Disentangling fluorescence signals from diffusing single molecules by independent component analysis

Kunihiko Ishii[1,2,*], Miyuki Sakaguchi[1,†], and Tahei Tahara[1,2]

[1]*Molecular Spectroscopy Laboratory, RIKEN, 2-1 Hirosawa, Wako 351-0198, Japan*

[2]*Ultrafast Spectroscopy Research Team, RIKEN Center for Advanced Photonics (RAP), 2-1 Hirosawa, Wako 351-0198, Japan*

[*]e-mail: kishii@riken.jp

[†]Present address: Graduate School of Science & Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan

Abstract

Multiparameter measurements offer new possibilities of single-molecule fluorescence experiments for detecting and quantifying heterogeneity in freely diffusing molecules. However, data analysis remains challenging due to limited photon rates from individual molecules. Here, we present independent fluorescence component analysis (IFCA), a universal analytical framework that applies independent component analysis (ICA) to unmix multiparameter fluorescence signals. In this work, we developed an efficient algorithm of ICA for multiparameter photon data based on third-order cumulant tensor decomposition. To validate its performance, we applied IFCA to a static mixture of fluorescent dyes and found that IFCA enables robust separation of signals from five species in a nanomolar mixture solution with a 5 μs binning time. We further analyzed a photon dataset from a FRET-labeled DNA construct using IFCA, demonstrating its capability to quantitatively characterize signals from subpopulations exchanging on a sub-millisecond time scale in a model-independent manner. These findings highlight the high ability of IFCA for quantitative analysis of complex molecular systems exploiting these advantages.

Single-molecule fluorescence (SMF) methods, in particular single-molecule Förster resonance energy transfer (smFRET), are invaluable tools in current biophysical/bioanalytical chemistry research for detecting the heterogeneity and fluctuations of molecular structures, which are inaccessible with ensemble measurements.[1, 2] The major technical challenge in SMF experiments on freely diffusing molecules comes from the intrinsically weak signals that hamper the quantitative analysis of the fluorescence characteristics of constituent species and their relative populations. For deriving maximum information from a limited amount of fluorescence photons, intrinsic properties of fluorescence have been incorporated into SMF experiments, such as excitation/emission wavelengths, emission delay time after pulsed excitation (microtime), and polarization.[3-9] Photon data obtained with such multiparameter experiments are analyzed with the help of dedicated methods designed for individual experimental configurations, such as photon-by-photon hidden Markov modeling (H^2MM)[10-12] and FRET-lines.[13-15] However, there has been no general approach to photon data analysis that is compatible with an arbitrary combination of fluorescence parameters without assuming a physical model. Moreover, to apply these statistical methods, it is typically necessary to collect at least several tens of photons from each single molecule in a well-isolated condition. This requirement has been limiting the applicability of previous multiparameter SMF methods in terms of sample concentration (~10 pM)[16] and time resolution that is determined by temporal bin size or the duration of single-molecule bursts (>~1 ms).[17] Also, the brightness of fluorophores needs to be sufficiently high. These factors make it challenging to study molecular events in harsh conditions such as cellular environments.

To solve these problems, we report a new model-free method for unmixing signals in multiparameter SMF experiments based on a multivariate analysis method of signal fluctuations known as independent component analysis (ICA). ICA realizes blind separation of statistically independent components in an observed multiparameter signal by utilizing their non-gaussian

property. It has been widely adopted in other fields, e.g., feature extraction, brain imaging, and telecommunications.[18] Because the fluorescence signals from freely diffusing single molecules recorded with a confocal microscope exhibit Poisson-like non-gaussian behavior,[19] ICA is naturally expected to be applicable also to SMF data analysis. Here, we implemented an ICA for SMF by using the third-order cumulant tensor[20, 21] of temporal fluctuations in multiparameter fluorescence signals. We call this analysis IFCA (independent fluorescence component analysis) hereafter.

In this paper, after describing the principle and actual procedures of IFCA, we demonstrate the ability of IFCA by applying it to photon data obtained with multiparameter fluorescence measurements. First, we analyze photon data of a mixture of five fluorescent dyes in nanomolar concentrations. This example illustrates the capability of IFCA in separating individual species and characterizing the fluorescence signal of each species in a high concentration regime. Then, we examine photon data of a FRET-labeled DNA construct that undergoes transitions between two conformational states in a sub-millisecond time scale,[22] in which transient species in a dynamic system are identified utilizing the microsecond time resolution achievable with IFCA.

Results and Discussion

Multiparameter photon data for IFCA

In a multiparameter single-molecule fluorescence measurement (Fig. 1a), a dilute solution of fluorescent molecules is irradiated under a confocal microscope by light pulses of either a single color or multiple colors with temporal interleaving. The emitted fluorescence photons are sorted according to their colors, λ_{em}, using dichroic mirrors and/or their polarizations using a polarizer and counted by separate detectors. The microtime, t, and the excitation wavelength in case of multi-color excitation, λ_{ex}, of each detected photon are determined from the detection delay

time measured with a time-correlated single-photon counting (TCSPC) device. All the photons are also tagged with their absolute arrival times (macrotime), T, and stored as multiparameter photon data (Fig. 1a, right panel).[23] In this data, λ_{em} and polarization are indicated by the detector index c (for example, $c = 0$ for the donor detector and $c = 1$ for the acceptor detector in a FRET experiment). In case of multi-color excitation, λ_{ex} is judged from the value of t recorded by the TCSPC module (for example, $t \leq t_1$ corresponds to a photon emitted after excitation with the first color and $t > t_1$ corresponds to a photon emitted after excitation with the second color). We collectively denote these fluorescence parameters as a one-dimensional variable, $x = ct_0 + t$, where t_0 is the number of TCSPC channels, and x takes an integer value between 1 and $x_0 = c_0 t_0$ (c_0 is the number of detectors). The goal of IFCA is to obtain the characteristic fluorescence patterns of independent fluorescence components (IFCs)

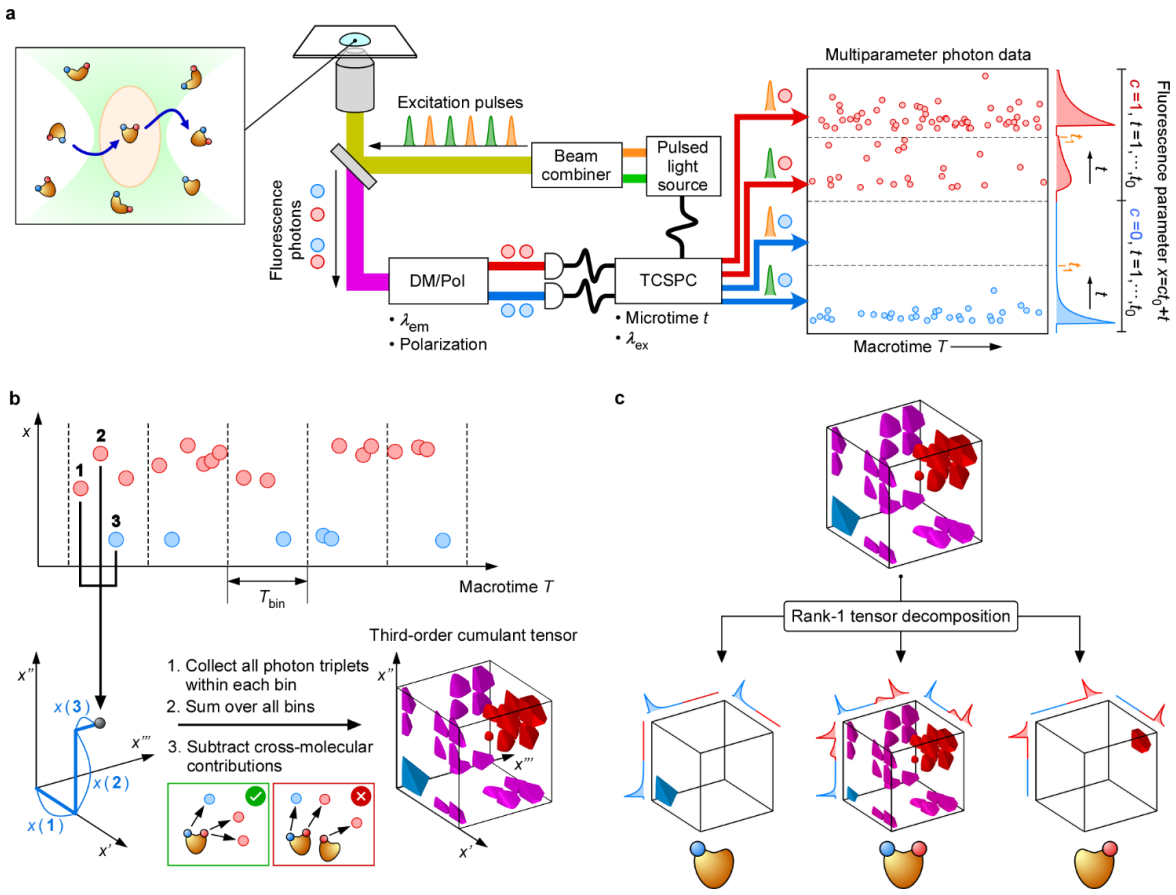

Figure 1. Principle of IFCA. (a) Schematic illustration of a multi-parameter SMF measurement of freely diffusing molecules. DM: dichroic mirror, Pol: polarizer. (b) Construction of a third-order cumulant tensor. (c) Decomposition of a third-order cumulant tensor into a sum of independent fluorescence components.

represented as functions of x.

Principle of IFCA

IFCA analyzes the fluctuation of fluorescence intensity at channel x, $I_x(T)$, as a function of macrotime *T*. For this purpose, we utilize third-order cross-cumulants of $I_x(T)$ that are defined as:[18]

$$\begin{aligned} &\mathrm{cum}\big[I_i(T), I_j(T), I_k(T)\big] \\ &= \mathrm{E}\big[\{I_i(T) - \bar{I}_i\}\{I_j(T) - \bar{I}_j\}\{I_k(T) - \bar{I}_k\}\big] \\ &= \frac{1}{T_0}\int_0^{T_0}[I_i(T) - \bar{I}_i]\big[I_j(T) - \bar{I}_j\big][I_k(T) - \bar{I}_k]\, dT\,, \end{aligned} \tag{1}$$

where $\mathrm{E}[\cdot]$ denotes an expectation value, $\bar{I}_x \equiv \mathrm{E}[I_x(T)] = \frac{1}{T_0}\int_0^{T_0} I_x(T)\, dT$ is the time-integrated average of $I_x(T)$, and T_0 is the total measurement time. The third-order cumulant tensor is defined as a three-dimensional array $\boldsymbol{\mathcal{T}}$, whose (i,j,k)-element is $\mathcal{T}_{ijk} \equiv \mathrm{cum}\big[I_i(T), I_j(T), I_k(T)\big]$ (Fig. 1b). In IFCA, we exploit an important property of cumulant tensors: the additivity.[18] That is, a third-order cumulant tensor is decomposed as a sum of multiple rank-1 tensors, each of which is represented as the triple direct product of the vector representing the characteristic fluorescence pattern of a constituent species in the sample (Fig. 1c):

$$\mathcal{T}_{ijk} = \sum_{s=1}^{n} \mathcal{T}_{ijk}^{s} = \sum_{s=1}^{n} d_i^s d_j^s d_k^s N_s\,, \tag{2}$$

where n is the number of species, N_s is the average number of molecules of species s in the observation volume, and d_x^s is the characteristic fluorescence pattern of a single molecule of species s. For third-order cumulant tensors, there exists another useful property, i.e., uniqueness of decomposition.[24] Because of this property, the above decomposition of $\boldsymbol{\mathcal{T}}$ into the contributions of independent sources, $\boldsymbol{\mathcal{T}}^s$, is uniquely defined. Based on these properties of cumulant tensors, we realize IFCA in two steps: Construction of the third-order cumulant tensor

$\boldsymbol{\mathcal{T}}$ from multiparameter photon data (Fig. 1b) and decomposition of $\boldsymbol{\mathcal{T}}$ into independent components (Fig. 1c).

Procedure of IFCA

Construction of the third-order cumulant tensor from multiparameter photon data. To derive third-order cumulants as defined in Eq. 1 from multiparameter photon data, it is necessary to account for the quantum nature of light.[25] In the fluctuation analyses of fluorescence photons, it is known that the factorial cumulants of photon counts should be employed instead of ordinary cumulants of fluorescence intensity.[19, 26] In the present case, this can be accomplished by evaluating $\boldsymbol{\mathcal{T}}$ from photon counts as follows:

$$\mathcal{T}_{ijk} = \mathrm{E}\left[n_i\left(n_j - \delta_{ij}\right)\left(n_k - \delta_{ik} - \delta_{jk}\right)\right] - \mathrm{E}[n_i]\mathrm{E}\left[n_j\left(n_k - \delta_{jk}\right)\right] - \mathrm{E}\left[n_j\right]\mathrm{E}[n_i(n_k - \delta_{ik})]$$
$$-\mathrm{E}[n_k]\mathrm{E}\left[n_i\left(n_j - \delta_{ij}\right)\right] + 2\mathrm{E}[n_i]\mathrm{E}\left[n_j\right]\mathrm{E}[n_k], \quad (3)$$

where n_x indicates the photon counts per bin time at channel x and $\delta_{xx'}$ is the Kronecker's delta. Computation of these elements are performed as follows (Fig. 1b).

(i) The macrotime axis of the photon data is binned into a preset temporal width, T_{bin}.

(ii) In each bin, photon triplets are searched, and once found, their x values are histogrammed in a three-dimensional coordinate space.

(iii) The obtained three-dimensional histograms for all bins are summed up and then divided by the total number of bins.

(iv) If there are multiple data files recorded in the same measurement condition, the three-dimensional histograms may be averaged over these files.

(v) Then, all the possible permutations of the indices (i, j, k) are considered, and the obtained six histograms are added up.

It is easy to see that the resultant three-dimensional array corresponds to the first term of Eq. 3. Note that the decrements of n_x by $\delta_{xx'}$ in Eq. 3 are equivalent to avoiding double (or triple) counting of the same photon in a photon triplet in step (ii). The second to fifth terms

of Eq. 3 can be evaluated and subtracted from the three-dimensional array by using the first- and second-order cumulant tensors, $\mathbf{v}$ and $\mathbf{M}$:

$$v_i \equiv \mathrm{cum}[I_i(T)] = \mathrm{E}[n_i], \tag{4}$$

$$M_{ij} \equiv \mathrm{cum}\big[I_i(T), I_j(T)\big] = \mathrm{E}\big[n_i\big(n_j - \delta_{ij}\big)\big] - \mathrm{E}[n_i]\mathrm{E}\big[n_j\big], \tag{5}$$

which are computed by histogramming photons and photon pairs, respectively, in a similar manner to the above steps (i)-(v) for deriving $\mathcal{T}_{ijk}$. Intuitively, the second to fifth terms of Eq. 3 and the second term of Eq. 5 are corresponding to cross-molecular contributions, i.e., photon triplets (pairs) consisting of photons from different molecules (Fig. 1b), and those involving uncorrelated background counts.[27] After subtracting these terms in Eq. 3, a third-order cumulant tensor is obtained that can be decomposed into independent components. In practice, it is sometimes necessary for obtaining the best results to consider additional effects due to molecular dynamics within each bin and detector nonideality (Computational Methods 1 in Supporting Information).

Decomposition of the third-order cumulant tensor into independent components. Next, we decompose the third-order cumulant tensor into the sum of rank-1 tensors. Though this decomposition is known to be unique provided that all the fluorescence patterns are linearly independent of each other,[24] the numerical procedure is not as well established as its matrix analogues such as singular value decomposition. Here, we employ a standard method in ICA for preprocessing of higher-order cumulant tensors[28] utilizing the eigenvalue decomposition (EVD) of the second-order cumulant tensor. In this procedure, we first apply noise normalization across channels to make the noise level uniform over x. Because the noise level at a certain channel is proportional to the square root of the counts registered, the normalization factor is set to $\sqrt{v_x}$ by using the first-order cumulant tensor $\mathbf{v}$, i.e., the ensemble-averaged fluorescence pattern. Thus, the raw cumulant tensors $\boldsymbol{\mathcal{T}}, \mathbf{M}$ derived from experimental photon data are scaled as

$$\mathcal{T}'_{ijk} \equiv \frac{\mathcal{T}_{ijk}}{\sqrt{v_i v_j v_k}}, \tag{6}$$

$$M'_{ij} \equiv \frac{M_{ij}}{\sqrt{v_i v_j}}. \tag{7}$$

EVD of the scaled second-order cumulant tensor is

$$\mathbf{M}' = \mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}}, \tag{8}$$

where $\mathbf{B}$ is a matrix combining the eigenvectors as $\mathbf{B} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{x_0})$ and $\mathbf{W}$ is a diagonal matrix with the diagonal elements W_p being eigenvalues in the descending order, and superscript T indicates matrix transpose. In the following analysis, only the eigen components with the largest n $(n < x_0)$ eigenvalues are considered to reduce computational cost and noise. n determines the number of detectable IFCs and selected on the basis of the comparison of the first n eigenvalues and their associated eigenvectors with the noise components (see Fig. 2c,2d,3c,3d). Using $\mathbf{B}$, $\mathbf{W}$, and n, it can be shown that an element of the scaled third-order cumulant tensor $\boldsymbol{\mathcal{T}}'$ is approximated as a sum over n independent components (Computational Methods 2 in Supporting Information):

$$\mathcal{T}'_{ijk} \approx \sum_{s=1}^{n} \left(\sum_{p=1}^{n} B_{ip} \sqrt{W_p} U_{sp} \right) \left(\sum_{q=1}^{n} B_{jq} \sqrt{W_q} U_{sq} \right) \left(\sum_{r=1}^{n} B_{kr} \sqrt{W_r} U_{sr} \right) h_s, \tag{9}$$

where h_s is a coefficient representing the contribution of the s-th independent component, which contains information of the average number of molecules in the observation volume, and U_{sp}, U_{sq}, U_{sr} are elements of an $n \times n$ orthogonal matrix $\mathbf{U}$. Hence, the goal of IFCA is to determine $h_1, \cdots, h_n$ and $\mathbf{U}$ by comparing Eq. 9 and the experimentally obtained cumulant tensor. To carry out this analysis efficiently, the dimension of the third-order cumulant tensor is reduced by multiplying the eigenvectors:

$$\mathcal{T}''_{pqr} \equiv \sum_{i=1}^{x_0} \sum_{j=1}^{x_0} \sum_{k=1}^{x_0} B_{ip} B_{jq} B_{kr} \mathcal{T}'_{ijk} \tag{10}$$

$$\approx \sum_{s=1}^{n} \sqrt{W_p W_q W_r} U_{sp} U_{sq} U_{sr} h_s \,, \tag{11}$$

where $1 \leq p, q, r \leq n$. With this conversion, the original dimension of $\mathcal{T}'_{ijk}$, x_0^3 (e.g., $x_0 = 256$), is reduced to n^3, where n is typically smaller than 10. $h_1, \cdots, h_n$ and $\mathbf{U}$ are determined by fitting the experimentally obtained $\boldsymbol{\mathcal{T}}''$ (Eq. 10) with Eq. 11 by using $h_1, \cdots, h_n$ and $\mathbf{U}$ as fitting parameters (Figs. S2, S3 in Supporting Information). Fitting involving an orthogonal matrix as a fitting parameter can be performed by a Jacobi-type algorithm[28] (Computational Methods 3 in Supporting Information). Finally, the fluorescence patterns of IFCs are reconstructed from the obtained parameters (Computational Methods 2 in Supporting Information):

$$d_x^s \approx \sqrt{v_x} \left(\sum_{p=1}^{n} B_{xp} \sqrt{W_p} U_{sp} \right) h_s. \tag{12}$$

Here, the factor $\sqrt{v_x}$ is multiplied to compensate for the noise normalization (Eqs. 6 and 7). This d_x^s represents fluorescence pattern of a single molecule of component s, and its sum over x corresponds to molecular brightness. Note that the shape of d_x^s potentially depends on the choice of n.

The concentrations of fluorescent species representing IFCs are evaluated as the average numbers of molecules in the observation volume, N_s, which is obtained as (Computational Methods 2 in Supporting Information)

$$N_s \approx \frac{1}{h_s^2}. \tag{13}$$

Note that this is the average number of molecules in the observation volume having a well-defined boundary. In SMF measurements under confocal microscope such as fluorescence correlation spectroscopy (FCS), the observation volume is often approximated by a 3-D gaussian profile. In that case, the value obtained with Eq. 13 differs from that defined in FCS,

N_s^{FCS}, as:[19]

$$N_s^{\mathrm{FCS}} = \frac{64}{27} N_s. \tag{14}$$

Details of the IFCA analysis on actual photon data are described in Supporting Information, Computational Methods 4.

Application to a mixture of fluorescent dyes

As the first example, we applied IFCA to experimental photon data of a static mixture of five fluorescent dyes. A mixture solution of tetramethyl rhodamine (TMR), rhodamine 6G (R6G), Alexa Fluor 647 (Alexa647), ATTO647N, and Texas Red (TR) was prepared with a concentration of 1-10 nM for each dye (Experimental Section 1 in Supporting Information). These dyes have different absorption/emission wavelengths and/or fluorescence lifetimes.

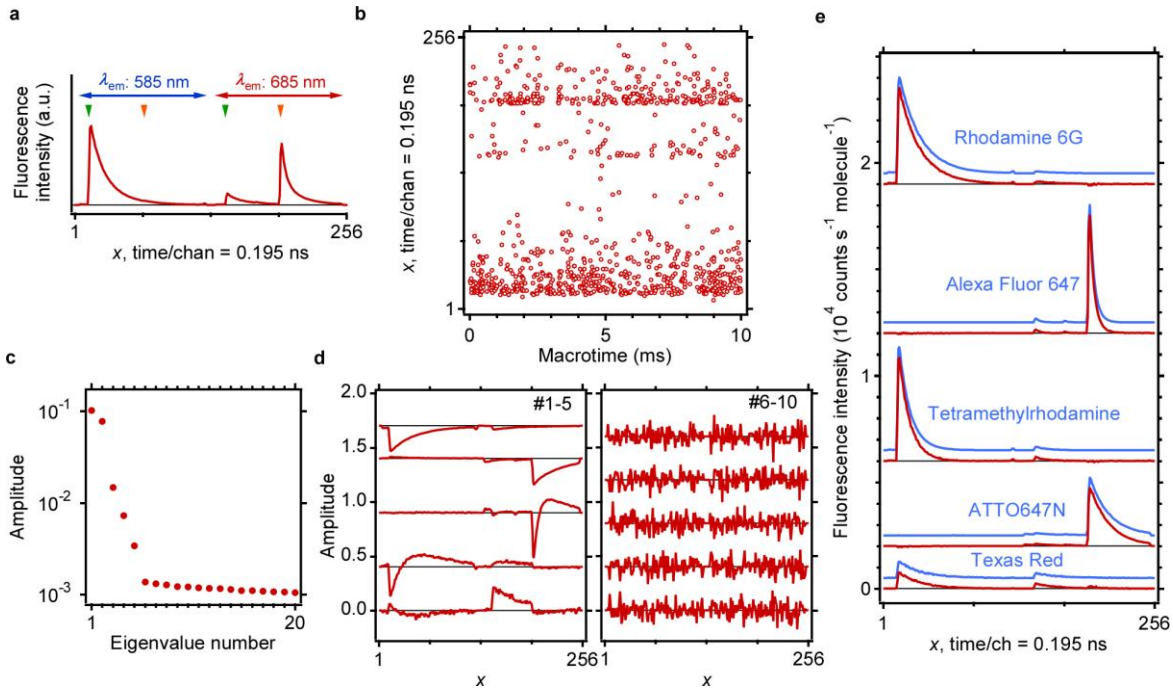

Figure 2. Application of IFCA to a dye mixture. (a) Ensemble-averaged fluorescence pattern of a mixture of five fluorescent dyes. The arrival times of two-color excitation pulses are indicated by downwards arrowheads (green: 525 nm, orange: 640 nm). (b) Representative portion of the photon data of the mixture sample. In this panel, 910 photons are plotted in total. (c) The largest 20 eigenvalues of the scaled second-order cumulant tensor. (d) The eigenvectors associated with the largest 10 eigenvalues (#1-5 (left) and #6-10 (right), from top to bottom). Traces are vertically offset for clarity. (e) Fluorescence patterns of five IFCs retrieved by the IFCA analysis (red). Traces are vertically offset for clarity. Fluorescence patterns of individual fluorophores are shown in blue, which are scaled and vertically offset for comparison.

Multiparameter fluorescence measurement was performed on this mixture solution with a confocal microscope equipped with pulsed-interleaved excitation (PIE)[7, 29] and TCSPC (Experimental Section 2 in Supporting Information).[30] Figure 2a shows the fluorescence pattern of the mixture solution integrated over the entire acquisition period (3313 sec.). This pattern is essentially the sum of contributions from individual dyes having specific fluorescence patterns. The multiparameter photon data obtained in this experiment (Fig. 2b) exhibits two notable features. First, photons seem to be detected mostly at random and distributed uniformly. In other words, no time regions are clearly discernible that exhibit locally dense photons. The lack of such 'burst-like' features is due to the relatively high dye concentrations (nM), making the observation volume almost always occupied with at least one molecule. Such a high concentration regime is usually not compatible with ordinary single-molecule experiments, where separation of individual molecules is of paramount importance. Secondly, the average interval of photon arrival times is about 10 μs. If we performed a typical smFRET experiment with such a count rate, it would take ~1 ms to collect sufficient number of photons to derive a reliable FRET efficiency. Therefore, even if a lower concentration were employed in this experiment, it would be impossible to probe transient states with a microsecond lifetime.

Figures 2c,d show the results of EVD analysis (Eq. 8) of the scaled second-order cumulant tensor made from the recorded photon data employing the temporal binning width of 5 μs. In the eigenvalue plot (Fig. 2c), it is clearly seen that the first five largest eigenvalues are beyond the noise level. Correspondingly, the associated eigenvectors of these eigenvalues (Fig. 2d, #1-5) display meaningful features, while other eigenvectors (#6-10) are dominated by noise. Hence, we selected the first five eigenvectors for reducing the third-order cumulant tensor (Eq. 10), and then the reduced tensor was fitted using Eq. 11 to obtain $h_1, \cdots, h_5$ and $\mathbf{U}$. The fitting results satisfactorily match with the elements of the reduced tensor (Fig. S2 in Supporting Information), justifying the use of fitting algorithm employed in this work.

The five IFCs recovered by Eq. 12 are compared with reference patterns determined by individually measuring solutions containing only one of the five dyes (Table 1). In this table, the cosine similarity between a recovered IFC, s, and a reference pattern, r, are evaluated as

$$S_{\mathrm{C}}(s,r) = \frac{\sum_{x=1}^{x_0} d_x^s d_x^r}{\sqrt{\sum_{x=1}^{x_0} (d_x^s)^2 \sum_{x=1}^{x_0} (d_x^r)^2}}. \tag{15}$$

For each of the recovered IFCs, a reference pattern exhibiting a high similarity ($S_c > 0.99$) is found. According to Table 1, we made assignments of IFCs #1, #2, #3, #4, and #5 as TR, ATTO647N, TMR, Alexa647, and R6G, respectively. The fluorescence patterns of these IFCs and assigned reference patterns (Fig. 2e) are visually almost indistinguishable, confirming the successful decomposition of IFCs using IFCA. The average numbers of molecules in the observation volume derived by Eq. 13 are $N = 0.343$ (TR), 0.095 (ATTO647N), 0.089 (TMR), 0.088 (Alexa647), and 0.070 (R6G). The relative concentrations estimated from these values are consistent with the actual concentration (5 nM for TR and 1 nM for ATTO647N, TMR, Alexa647, R6G) within experimental error in sample preparation. Furthermore, the fluorescence patterns of IFCs and their relative concentrations are reproducible for three measurements (Fig. S1f,g in Supporting Information).

Table 1. Cosine similarities between the IFCs retrieved by IFCA and reference fluorescence patterns of the five fluorophores. Values in parentheses are the standard deviations (multiplied by 10^4) estimated from three measurements.

IFC #	TR	ATTO647N	TMR	Alexa647	R6G
1	0.9947(10)	0.1021(93)	0.9066(46)	0.0951(108)	0.9572(20)
2	0.0373(19)	0.9998(1)	-0.0084(14)	0.7889(94)	-0.0076(17)
3	0.9067(24)	-0.0035(26)	0.9997(1)	-0.0041(43)	0.9429(21)
4	0.0290(9)	0.7821(72)	-0.0060(2)	0.9996(1)	-0.0061(4)
5	0.9647(6)	-0.0078(75)	0.9540(20)	-0.0094(109)	0.9989(9)

We have also examined the dependence of IFCs on the number of components (n) selected for reducing the third-order cumulant tensor (Eq. 10). The results for setting n = 4 to 6 are presented in Fig. S1a-c in Supporting Information. When n is decremented (n = 4), the IFC assignable to TR disappears, while when n is incremented (n = 6), an unphysical noisy component appears. These are naturally expected behaviors and regarded as representing the practicality of IFCA. On the other hand, the obtained fluorescence patterns are essentially invariable on extending the binning width T_{bin} up to 50 μs (Fig. S1d, e in Supporting Information), which implies that the result for T_{bin} = 5 μs is reliable, as this measurement was performed on a static mixture sample, and hence, there should be no binning-width dependence. At the same time, an increase in the N values and concomitant decrease of the amplitudes of fluorescence intensity, which corresponds to the molecular brightness, are observed on increasing T_{bin}. This behavior is ascribed to the effect of molecular diffusion across the observation volume.[31] Notably, the binning width employed in this analysis, T_{bin} = 5 μs, is even shorter than the average photon interval, ~10 μs (Fig. 2b), which is in a stark contrast with conventional methods of smFRET analysis.

Application to a FRET-labeled DNA construct

In the second example, we focused on application of IFCA to a molecular system undergoing dynamic transitions between conformational states distinguished by their FRET efficiencies. For this purpose, we analyzed photon data sets that were obtained by Schröder et al.[22] and deposited on a public repository.[32] The data contain photons recorded with two-color TCSPC-FCS measurements on a single-stranded DNA, which was labeled with Cy3B as a FRET donor and tethered on a small DNA origami structure as illustrated in Fig. 3a. In addition to this donor-labeled single-stranded DNA (*pointer* strand), the DNA origami structure was attached with two short single-stranded DNA (*staple* strands) having identical 5-nt sequences complementary to the end of the pointer strand. With this design, the pointer strand may form a duplex with

either of the two staple strands with almost equal probability. The FRET acceptor (ATTO647N) was fixed on the DNA origami structure near one of these two staple strands, so that switching between the two states could be detected in an equilibrium condition through the change of FRET efficiency. The relaxation time constant between these two states, high-FRET (HF) and low-FRET (LF), was reported to be $\tau = (1/\tau_{HF\rightarrow LF} + 1/\tau_{LF\rightarrow HF})^{-1} = 220 \pm 20$ μs.[22] The entire DNA construct was freely diffusing in a buffer with a concentration of 100 pM.

Figure 3b shows a concatenated fluorescence pattern of the donor fluorescence (x = 1-128) and acceptor fluorescence (x = 129-256) integrated over 600 s. The decay time of the donor fluorescence and the ratio of integrated intensities of the donor (I_D) and acceptor (I_A) fluorescence reflect the FRET efficiency of the labeled DNA. This ensemble pattern is the superposition of the HF and LF patterns, which is characterized by faster donor decay with

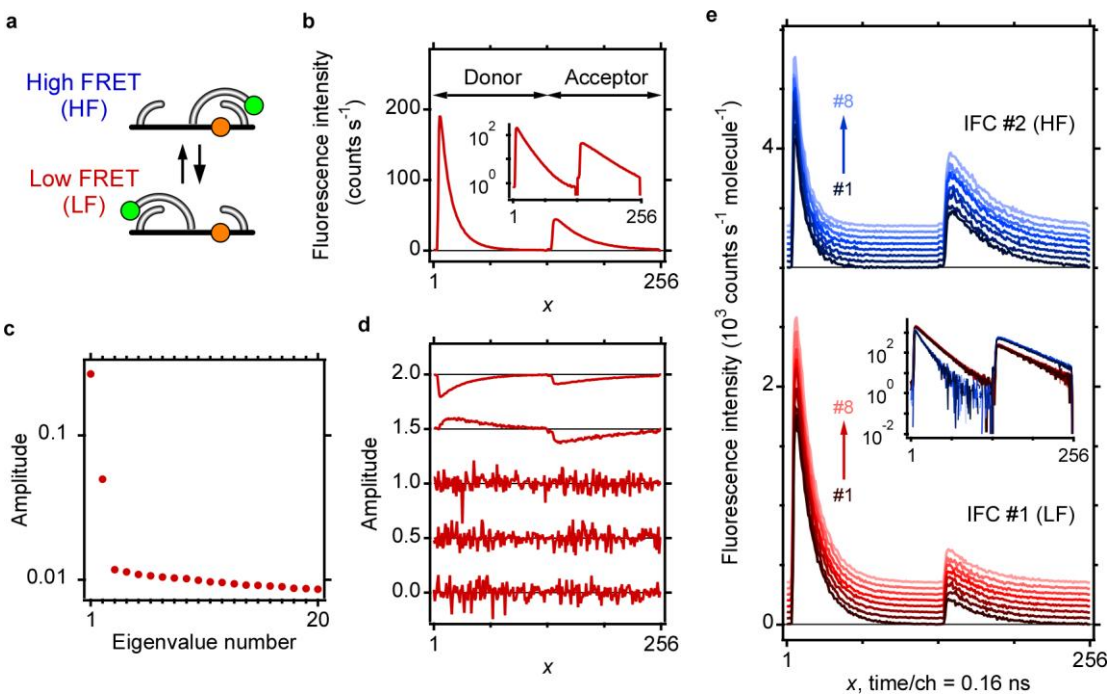

Figure 3. Applications of IFCA to a FRET-labeled DNA construct. (a) Schematic illustration of the FRET-labeled DNA construct. Green and orange circles: donor and acceptor dyes, black bars: DNA origami structure. (b) Ensemble-averaged fluorescence pattern. Inset: The same pattern in log scale. (c) The largest 20 eigenvalues of the scaled second-order cumulant tensor. (d) Eigenvectors associated with the largest five eigenvalues (from top to bottom). Traces are vertically offset for clarity. (e) Fluorescence patterns of IFCs retrieved by the IFCA analysis. Two IFCs obtained for the eight data sets are plotted separately with vertical offsets for clarity. Inset: All IFCs plotted together in log scale.

higher acceptor-to-donor ratio, and slower donor decay with lower acceptor-to-donor ratio, respectively. To separate these contributions, we applied IFCA to the raw photon data by employing a binning width T_{bin} = 20 μs that is much shorter than the relaxation time scale.

Being consistent with the two-state model in Fig. 3a, two significant eigen components were identified in an EVD analysis (Figs. 3c,d). It indicates that a coexisting impurities such as donor-only population is negligible, in agreement with the result of a strict single molecule analysis by Schröder et al.[22] Therefore, we set n = 2 and performed the dimensionality reduction and fitting of the third-order cumulant tensor using Eqs. 10 and 11 (Fig. S3 in Supporting Information). The obtained fluorescence patterns (Fig. 3e) are undoubtedly assignable to the LF (IFC #1) and HF (IFC #2) states due to rapid donor decay and relatively high acceptor intensity of IFC #2. This result is highly reproducible over eight data sets, allowing us to precisely determine the mean donor fluorescence lifetimes (τ_{donor}) and the apparent FRET efficiencies ($E = I_{\mathrm{A}}/(I_{\mathrm{D}} + I_{\mathrm{A}})$). These values were determined as $\tau^{1}_{\mathrm{donor}} = 2.50\ (0.02)$ ns, $\tau^{2}_{\mathrm{donor}} = 1.45\ (0.03)$ ns and $E^{1} = 0.179\ (0.005)$, $E^{2} = 0.594\ (0.011)$ for the two IFCs, where values in parentheses are standard deviations. The concentrations were obtained for these IFCs as N_1 = 0.0238 (0.0016) and N_2 = 0.0205 (0.0015), leading to the equilibrium constant K = [HF]/[LF] = N_2/N_1 = 0.87 (0.08). Deviation of K from unity is consistent with the previous work by Schröder et al.,[22] who suggested that LF is slightly more stable in similar DNA constructs.

In the present analysis, the binning width of 20 μs allows us to distinguish two forms interconverting on the sub-millisecond time scale. Assuming an exponential dynamics with the relaxation time constant of τ = 220 μs, the effect of interconversion in the retrieved IFCs with T_{bin} = 20 μs is estimated to be ~4%. Thus, this analysis highlights the feasibility of IFCA interrogating dynamic FRET systems with a microsecond time resolution. The results are generally consistent with the previous work by Schröder et al.[22] Moreover, IFCA provides

additional sources of information such as number of IFCs, acceptor intensity, and the precise shapes of decay curves. Interestingly, the donor decays are not exactly single exponential but double exponential functions are required to reproduce them (Fig. S4 in Supporting Information). Such a detailed analysis is made feasible by the model-free nature of IFCA.

Discussion

The two applications explored in this work represent the high potential of IFCA in resolving mixed fluorescence signals: The first example of the dye mixture illustrates accurate and robust separation of five IFCs and determination of their characteristic fluorescence patterns and relative concentrations on a nanomolar range. The second application to the FRET-labeled DNA construct demonstrates the high time resolution of IFCA to precisely identify microsecond transients and its model-free nature, enabling detailed analysis of the obtained fluorescence patterns. The high time resolution of IFCA and its compatibility with relatively high sample concentrations originate from its basis on a correlation analysis. It is analogous to FCS, in which a nanosecond time resolution can be achieved for observing molecular dynamics.[33] IFCA collects photon triplets emitted from individual molecules and analyzes their statistical properties (Fig. 1b). Because only three photons are required per molecule—far fewer than ~100 photons typically needed for smFRET—short binning times down to a few microseconds can be employed. On the other hand, the advantage at high-concentrations stems from the ability of IFCA to accurately eliminate cross-molecular contributions caused by multiple molecules coexisting within the observation volume. This is accomplished through the construction of cumulant tensors (Eqs. 3 and 5), which automatically subtract uncorrelated backgrounds including cross-molecular effects.[27] Such automatic subtraction is also convenient for removing background signals originating from water Raman scattering, autofluorescence of optical elements, and detector dark counts, without resorting to additional control measurements (compare insets of Fig. 3b and 3e).

Existing methods for multiparameter SMF analysis of freely diffusing molecules can be categorized into strict single-molecule methods and correlation-based approaches. The former, such as H^2MM[10-12] and FRET lines,[13, 14] require low concentrations (~10 pM) and high molecular brightness (~100 photons per bin/burst). Moreover, when applied to study microsecond transients, it is necessary to assume appropriate physical models, such as the relation between the donor/acceptor ratio and fluorescence lifetime[15] or Markovian property of state transitions. In contrast, IFCA is entirely model-independent except for the assumption of statistical independence of individual components and is compatible with nanomolar concentrations and low brightness (three photons per bin). A representative correlation-based method is two-dimensional fluorescence lifetime correlation spectroscopy (2D FLCS).[34, 35] Since 2D FLCS is based on FCS, it shares the advantages of IFCA in sample concentration (nM) and time resolution (microseconds).[36] However, 2D FLCS solely relies on differences in fluorescence lifetimes for component separation. Hence, it is not straightforward to fully utilize information obtainable with multiparameter measurements, though a few techniques have been developed that incorporate additional detection channels.[30, 37] Therefore, in terms of component separation, IFCA can be regarded as a generalization of 2D FLCS to model-free analysis via employing third-order cumulant tensors. Other correlation-based approaches utilizing higher-order cumulants have been reported by Müller et al.;[19, 31, 38] however, they were restricted to photon data of up to two detection channels. In contrast, the diagonalization of third-order cumulant tensors introduced in this work enables the separation of general multiparameter signals.

Although IFCA is highly effective for separating multiple components, there are a few specific situations warranting caution. First, fluorescence patterns of two or more components are not always linearly independent. In such cases, the number of mutually orthogonal eigenvectors obtained with EVD of the scaled second-order cumulant tensor (Eq. 8) does not

correspond to the true number of components, thereby precluding the application of ICA. Examples include oligomer formation accompanied by changes in brightness without altering fluorescence patterns, and FRET measurements with coexisting donor-only, acceptor-only, and zero-FRET species.[30] Dedicated analytical protocols are required to be established for these cases. Second, when rapid interconversion occurs between components, it would be difficult to employ a sufficiently short binning width (e.g., less than 10% of the time scale of the dynamics) to minimize the dynamical effect. If the binning width approaches the time scale of dynamics, IFCA yields mixed fluorescence patterns rather than distinct components. Consequently, investigations of fast dynamics require methods to estimate mixing effects accurately, e.g., through systematic analysis of binning-width dependence. Finally, when the number of components n is large, it would become challenging to achieve a sufficiently low noise level to extract n significant components via EVD (Eq. 8). In such cases, alternative approaches that do not rely on dimensionality reduction are necessary.

The advantages of IFCA demonstrated in this work are expected to enable a broad range of novel measurement modalities. Its high time resolution can be exploited not only for characterizing short-lived transient species but also for measurements under continuous flow in microfluidic systems, where the observation window is constrained by the brief passage time through the detection volume.[39, 40] Compatibility with high sample concentrations facilitates efficient data acquisition and allows detection of complex formation with moderate (~nM) affinity via intermolecular FRET.[41, 42] Most importantly, IFCA is entirely independent of any specific theoretical model for reproducing fluorescence patterns. As a result, it can be applied to arbitrary combinations of fluorescence parameters, i.e., excitation/emission wavelengths, emission delay time, and polarization. The model-free nature enables direct integration with advanced multiparameter SMF techniques such as polarization-resolved measurement[8] and multi-color FRET.[43] Beyond single-molecule studies, the same framework is potentially

applicable to multiparameter single-particle[44] or single-cell spectroscopic measurements such as spectral flow cytometry.[45]

Conclusion

In this work, we applied ICA to multiparameter SMF photon data from freely diffusing molecules and introduced IFCA as a novel analytical framework, based on the decomposition of third-order cumulant tensors generated from photon data. To assess its performance, we analyzed two representative experimental datasets: a static dye mixture and a dynamic FRET-labeled DNA construct. The results demonstrate that IFCA enables robust separation of fluorescence signals from multiple species in samples with concentrations up to the nanomolar range, achieving time resolutions on the order of several microseconds, all within a fully model-free approach. These findings establish IFCA as a versatile and powerful method that holds significant promise for advancing the scope and precision of SMF experiments, paving the way for the quantitative analysis of complex molecular systems.

Supporting Information

The authors have cited additional references within the Supporting Information.[46-48]

Data availability. Data for the dye mixture experiment have been deposited at Zenodo and are publicly available at https://doi.org/10.5281/zenodo.14869092. Data for the FRET-labeled single-stranded DNA construct are made publicly available by Schröder et al. and accessible at https://doi.org/10.5281/zenodo.7254095.

Code availability. A Python module for IFCA analysis has been deposited at Zenodo and is publicly available at https://doi.org/10.5281/zenodo.14869092.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number JP23K21102 (to K. I.).

References

(1) Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S. Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **2018**, *359*, eaan1133. DOI: 10.1126/science.aan1133.
(2) Nettels, D.; Galvanetto, N.; Ivanović , M. T.; Nüesch, M.; Yang, T.; Schuler, B. Single-molecule FRET for probing nanoscale biomolecular dynamics. *Nat. Rev. Phys.* **2024**, *6*, 587-605.
(3) Kapanidis, A. N.; Lee, N. K.; Laurence, T. A.; Doose, S.; Margeat, E.; Weiss, S. Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (24), 8936-8941. DOI: 10.1073/pnas.0401690101.
(4) Sisamakis, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A. M. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* **2010**, *475*, 455-514. DOI: 10.1016/s0076-6879(10)75018-7.
(5) Eggeling, C.; Berger, S.; Brand, L.; Fries, J. R.; Schaffer, J.; Volkmer, A.; Seidel, C. A. M. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **2001**, *86* (3), 163-180.
(6) Widengren, J.; Kudryavtsev, V.; Antonik, M.; Berger, S.; Gerken, M.; Seidel, C. A. M. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Anal. Chem.* **2006**, *78*, 2039-2050. DOI: 10.1021/ac0522759.
(7) Hendrix, J.; Lamb, D. C. Pulsed interleaved excitation: Principles and applications. *Methods Enzymol.* **2013**, *518*, 205-243.
(8) Kolimi, N.; Ghimire, S.; Duffy, F.; Peulen, T.; Medina, E.; Sanabria, H. Time-resolved fluorescence anisotropy from single molecules for characterizing local flexibility in biomolecules. *J. Vis. Exp.* **2025**, *218*, e67802.
(9) Chu, J.; Ejaz, A.; Lin, K. M.; Joseph, M. R.; Coraor, A. E.; Drummond, D. A.; Squires, A. H. Single-molecule fluorescence multiplexing by multi-parameter spectroscopic detection of nanostructured FRET labels. *Nat. Nanotechnol.* **2024**, *19*, 1150-1157.
(10) Pirchi, M.; Tsukanov, R.; Khamis, R.; Tomov, T. E.; Berger, Y.; Khara, D. C.; Volkov, H.; Haran, G.; Nir, E. Photon-by-photon hidden Markov model analysis for microsecond single-molecule FRET kinetics. *J. Phys. Chem. B* **2016**, *120* (51), 13065-13075. DOI: 10.1021/acs.jpcb.6b10726.
(11) Harris, P. D.; Narducci, A.; Gebhardt, C.; Cordes, T.; Weiss, S.; Lerner, E. Multi-parameter photon-by-photon hidden Markov modeling. *Nat. Commun.* **2022**, *13*, 1000. DOI: 10.1038/s41467-022-28632-x.
(12) Harris, P. D.; Lerner, E. Identification and quantification of within-burst dynamics in singly labeled single-molecule fluorescence lifetime experiments. *Biophys. Rep.* **2022**, *2*, 100071. DOI: 10.1016/j.bpr.2022.100071.
(13) Barth, A.; Opanasyuk, O.; Peulen, T.-O.; Felekyan, S.; Kalinin, S.; Sanabria, H.; Seidel, C. A. M. Unraveling multi-state molecular dynamics in single-molecule FRET experiments. I. Theory of FRET-lines. *J. Chem. Phys.* **2022**, *156*, 141501. DOI: 10.1063/5.0089134.
(14) Opanasyuk, O.; Barth, A.; Peulen, T.-O.; Felekyan, S.; Kalinin, S.; Sanabria, H.; Seidel, C. A. M. Unraveling multi-state molecular dynamics in single-molecule FRET experiments. II. Quantitative analysis of multi-state kinetic networks. *J. Chem. Phys.* **2022**, *157*, 031501. DOI: 10.1063/5.0095754.
(15) Gopich, I. V.; Szabo, A. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 7747-7752. DOI: 10.1073/pnas.1205120109.
(16) Gopich, I. V. Concentration effects in "single-molecule" spectroscopy. *J. Phys. Chem. B*

2008, *112*, 6214-6220.
(17) Nir, E.; Michalet, X.; Hamadani, K. M.; Laurence, T. A.; Neuhauser, D.; Kovchegov, Y.; Weiss, S. Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J. Phys. Chem. B* **2006**, *110*, 22103-22124.
(18) Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; Wiley-Interscience, 2001.
(19) Müller, J. D. Cumulant analysis in fluorescence fluctuation spectroscopy. *Biophys. J.* **2004**, *86*, 3981-3992. DOI: 10.1529/biophysj.103.037887.
(20) Li, J.; Usevich, K.; Comon, P. Globally convergent Jacobi-type algorithms for simultaneous orthogonal symmetric tensor diagonalization. *SIAM J. Matrix Anal. Appl.* **2018**, *39* (1), 1-22. DOI: 10.1137/17M1116295.
(21) Song, L.; Lu, H. EcoICA: Skewness-based ICA via eigenvectors of cumulant operator. *Proceedings of Machine Learning Research* **2016**, *63*, 445-460.
(22) Schröder, T.; Bohlen, J.; Ochmann, S. E.; Schüler, P.; Krause, S.; Lamb, D. C.; Tinnefeld, P. Shrinking gate fluorescence correlation spectroscopy yields equilibrium constants and separates photophysics from structural dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (4), e2211896120. DOI: 10.1073/pnas.2211896120.
(23) Becker, W. *The bh TCSPC Handbook*; Becker & Hickl, 2026.
(24) Kruskal, J. B. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Its Appl.* **1977**, *18*, 95-138. DOI: 10.1016/0024-3795(77)90069-6.
(25) Mandel, L. Fluctuations of photon beams and their correlations. *Proc. Phys. Soc.* **1958**, *72* (6), 1037-1048. DOI: 10.1088/0370-1328/72/6/312.
(26) Abdollah-Nia, F. Cumulant-based formulation of higher-order fluorescence correlation spectroscopy. *arXiv* **2018**, 1612.05388v05382.
(27) Ishii, K.; Tahara, T. Extracting decay curves of the correlated fluorescence photons measured in fluorescence correlation spectroscopy. *Chem. Phys. Lett.* **2012**, *519-520*, 130-133.
(28) De Lathauwer, L. Algebraic methods after prewhitening. In *Handbook of Blind Source Separation*, Comon, P., Jutten, C. Eds.; Academic Press, 2010; pp 155-177.
(29) Müller, B. K.; Zaychikov, E.; Bräuchle, C.; Lamb, D. C. Pulsed interleaved excitation. *Biophys. J.* **2005**, *89*, 3508-3522. DOI: 10.1529/biophysj.105.064766.
(30) Sarkar, B.; Ishii, K.; Tahara, T. Pulsed-interleaved-excitation two-dimensional fluorescence lifetime correlation spectroscopy. *J. Phys. Chem. B* **2024**, *128* (19), 4685-4695.
(31) Wu, B.; Müller, J. D. Time-integrated fluorescence cumulant analysis in fluorescence fluctuation spectroscopy. *Biophys. J.* **2005**, *89*, 2721-2735.
(32) Schröder, T.; Bohlen, J.; Ochmann, S. E.; Schüler, P.; Krause, S.; Lamb, D. C.; Tinnefeld, P. Data for "Shrinking gate fluorescence correlation spectroscopy yields equilibrium constants and separates photophysics from structural dynamics". Zenodo, **2022**. DOI: 10.5281/zenodo.7254095.
(33) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2655-2660. DOI: 10.1073/pnas.0611093104.
(34) Ishii, K.; Tahara, T. Two-dimensional fluorescence lifetime correlation spectroscopy. 1. Principle. *J. Phys. Chem. B* **2013**, *117*, 11414-11422. DOI: 10.1021/jp406861u.
(35) Ishii, K.; Tahara, T. Two-dimensional fluorescence lifetime correlation spectroscopy. 2. Application. *J. Phys. Chem. B* **2013**, *117*, 11423-11432. DOI: 10.1021/jp406864e.
(36) Otosu, T.; Ishii, K.; Tahara, T. Microsecond protein dynamics observed at the single-molecule level. *Nat. Commun.* **2015**, *6*, 7685. DOI: 10.1038/ncomms8685.
(37) Cheng, C.-H.; Ishii, K.; Tahara, T. Microsecond equilibrium dynamics of hairpin-forming

oligonucleotides quantified by two-color two-dimensional fluorescence lifetime correlation spectroscopy. *J. Phys. Chem. B* **2020**, *124*, 10673-10681.
(38) Wu, B.; Chen, Y.; Müller, J. D. Dual-color time-integrated fluorescence cumulant analysis. *Biophys. J.* **2006**, *91*, 2687-2698.
(39) Lipman, E. A.; Schuler, B.; Bakajin, O.; Eaton, W. A. Single-molecule measurement of protein folding kinetics. *Science* **2003**, *301*, 1233-1235. DOI: 10.1126/science.1085399.
(40) Gambin, Y.; VanDelinder, V.; Ferreon, A. C. M.; Lemke, E. A.; Groisman, A.; Deniz, A. A. Visualizing a one-way protein encounter complex by ultrafast single-molecule mixing. *Nat. Methods* **2011**, *8* (3), 239-241. DOI: 10.1038/nmeth.1568.
(41) Zijlstra, N.; Dingfelder, F.; Wunderlich, B.; Zosel, F.; Benke, S.; Nettels, D.; Schuler, B. Rapid microfluidic dilution for single-molecule spectroscopy of low-affinity biomolecular complexes. *Angew. Chem. Int. Ed.* **2017**, *56*, 7126-7129. DOI: 10.1002/anie.201702439.
(42) White, D. S.; Smith, M. A.; Chanda, B.; Goldsmith, R. H. Strategies for overcoming the single-molecule concentration barrier. *ACS Meas. Sci. Au* **2023**, *3* (4), 239-257. DOI: 10.1021/acsmeasuresciau.3c00002.
(43) Yoo, J.; Louis, J. M.; Gopich, I. V.; Chung, H. S. Three-color single-molecule FRET and fluorescence lifetime analysis of fast protein folding. *J. Phys. Chem. B* **2018**, *122* (49), 11702-11720. DOI: 10.1021/acs.jpcb.8b07768.
(44) Sych, T.; Schlegel, J.; Barriga, H. M. G.; Ojansivu, M.; Hanke, L.; Weber, F.; Bostancioglu, R. B.; Ezzat, K.; Stangl, H.; Plochberger, B.; et al. High-throughput measurement of the content and properties of nano-sized bioparticles with single-particle profiler. *Nat. Biotechnol.* **2024**, *42*, 587-590.
(45) Nolan, J. P.; Condello, D. Spectral flow cytometry. *Curr. Protoc. Cytom.* **2013**, *63*, 1.27.21-21.27.13.
(46) Heo, W.; Hasegawa, K.; Okamoto, K.; Sako, Y.; Ishii, K.; Tahara, T. Scanning two-dimensional fluorescence lifetime correlation spectroscopy: conformational dynamics of DNA Holliday junction from microsecond to subsecond. *J. Phys. Chem. Lett.* **2022**, *13*, 1249-1257.
(47) Abdollah-Nia, F.; Gelfand, M. P.; Van Orden, A. Artifact-free and detection-profile-independent higher-order fluorescence correlation spectroscopy for microsecond-resolved kinetics. 1. Multidetector and sub-binning approach. *J. Phys. Chem. B* **2017**, *121* (11), 2373-2387.
(48) Ingargiola, A.; Laurence, T.; Boutelle, R.; Weiss, S.; Michalet, X. Photon-HDF5: an open file format for timestamp-based single-molecule fluorescence experiments. *Biophys. J.* **2016**, *110* (1), 26-33.