantified rigorously. Single-channel PAID would allow brightness and diffusion time to be quantified, providing two observables for the degree of aggregation.

Dual-channel methods provide improved sensitivity over single-channel methods for the analysis of the interactions using FFS.<sup>2,15–17</sup> We demonstrated PAID-based extraction of coincidence, diffusion time, brightness, and occupancy of several species in a single data set, which has not been done with previous methods. We used control experiments both to check consistency in unrestricted fits and to restrict certain parameters, increasing confidence in the remaining fitted parameters. The PAID analysis of the RNAP–DNA interaction demonstrated the ability of PAID to analyze complex systems, detecting concentration, brightness, and diffusion time for multiple species present in a mixture. This ability allows quantitation of the various free and bound species present in equilibrium binding reactions, thus paving the way for generating binding constants between interacting partners in a high-throughput, low-volume assay format (with suitable simplifications and optimizations of the PAID method).

Experimental complications were encountered that may skew results: likely adhesion of molecules to the surface, nonfluorescent (possibly photobleached) Cy5, and photoinduced isomerization. Improved results are expected with the use of passivated surfaces and improved fluorophores.

By extending FCCS to PAID, we can use brightness information to discriminate between species and background. In FCCS, leakage and background can hamper extraction of the occupancy of complexes;<sup>2</sup> although control experiments can extract leakage and background, it is often essential to extract all parameters from one data set, requiring control experiments only for checking consistency. PAID can often extract all parameters simultaneously and will be indispensable in cases where precise control experiments are impractical or impossible. For example, cellular autofluorescence changes as a function of position, preventing background measurements for a particular spatial position in a cell that contains different fluorescent species. FCCS was used to monitor the endocytosis of Cy2- and Cy5-labeled cholera toxin (CTX) into cells.<sup>53</sup> PAID can improve the characterization of the species most importantly by determining the brightness in each channel and the number of CTX subunits per vesicle.

PAID provides a visual representation that focuses on critical features of diffusing species over a large dynamic range. This is vital, since clear visual features in histograms provide the first clues to interesting findings (such as additional subpopulations) or experimental problems. The PAID histogram emphasizes photon-rich time intervals, retaining the intuitive nature of correlation functions, while simultaneously providing the information in photon-counting histograms. In contrast, the photon-counting histograms used in FIDA/PCH and FIMDA weigh all time intervals equally; at low occupancy, most of the bins correspond to time intervals when no molecule is present and few photons are counted.

The modeling technique we developed permits the use of PAID over a large dynamic range both in the temporal and photon count dimensions. The PAID model is able to account for all the possible trajectories through an arbitrary detection volume. The model used for the detectivity has a large impact on the accuracy of PAID. Using our numerically calculated detectivity rather than a Gaussian detectivity improved the value of  $\chi^2$  considerably. Similar effects are found in FCS and other FFS methods.<sup>45</sup> Improvement in the quality of PAID fits will come from better characterization of the detection volume. FIDA models the detection volume as a volume density per intensity level; a polynomial expression is used whose parameters are determined using experiments on a standard sample. Although

attractive for its simplicity, this prevents the modeling of possible trajectories through the detection volume used in PAID. Comparisons between how FIDA and FIMDA, PCH, and PAID account for non-Gaussian detection volumes differently will be interesting and will be the subject of a future study.

PAID can impact measurements where simultaneous determination of coincidence and diffusion (or other temporal dynamics) are critical. We plan to use PAID to monitor protein–protein interactions and protein–DNA interactions. We also plan to extend the PAID model to include photophysical properties of the dyes (triplet-state induced blinking, singlet and triplet-state saturation, photobleaching), and to incorporate an experimentally measured detectivity; these improvements will increase confidence in additional dynamics or species found during fitting. We will extend the model to account for two monitor channels (see sections 2.1 and 7.2S) and to become compatible with studies of immobilized molecules and systems involving flow. Such improvements will allow in vitro analysis of protein–protein interactions and protein–DNA interactions, as well as analysis of interactions in a cellular environment.

**Acknowledgment.** We thank Emmanuel Margeat for assistance with the analysis of RNAP–DNA interactions, Xavier Michalet and Sören Doose for useful discussions, and Professor Tomasz Heyduk for the plasmid used for expressing  $\sigma^{Cys596}$ . This research was supported by the Biological and Environmental Research Program (BER), U.S. Department of Energy, Grant No. DE-FG03-02ER63339.

**Supporting Information Available:** Bin specification and normalization for PAID histograms; algorithms for constructing PAID histograms; PAID function model; figure of numerically approximated detection volume; single-channel, intermediate-occupancy simulations and modified FIMDA; single-channel, intermediate-occupancy experiments and initial experiments using Cy3; simulations of quantitative analysis of binding using dual-channel PAID and PAID histograms with two monitor channels; correlations for dual-channel, low-occupancy experiments and PAID and FCS analysis of dual-channel, intermediate-occupancy experiments. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Magde, D.; Elson, E.; Webb, W. W. *Phys. Rev. Lett.* **1972**, *29*, 705.
- (2) Schwill, P.; Meyer-Almes, F. J.; Rigler, R. *Biophys. J.* **1997**, *72*, 1878.
- (3) Palmer, A. G., III; Thompson, N. L. *Biophys. J.* **1987**, *52*, 257.
- (4) Qian, H.; Elson, E. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 5479.
- (5) Qian, H.; Elson, E. *Appl. Polym. Symp.* **1989**, *43*, 305.
- (6) Chen, Y.; Müller, J. D.; So, P. T.; Gratton, E. *Biophys. J.* **1999**, *77*, 553.
- (7) Kask, P.; Palo, K.; Ullmann, D.; Gall, K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13756.
- (8) Brooks Shera, E.; Seitzinger, N. K.; Davis, L. M.; Keller, R. A.; Soper, S. A. *Chem. Phys. Lett.* **1990**, *174*, 553.
- (9) Nie, S.; Chiu, D. T.; Zare, R. N. *Science* **1994**, *266*, 1018.
- (10) Levene, M. J.; Korlach, J.; Turner, S. W.; Foquet, M.; Craighead, H. G.; Webb, W. W. *Science* **2003**, *299*, 682.
- (11) Qian, H.; Elson, E. L. *Biophys. J.* **1990**, *57*, 375.
- (12) Muller, J. D.; Chen, Y.; Gratton, E. *Biophys. J.* **2000**, *78*, 474.
- (13) Chen, Y.; Muller, J. D.; Ruan, Q. Q.; Gratton, E. *Biophys. J.* **2002**, *82*, 133.
- (14) Palo, K.; Mets, U.; Jäger, S.; Kask, P.; Gall, K. *Biophys. J.* **2000**, *79*, 2858.
- (15) Kask, P.; Palo, K.; Fay, N.; Brand, L.; Mets, U.; Ullmann, D.; Jungmann, J.; Pschorr, J.; Gall, K. *Biophys. J.* **2000**, *78*, 1703.
- (16) Dahan, M.; Deniz, A. A.; Ha, T.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Chem. Phys. (Netherlands)* **1999**, *247*, 85.

- (17) Eggeling, C.; Fries, J. R.; Brand, L.; Günther, R.; Seidel, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1556.
- (18) Heinze, K. G.; Koltermann, A.; Schwille, P. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10377.
- (19) Keller, R. A.; Ambrose, W. P.; Goodwin, P. M.; Jett, J. H.; Martin, J. C.; Wu, M. *Appl. Spectrosc.* **1996**, *50*, A12.
- (20) Tellinghuisen, J.; Goodwin, P. M.; Ambrose, W. P.; Martin, J. C.; Keller, R. A. *Anal. Chem.* **1994**, *66*, 64.
- (21) Fries, J. R.; Brand, L.; Eggeling, C.; Kollner, M.; Seidel, C. A. M. *J. Phys. Chem. A* **1998**, *102*, 6601.
- (22) 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.
- (23) Reynaud, S. *Ann. Phys.* **1983**, *8*, 315.
- (24) Edman, L.; Rigler, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8266.
- (25) Rigler, R.; Mets, U.; Widengren, J.; Kask, P. *Eur. Biophys. J.* **1993**, *22*, 169.
- (26) Ehrenberg, M.; Rigler, R. *Chem. Phys. (Netherlands)* **1974**, *4*, 390.
- (27) Widengren, J.; Mets, U.; Rigler, R. *J. Phys. Chem.* **1995**, *99*, 13368.
- (28) Widengren, J.; Rigler, R. *Bioimaging* **1996**, *4*, 149.
- (29) Dittrich, P. S.; Schwille, P. *Appl. Phys. B* **2001**, *73*, 829.
- (30) Creighton, T. E. *Proteins: Structures and Molecular Principles*; W. H. Freeman: New York, 1983.
- (31) Kimble, H. J.; Dagenais, M.; Mandel, L. *Phys. Rev. Lett.* **1977**, *39*, 691.
- (32) Mandel, L. *Proc. Phys. Soc.* **1958**, *72*, 1037.
- (33) Mandel, L. *Proc. Phys. Soc.* **1959**, *74*, 233.
- (34) Enderlein, J.; Robbins, D. L.; Ambrose, W. P.; Goodwin, P. M.; Keller, R. A. *J. Phys. Chem. B* **1997**, *101*, 3626.
- (35) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.
- (36) Mukhopadhyay, J.; Kapanidis, A. N.; Mekler, V.; Kortkhonjia, E.; Ebricht, Y. W.; Ebricht, R. H. *Cell* **2001**, *106*, 453.
- (37) Murakami, K. S.; Masuda, S.; Campbell, E. A.; Muzzin, O.; Darst, S. A. *Science* **2002**, *296*, 1285.
- (38) Wohland, T.; Rigler, R.; Vogel, H. *Biophys. J.* **2001**, *80*, 2987.
- (39) Wolf, E. *Proc. Phys. Soc. A* **1959**, *253*, 349.
- (40) Hell, S.; Reiner, G.; Cremer, C.; Stelzer, E. H. K. *J. Microsc. (Oxford)* **1993**, *169*, 391.
- (41) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; W. H. Freeman: San Francisco, 1980.
- (42) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Clarendon Press and Oxford University Press: Oxford, U.K. and NY, 1988.
- (43) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 3rd electronic ed.; CRC Press: Boca Raton, FL, 2001.
- (44) Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman and Hall: New York, 1993.
- (45) Hess, S. T.; Webb, W. W. *Biophys. J.* **2002**, *83*, 2300.
- (46) 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.
- (47) Widengren, J.; Schwille, P. *J. Phys. Chem. A* **2000**, *104*, 6416.
- (48) Milligan, G. *Science* **2000**, *288*, 65.
- (49) Bieschke, J.; Giese, A.; Schulz-Schaeffer, W.; Zerr, I.; Poser, S.; Eigen, M.; Kretzschmar, H. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5468.
- (50) Cohen, F. E. *J. Mol. Biol.* **1999**, *293*, 313.
- (51) Tjernberg, L. O.; Pramanik, A.; Bjorling, S.; Thyberg, P.; Thyberg, J.; Nordstedt, C.; Berndt, K. D.; Terenius, L.; Rigler, R. *Chem. Biol.* **1999**, *6*, 53.
- (52) Pitschke, M.; Prior, R.; Haupt, M.; Riesner, D. *Nat. Med.* **1998**, *4*, 832.
- (53) Bacia, K.; Majoul, I. V.; Schwille, P. *Biophys. J.* **2002**, *83*, 1184.