Independent component analysis disentangles fluorescence signals from diffusing single molecules

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

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

Abstract

Multiparameter single-molecule fluorescence measurement is a promising technique for detecting and quantifying heterogeneity in freely diffusing molecules, albeit with challenges in data analysis due to limited photon numbers. Here, we present a universal analytical framework that leverages independent component analysis for unmixing multiparameter fluorescence signals. Applications to static and dynamic mixture systems demonstrate its potential allowing model-free separation of subpopulations with microsecond time resolution in nanomolar concentration regime.

Main text

Single-molecule fluorescence (SMF) methods, in particular single-molecule Förster resonance energy transfer (smFRET), are invaluable tools in current biophysical research for detecting the heterogeneity and fluctuations of molecular structures, which are inaccessible with ensemble measurements. Nowadays, they are utilized for elucidating a variety of subjects[1]. 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 (Fig. 1a)[2, 3]. Photon data obtained with such multiparameter experiments are analyzed with the help of dedicated methods designed for individual experimental configurations[4, 5]. However, there has been no general approach of photon data analysis that is compatible with arbitrary combination of fluorescence parameters. 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, limiting their applicability in terms of the sample concentration (~10 pM), the time resolution (>~1 ms), and the brightness of fluorophores.

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 utilizing its non-gaussian property. It has been widely applied in other fields, e.g., feature extraction, brain imaging, and telecommunications[6]. Because the fluorescence signals from freely diffusing single molecules recorded with a confocal microscope exhibit Poisson-like non-gaussian behavior[7], ICA is

expected to be applicable also to SMF data analysis. Here, we implemented an ICA for SMF, which is called IFCA (independent fluorescence component analysis) hereafter, by using the third-order cumulant tensor[8] of temporal fluctuations in multiparameter fluorescence signals, $\boldsymbol{\mathcal{T}}$ (Fig.1b; Eq. (1) in *Online methods*). Cumulants of fluorescence fluctuations are known to be represented as the sum of those of individual species contained in the sample[7]. Therefore, unmixing of an observed signal is achieved by decomposition of the cumulant tensors into rank-1 tensors of independent fluorescence components (IFCs; Fig. 1c) as $\mathcal{T}_{ijk} = \sum_{\mathrm{IFC}} d_i^{\mathrm{IFC}} d_j^{\mathrm{IFC}} d_k^{\mathrm{IFC}} N_{\mathrm{IFC}}$, where N_{IFC} is the number of molecules in the observation volume of an IFC and d_x^{IFC} is its signal intensity at channel x. The key feature of ICA utilized in this work is that, while the decomposition into rank-1 tensors is generally not unique for second-order matrices, third-order tensors can be decomposed into unique sets of rank-1 tensors[9].

Application of IFCA to experimental photon data is demonstrated in Fig. 2 and Extended Data Fig 1-3. In the first example, we prepared a mixture of five fluorophores with different absorption/emission wavelengths and/or fluorescence lifetimes with 1-10 nM concentration for each fluorophore (Fig. 2a). The mixture solution was measured with a confocal microscope equipped with pulsed-interleaved excitation (PIE)[10] and time-correlated single-photon counting (TCSPC)[11]. The five fluorescence patterns recovered by IFCA (Fig. 2b) are in good agreement with those individually obtained for solutions containing only one of the five fluorophores. This experiment demonstrates that IFCA can identify five IFCs and determine their characteristic fluorescence patterns and concentrations. Notably, the temporal binning width of photon data employed in this analysis is 5 μs, which is even shorter than the average photon interval over the whole photon data (12 μs, Fig. 2c). In the second example (Fig. 2d), we analyzed photon data sets that were obtained by Schröder et al.[12] and deposited on a public repository[13]. They contain photons recorded with two-color TCSPC measurement on FRET-labeled single-stranded DNA immobilized on freely diffusing DNA origami structures in

100 pM concentration. This single-stranded DNA system takes two metastable forms with different donor-acceptor distances (Fig. 2d, inset), and switching occurs between them in the sub-millisecond time scale[12]. The results of IFCA clearly resolve two IFCs assignable to the high-FRET and the low-FRET forms with high reproducibility over eight data sets. The binning width of 20 μs employed in this analysis allows us to distinguish these forms interconverting in the sub-millisecond time scale, highlighting the feasibility of IFCA for interrogating dynamic FRET systems with a microsecond time resolution.

Thus, IFCA enables us to separate fluorescence signals from different species in a sample with a high concentration up to the nanomolar range with a high time resolution down to several microseconds. The high time resolution of IFCA will facilitate characterization of a short-lifetime transient species or a measurement under continuous flow using microfluidics[14]. High sample concentration allows efficient data collection as well as detection of complex formation using intermolecular FRET[15]. Most importantly, IFCA is entirely free from a specific theoretical model for reproducing the fluorescence pattern. Therefore, it can be applied to arbitrary combinations of fluorescence parameters, i.e., excitation/emission wavelengths, emission delay time, and polarization. The model-free nature of IFCA enables its direct application to state-of-the-art SMF methods such as multiparameter fluorescence detection[3] and multi-color FRET[16]. Thus, IFCA can be a versatile method that will find many applications in SMF experiments.

References

1. Lerner, E. et al. Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **359**, eaan1133 (2018).
2. Kapanidis, A.N. et al. Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8936-8941 (2004).
3. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P.J. & Seidel, C.A.M. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.*

475, 455-514 (2010).
4. Harris, P.D. et al. Multi-parameter photon-by-photon hidden Markov modeling. *Nat. Commun.* **13**, 1000 (2022).
5. Barth, A. et al. Unraveling multi-state molecular dynamics in single-molecule FRET experiments. I. Theory of FRET-lines. *J. Chem. Phys.* **156**, 141501 (2022).
6. Hyvärinen, A., Karhunen, J. & Oja, E. Independent Component Analysis. (Wiley-Interscience, 2001).
7. Müller, J.D. Cumulant analysis in fluorescence fluctuation spectroscopy. *Biophys. J.* **86**, 3981-3992 (2004).
8. Li, J., Usevich, K. & Comon, P. Globally convergent Jacobi-type algorithms for simultaneous orthogonal symmetric tensor diagonalization. *SIAM J. Matrix Anal. Appl.* **39**, 1-22 (2018).
9. Kruskal, J.B. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Its Appl.* **18**, 95-138 (1977).
10. Müller, B.K., Zaychikov, E., Bräuchle, C. & Lamb, D.C. Pulsed interleaved excitation. *Biophys. J.* **89**, 3508-3522 (2005).
11. Sarkar, B., Ishii, K. & Tahara, T. Pulsed-interleaved-excitation two-dimensional fluorescence lifetime correlation spectroscopy. *J. Phys. Chem. B* **128**, 4685-4695 (2024).
12. Schröder, T. et al. Shrinking gate fluorescence correlation spectroscopy yields equilibrium constants and separates photophysics from structural dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2211896120 (2023).
13. Schröder, T. et al. Data for "Shrinking gate fluorescence correlation spectroscopy yields equilibrium constants and separates photophysics from structural dynamics", Zenodo, https://doi.org/10.5281/zenodo.7254095 (2022).
14. Lipman, E.A., Schuler, B., Bakajin, O. & Eaton, W.A. Single-molecule measurement of protein folding kinetics. *Science* **301**, 1233-1235 (2003).
15. Zijlstra, N. et al. Rapid microfluidic dilution for single-molecule spectroscopy of low-affinity biomolecular complexes. *Angew. Chem. Int. Ed.* **56**, 7126-7129 (2017).
16. 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* **122**, 11702-11720 (2018).

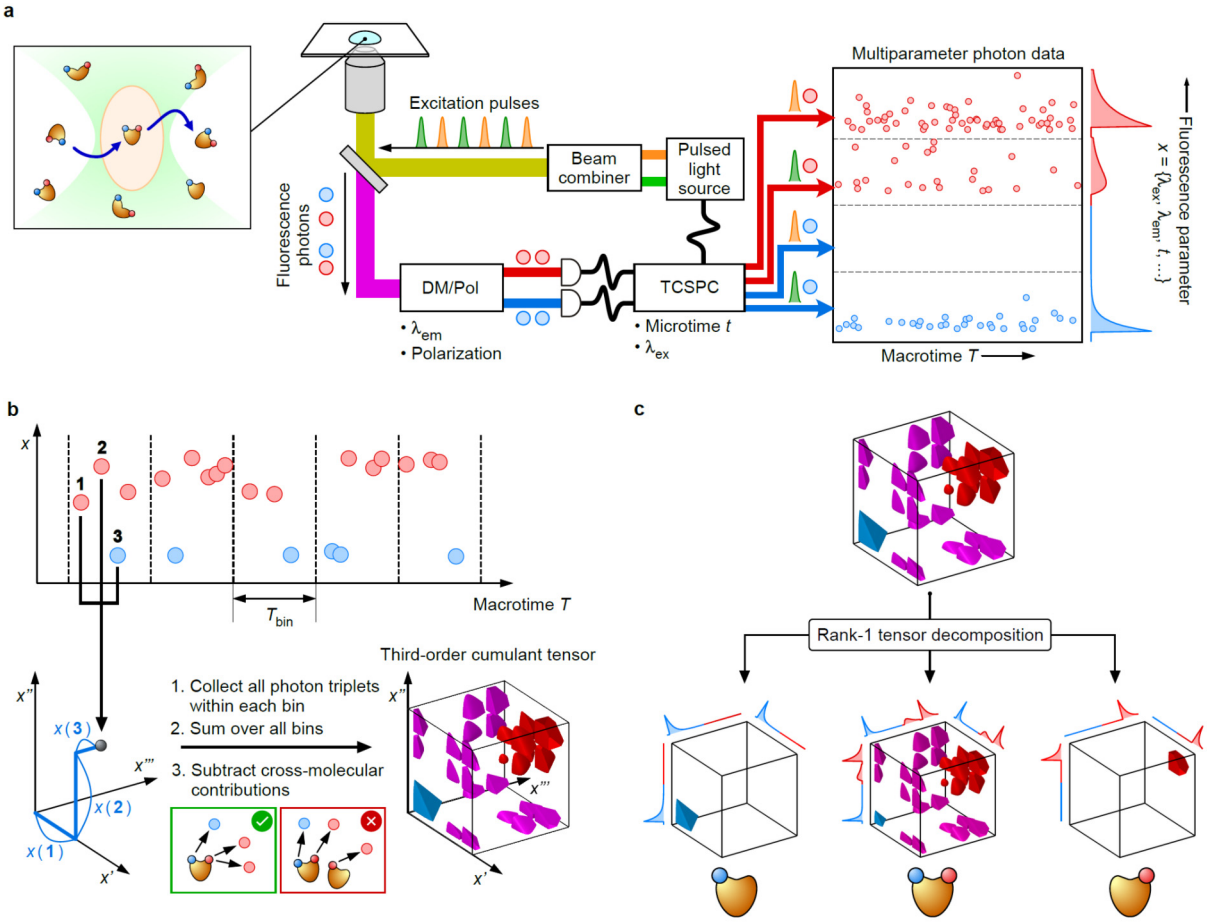


Fig. 1 Principle of IFCA

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 the third-order cumulant tensor into a sum of independent fluorescence components.

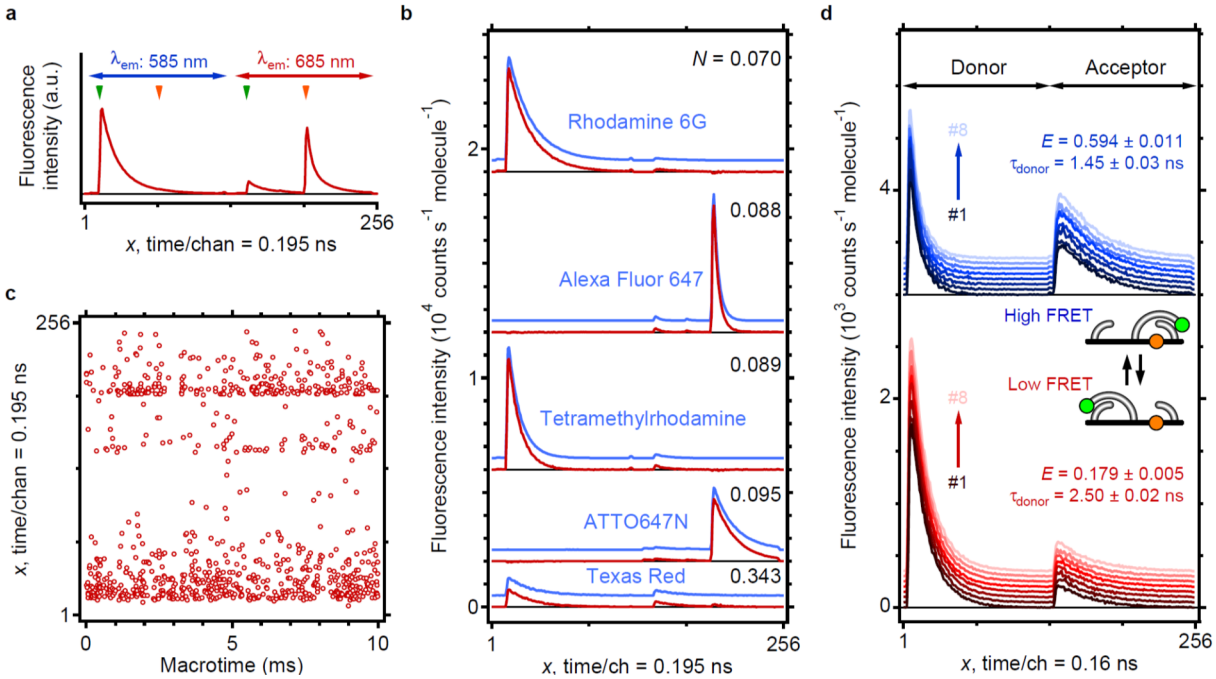


Fig. 2 Applications of IFCA

a, Ensemble-averaged fluorescence pattern of the mixture of five fluorophores. The arrival times of two-color excitation pulses are indicated by downwards arrowheads (green: 525 nm, orange: 640 nm). **b,** 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. **c,** Representative portion of the photon data of the mixture sample. **d,** Fluorescence patterns of IFCs retrieved by the IFCA analysis on a FRET-labeled single-stranded DNA. Two IFCs obtained for the eight data sets are plotted separately with vertical offsets for clarity. E and τ_{donor} are the apparent FRET efficiency and the mean donor lifetime calculated from each of the two fluorescence patterns and presented in the form of mean ± SD evaluated from the eight data sets.

Online methods

IFCA analysis. *Structure of multiparameter photon data.* In multiparameter single-molecule fluorescence measurements (Fig. 1a), a dilute solution of fluorescent molecules is photoexcited 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 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. Thus, actual data is a list of macrotimes in ascending order augmented with additional fluorescence parameters for each photon. In this data, λ_{em} and polarization are represented as the detector index c that is added by a routing module (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, $t \leq 64$ corresponds to a photon emitted after excitation with the first color and $t > 64$ corresponds to a photon emitted after excitation with the second color). We collectively denote these additional 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 IFCs represented as functions of x. The meaning of the conversion from c, t to x can be understood by considering a histogram of x for all detected photons, which becomes a concatenated trace of multiple fluorescence decay curves at different conditions as seen in Fig. 2a. However, note that the order of the individual decay curves (determined by c) and even the order of individual data points (determined by t) can be arbitrarily selected, since they do not

affect the performance of IFCA.

Use of the third-order cumulant tensor. IFCA analyzes the fluctuation of fluorescence intensity at channel x, $I_x(T)$, as a function of macrotime T. For this purpose, we utilize the third-order cross-cumulants of $I_x(T)$ that are defined as expectation values[6]:

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

where $\bar{I}_x \equiv \mathrm{E}[I_x(T)] = \frac{1}{T_0}\int_0^{T_0} I_x(T)\, dT$ 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[6]. 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[9]. Because of this property, the above decomposition of $\boldsymbol{\mathcal{T}}$ into the contributions of independent sources, $\boldsymbol{\mathcal{T}}^s$, is uniquely determined. 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).

Construction of the third-order cumulant tensor. 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[17]. In the fluctuation analyses of fluorescence photons, it is known that the factorial cumulants of photon counts should be used instead of ordinary cumulants of fluorescence intensity[7, 18]. In the present case, this can be accomplished by evaluating $\boldsymbol{\mathcal{T}}$ based on the 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 normalized with 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 a 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 $\mathbf{M}$:

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

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

which are computed by histogramming photon itself 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[19].

Effect of the bin width. The above formulation assumes that the fluorescence intensity $I_x(T)$ does not fluctuate within each bin. Therefore, in principle, T_{bin} should be short enough as compared with the dynamics of the molecule such as intramolecular conformational change or diffusion across the observation volume. However, such a short T_{bin} may deteriorate the performance of IFCA because it reduces the available number of photon triplets. Practically, it is recommended to choose a T_{bin} value either sufficiently shorter than or much longer than the time scale of intramolecular dynamics, so that the dynamical effect is well resolved or entirely averaged out, respectively. In this sense, T_{bin} determines the time resolution to detect transient species. On the other hand, the diffusion dynamics does not affect the separation of IFCs and only influences quantitative determination of the molecular brightness and the absolute concentrations (see Extended Data Fig. 2b,d,e). Therefore, a relatively long T_{bin} may be used when intramolecular dynamics is negligible. Nevertheless, if necessary, the effect of finite bin-size due to the molecular diffusion on the cumulants of fluorescence photons can be quantitatively estimated using previously reported methods[20, 21].

Workaround for detector nonideality. Another aspect that potentially deteriorates the performance of IFCA is detector nonideality, i.e., detection dead time and afterpulsing, which artificially modulates $I_x(T)$ in the nano- to microsecond time scale[7, 22]. As a workaround for this problem, we use only the photon triplets and photon pairs in which all the photon time intervals are longer than a threshold value T_{g} in the computation of $\boldsymbol{\mathcal{T}}$ and $\mathbf{M}$, respectively.

This selection process results in a decrease in the number of available photon triplets and photon pairs, which eventually affects the accuracy of calculations in Eqs. (3) and (5). To compensate for this effect, the obtained tensors, $\boldsymbol{\mathcal{T}}^{\mathrm{g}}$ and $\mathbf{M}^{\mathrm{g}}$, are rescaled as

$$\mathcal{T}_{ijk} = \mathcal{T}_{ijk}^{\mathrm{g}} \times \frac{T_{\mathrm{bin}}^{3}}{(T_{\mathrm{bin}} - 2T_{\mathrm{g}})(T_{\mathrm{bin}} - 2T_{\mathrm{g}} - T_{\mathrm{c}})(T_{\mathrm{bin}} - 2T_{\mathrm{g}} - 2T_{\mathrm{c}})}, \tag{6}$$

$$M_{ij} = M_{ij}^{\mathrm{g}} \times \frac{T_{\mathrm{bin}}^{2}}{(T_{\mathrm{bin}} - T_{\mathrm{g}})(T_{\mathrm{bin}} - T_{\mathrm{g}} - T_{\mathrm{c}})}. \tag{7}$$

Here, T_{c} indicates the macrotime clock period and is included for reflecting that two photons cannot possess the same macrotimes.

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[9], 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[23] 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

$$\mathcal{T}'_{ijk} \equiv \frac{\mathcal{T}_{ijk}}{\sqrt{v_i v_j v_k}}, \tag{8}$$

$$M'_{ij} \equiv \frac{M_{ij}}{\sqrt{v_i v_j}}. \tag{9}$$

EVD of the scaled second-order cumulant tensor is

$$\mathbf{M}' = \mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}}, \tag{10}$$

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 chosen based on the comparison of the eigenvalues and the noise level (see Extended Data Fig. 1a-c, 3b, 3c). Using $\mathbf{B}$, $\mathbf{W}$, and n, it can be shown that an element of third-order cumulant tensor $\boldsymbol{\mathcal{T}}'$ is approximated as a sum over n independent components (Supplementary Note 1):

$$\mathcal{T}'_{ijk} \approx \sum_{s=1}^{n} \left(\sum_{p=1}^{n} B_{ip}\sqrt{W_p}U_{sp} \right) \left(\sum_{q=1}^{n} B_{jq}\sqrt{W_q}U_{sq} \right) \left(\sum_{r=1}^{n} B_{kr}\sqrt{W_r}U_{sr} \right) h_s\,, \tag{11}$$

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 (Supplementary Note 1), and U_{sp} etc. 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. (11) 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:

$$\mathcal{T}''_{pqr} \equiv \sum_{i=1}^{x_0} \sum_{j=1}^{x_0} \sum_{k=1}^{x_0} B_{ip}B_{jq}B_{kr}\mathcal{T}'_{ijk} \tag{12}$$

$$\approx \sum_{s=1}^{n} \sqrt{W_pW_qW_r}U_{sp}U_{sq}U_{sr}h_s\,, \tag{13}$$

where $1 \le p, q, r \le n$. 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. (12)) with Eq. (13) by using $h_1, \cdots, h_n$ and $\mathbf{U}$ as fitting parameters (Extended Data Fig. 1d, 3d). Fitting involving an

orthogonal matrix as a fitting parameter can be performed by a Jacobi-type algorithm[23] (Supplementary Note 2). Finally, the fluorescence patterns of IFCs are reconstructed from the obtained parameters (Supplementary Note 1):

$$d_x^s \approx \sqrt{v_x}\left(\sum_{p=1}^{n} B_{xp}\sqrt{W_p}U_{sp}\right)h_s. \tag{14}$$

Here, the factor $\sqrt{v_x}$ is multiplied to compensate for the noise normalization (Eqs. (8,9)). 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. Its effect on the IFCs derived from the experimental data on a mixture sample of five fluorophores was examined and shown in Extended Data Fig. 2a-c.

Evaluation of concentrations. The concentrations of fluorescent species representing IFCs are evaluated as the average numbers of molecules in the observation volume, N_s. N_s is obtained in the above procedure as (Supplementary Note 1)

$$N_s \approx \frac{1}{h_s^2}. \tag{15}$$

Note that this is the average number of molecules in the observation volume having a well-defined boundary. However, 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 general, the value obtained with Eq. (15) differs from that defined in FCS, N_s^{FCS}:

$$N_s^{\mathrm{FCS}} = \frac{\gamma_3^2}{\gamma_2^4}N_s, \tag{16}$$

where γ_r are the shape factors characterizing the observation volume and $\gamma_r = r^{-\frac{3}{2}}$ for 3-D gaussian profiles[7].

Experiment on a mixture of fluorophores. *Multiparameter fluorescence measurement.* To

demonstrate the feasibility of IFCA, we performed a test experiment using a home-built multiparameter fluorescence measurement setup. The apparatus is based on the PIE scheme[10] and essentially the same as that was reported previously[11, 24]. The excitation light source is a supercontinuum laser (Fianium SC-400-4, 40 MHz). The visible part of its output was split into two by a dichroic mirror (Thorlabs DMLP567) and used as donor- and acceptor-excitation pulses. They passed through bandpass filters (Semrock, FF01-525/30-25 & FF01-640/14-25) and the timing were adjusted by adding an optical delay to the acceptor-excitation pulses using a 1-m long single-mode fiber (Thorlabs P1-460B-FC-1), before recombined by another dichroic mirror (Thorlabs DMLP567) and a single-mode fiber (Thorlabs P5-460B-PCAPC-1). They were introduced into a microscope (Nikon Eclipse Ti) and focused onto a sample solution through a water-immersion objective lens (Nikon CFI Plan Apo IR 60×, N.A.=1.27). The sample solution was filled in a cell made of two coverslips and a silicone spacer. Fluorescence light emitted from the sample was collected with the objective and separated from the scattering of excitation light by a dual-band dichroic mirror (Chroma ZT532/640rpc). After passing through a multi-mode fiber (Thorlabs FG050LGA, core diameter = 50 μm), which acted as a confocal pinhole, the donor and acceptor fluorescence were separated by a dichroic mirror (Chroma Technology ZT633rdc) and detected by two photon-counting detectors (Becker & Hickl HPM-100-40-C) after passing through bandpass filters (Semrock FF01-585/40-25 & FF02-685/40-25). For each photon detection event, the detector ID was tagged by a routing module (Becker & Hickl HRT-41) and the excitation-emission delay time (microtime) was measured by a TCSPC module (Becker & Hickl SPC-130-EM) in the FIFO mode, i.e., the absolute arrival time (macrotime) was also recorded. The list of macrotime, microtime, and detector ID was saved in the native file format of Becker & Hickl (.spc). In IFCA analysis, this file was loaded by using the *bhreader* module from a Python package *phconvert* [25].

Sample preparation. We measured a mixed dye solution of 5(6)-carboxy tetramethyl rhodamine

(TMR, BioReagent), rhodamine 6G (R6G, Sigma-Aldrich), Alexa Fluor 647 carboxylic acid tris(triethylammonium) salt (Alexa647, Invitrogen), ATTO647N free COOH (ATTO-TEC), and Texas Red-X succinimidyl ester mixed isomer (TR, Invitrogen). The mixture sample was prepared in concentrations of 1 nM, 1 nM, 1 nM, 1 nM, 5 nM for TMR, R6G, Alexa647, ATTO647N, TR, respectively, in 50 mM PBS buffer (pH7.2) containing 0.01% Triton X-100. Reference samples containing only single dyes were prepared in the same concentration and buffer conditions as the mixture sample.

IFCA analysis. With the above PIE-TCSPC setup, the sample was photoexcited by two interleaving pulse trains and the fluorescence signal was detected at two wavelengths, i.e., $c_0 = 2$. In the data analysis, 32 adjacent microtime channels in 4096 raw channels (6.1 ps/ch) were binned together, resulting in $t_0 = 128$ channels (0.195 ns/ch). Consequently, the fluorescence parameter x in IFCA consists of two parts; the first (x = 1-128) and second (x = 129-256) regions correspond to the microtime t of the photons detected at λ_{em} = 585 nm and 685 nm, respectively (Fig. 2a). The bin width T_{bin} was set to 5 μs and the data accumulation time was ~1 hr for each data set. The minimum photon interval T_g and the clock period T_c were 101 ns and 50.5 ns, respectively. Note that the afterpulsing effect is practically negligible in the detectors used in this experiment. Three independent data sets were obtained from the same sample and analyzed by IFCA, among which one is presented in Fig. 2a-c and Extended Data Fig. 1, 2a-e and the other two are shown in Extended Data Fig. 2f, g. The fluorescence patterns of individual fluorophores (Fig. 2b, blue lines) were obtained by applying IFCA to the reference samples containing only single dyes to remove the contributions of background and impurities, where we set $n = 2$ for each sample and only the predominant component is shown. All analyses were performed by using a custom code written in Python.

Analysis on photon data of FRET-labeled single-stranded DNA. *Data source.* The photon data analyzed in the second example (Fig. 2d) were obtained from a public repository, Zenodo[13].

These data sets were recorded by Schröder et al.[12] by a TCSPC-based FCS measurement on a single-stranded DNA, which was labeled with Cy3B as FRET donor and tethered on a small DNA origami structure. 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. Therefore, 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 through the change of FRET efficiency (Fig. 2d, inset). The switching time were reported to be 330 and 700 μs[12]. The whole DNA origami structure was freely diffusing in a buffer with 100 pM concentration. For IFCA analysis, photon data were loaded by using the *pqreader* module from the *phconvert* package.

IFCA analysis. The microtime resolution of original data, 1280, was reduced by binning 10 adjacent channels to make $t_0 = 128$ (0.16 ns/ch). The first half of the fluorescence pattern (x = 1-128) is from the donor detector and the second half (x = 129-256) is from the acceptor detector. The bin width T_{bin} was set to 20 μs. The minimum photon interval T_{g} and the clock period T_{c} were 2 μs and 20 ns, respectively. The number of components was fixed to $n = 2$. The analysis was repeated for eight data sets recorded with 600 s accumulation time for each. The results of EVD analysis (Eq. (10)) and the fitting to third-order cumulant tensor (Eqs. (12,13)) for the first data set are shown in Extended Data Fig. 3. From the fluorescence patterns of two components (d_x^1 and d_x^2), we calculated the apparent FRET efficiency E and the mean donor lifetime τ_{donor} as follows.

$$E^s = \frac{\sum_{x=129}^{256} d_x^s}{\sum_{x=1}^{256} d_x^s},$$

$$\tau_{\mathrm{donor}}^s = \left[\frac{\sum_{x=1}^{128} x d_x^s}{\sum_{x=1}^{128} d_x^s} - x_{\mathrm{IRF}}\right] \times 0.16\ \mathrm{ns},$$

where x_{IRF} is the time origin defined as the first moment of the instrument response function.

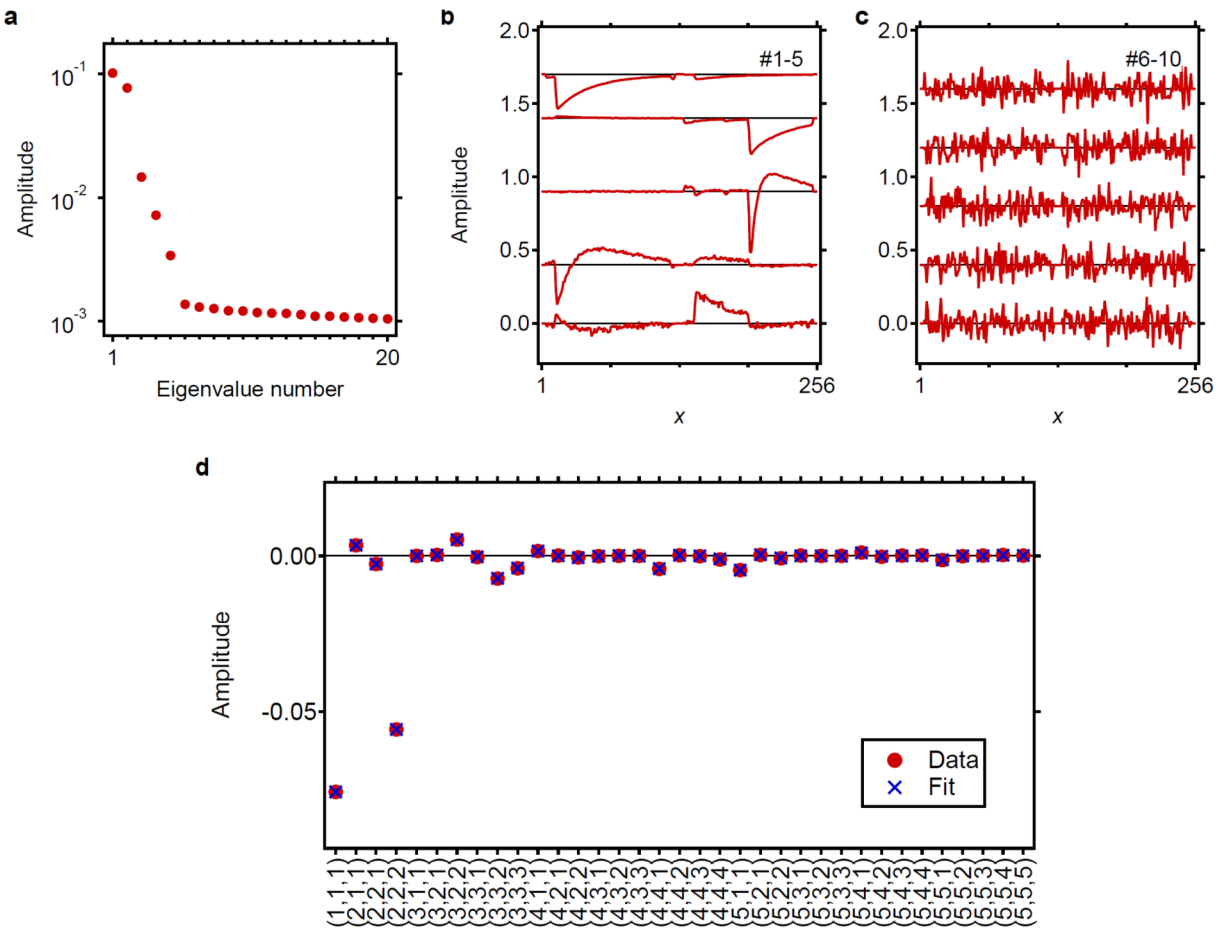
Data availability. Data for the dye mixture experiment have been deposited at Zenodo and are publicly available at https://doi.org/10.5281/zenodo.14869092. Data for the FRET-labeled single-stranded DNA are made publicly available by Schröder et al. and accessible at https://doi.org/10.5281/zenodo.7254095.

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.

References

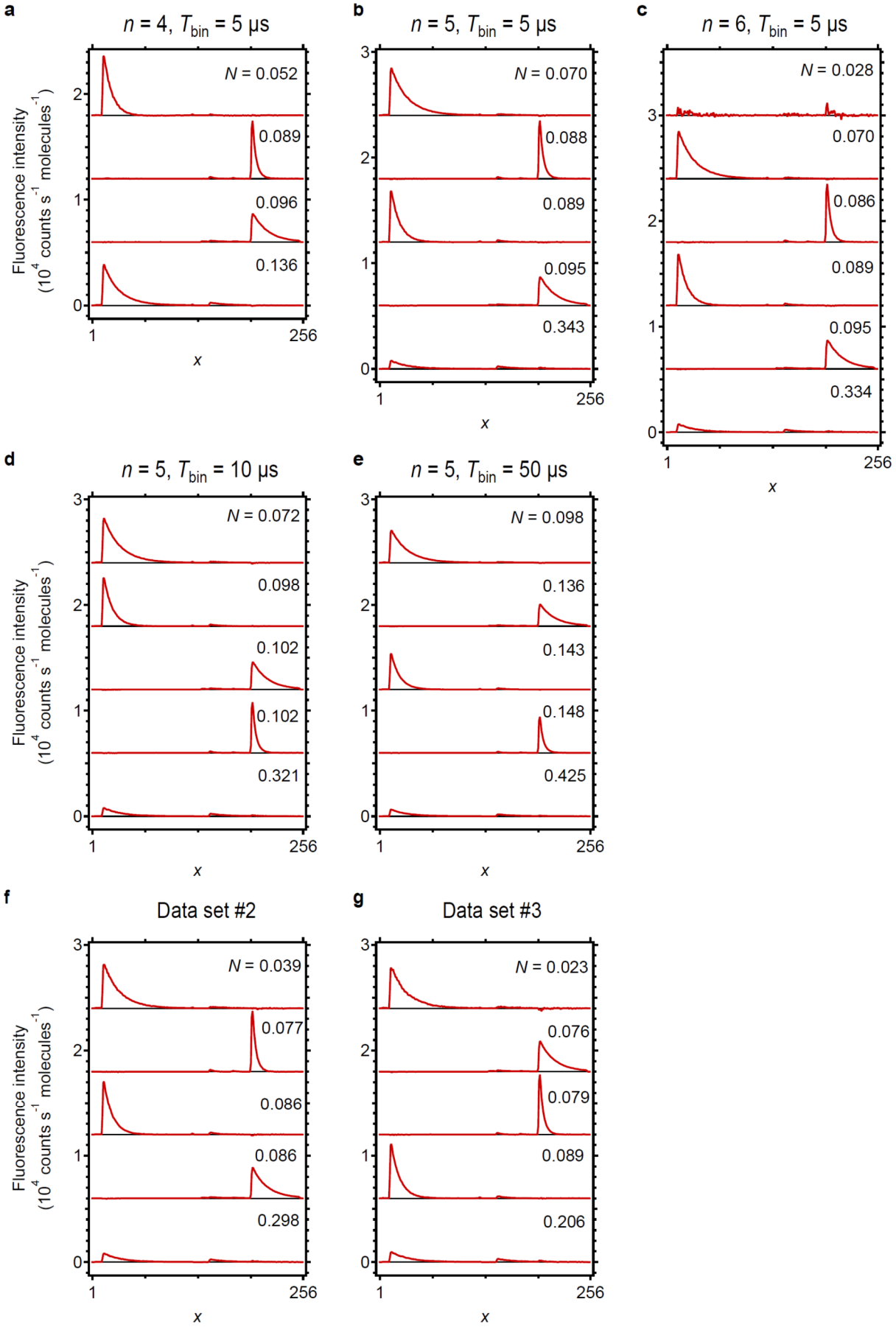
17. Mandel, L. Fluctuations of photon beams and their correlations. *Proc. Phys. Soc.* **72**, 1037-1048 (1958).
18. Abdollah-Nia, F. Cumulant-based formulation of higher-order fluorescence correlation spectroscopy. *arXiv*, 1612.05388v05382 (2018).
19. Ishii, K. & Tahara, T. Extracting decay curves of the correlated fluorescence photons measured in fluorescence correlation spectroscopy. *Chem. Phys. Lett.* **519-520**, 130-133 (2012).
20. Wu, B. & Müller, J.D. Time-integrated fluorescence cumulant analysis in fluorescence fluctuation spectroscopy. *Biophys. J.* **89**, 2721-2735 (2005).
21. Wu, B., Chen, Y. & Müller, J.D. Dual-color time-integrated fluorescence cumulant analysis. *Biophys. J.* **91**, 2687-2698 (2006).
22. 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* **121**, 2373-2387 (2017).
23. De Lathauwer, L. in Handbook of Blind Source Separation. (eds. P. Comon & C. Jutten) 155-177 (Academic Press, Oxford; 2010).
24. Heo, W. et al. Scanning two-dimensional fluorescence lifetime correlation spectroscopy: conformational dynamics of DNA Holliday junction from microsecond to subsecond. *J. Phys. Chem. Lett.* **13**, 1249-1257 (2022).
25. 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.* **110**, 26-33 (2016).

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

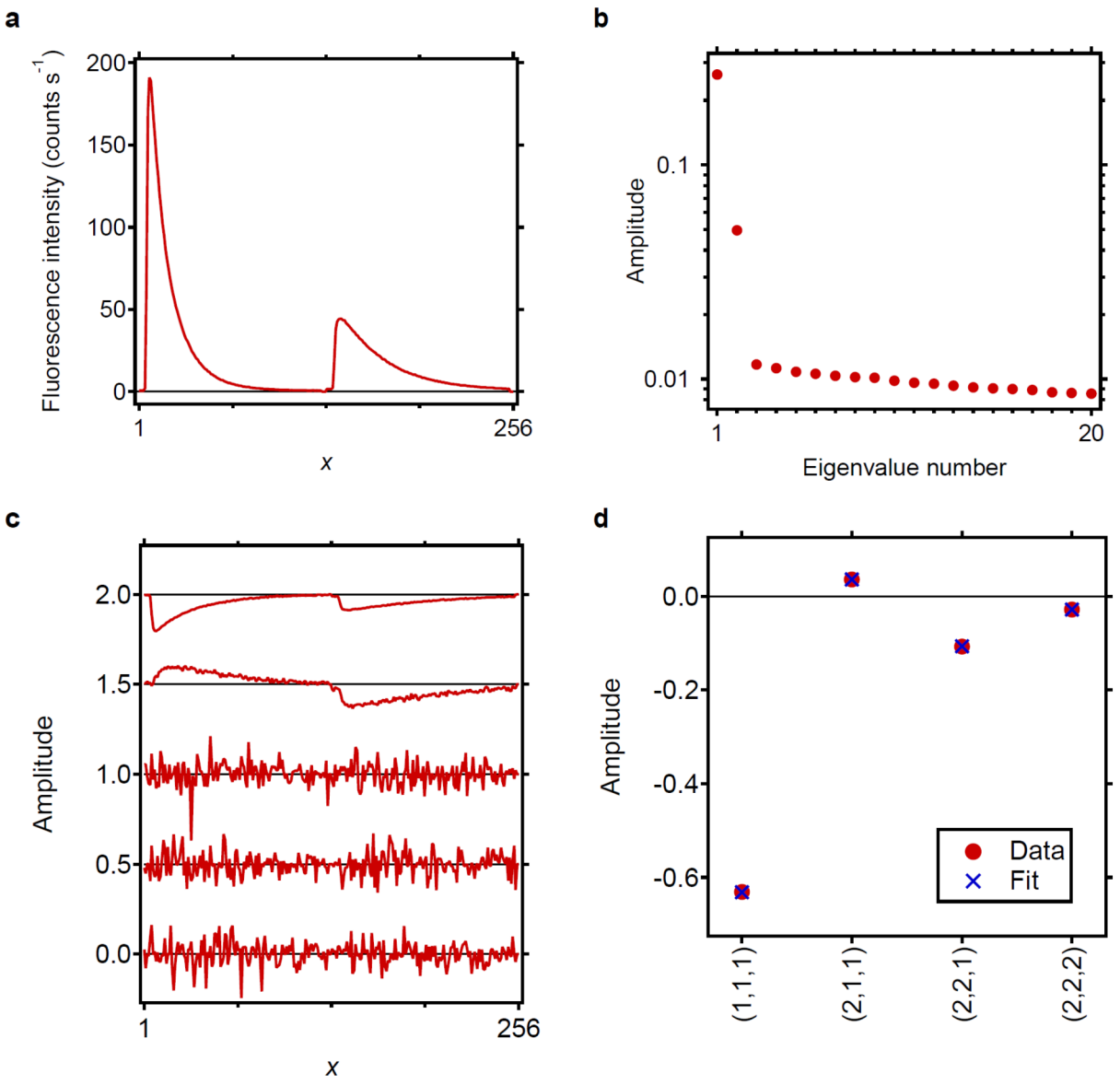


Extended Data Fig. 1: IFCA analysis of the dye mixture sample.

a, The largest 20 eigenvalues of the scaled second-order cumulant tensor. The first five eigenvalues are clearly beyond the noise level. **b,c,** Eigenvectors associated with the largest 10 eigenvalues (#1-5 (**b**) and #6-10 (**c**), from top to bottom). Traces are vertically offset for clarity. **d,** Result of the fitting to the elements of the reduced third-order cumulant tensors. Red filled circles: amplitudes calculated from photon data using Eq. (12). Blue crosses: fitted values based on Eq. (13).


a
n = 4, T_bin = 5 μs
N = 0.052
0.089
0.096
0.136
b
n = 5, T_bin = 5 μs
N = 0.070
0.088
0.089
0.095
0.343
c
n = 6, T_bin = 5 μs
N = 0.028
0.070
0.086
0.089
0.095
0.334
d
n = 5, T_bin = 10 μs
N = 0.072
0.098
0.102
0.102
0.321
e
n = 5, T_bin = 50 μs
N = 0.098
0.136
0.143
0.148
0.425
f
Data set #2
N = 0.039
0.077
0.086
0.086
0.298
g
Data set #3
N = 0.023
0.076
0.079
0.089
0.206
Fluorescence intensity
(10^4 counts s^-1 molecules^-1)
x
1
256
0
1
2
3

Extended Data Fig. 2: Validation of the stability of IFCA analysis on the mixture sample. **a-c,** Fluorescence patterns of IFCs obtained with different choices of the number of components, n. Binning width was set to $T_{\mathrm{bin}} = 5\ \mu\mathrm{s}$. Note that when n is incremented from 4 to 5, a new IFC (the lowest trace in **b**) newly appears, while one more increment to 6 results in appearance of an unphysical noisy IFC (the top trace in **c**). **d,e,** Fluorescence patterns of IFCs obtained with different binning widths ($n = 5$). **f,g,** Fluorescence patterns of IFCs obtained with other data sets ($T_{\mathrm{bin}} = 5\ \mu\mathrm{s}$ and $n = 5$). In all panels, traces are vertically offset for clarity.



Extended Data Fig. 3: IFCA analysis of the FRET-labeled single-stranded DNA sample. Details of the analysis are presented for the first data set (#1 in Fig. 2d). **a,** Ensemble-averaged fluorescence pattern. **b,** The largest 20 eigenvalues of the scaled second-order cumulant tensor. **c,** Eigenvectors associated with the largest five eigenvalues (from top to bottom). Traces are vertically offset for clarity. **d,** Elements of the reduced third-order cumulant tensor (red filled circles) and the fitting results (blue crosses).