Supplementary Information

High Resolution Fluorescence Lifetime Maps from Minimal Photon Counts

Mohamadreza Fazel[1], Sina Jazani[1], Lorenzo Scipioni[2,3], Alexander Vallmitjana[2,3], Enrico Gratton[2,3], Michelle A. Digman[2,3] and Steve Pressé[1,4]

[1] Center for Biological Physics, Department of Physics, Arizona State University, Tempe, AZ 85287, USA

[2] Department of Biomedical Engineering, University of California Irvine, Irvine, CA 92697, USA

[3] Laboratory of Fluorescence Dynamics, The Henry Samueli School of Engineering, University of California, Irvine, CA 92697, USA

[4] School of Molecular Science, Arizona State University, Tempe, AZ 85287, USA

January 24, 2022

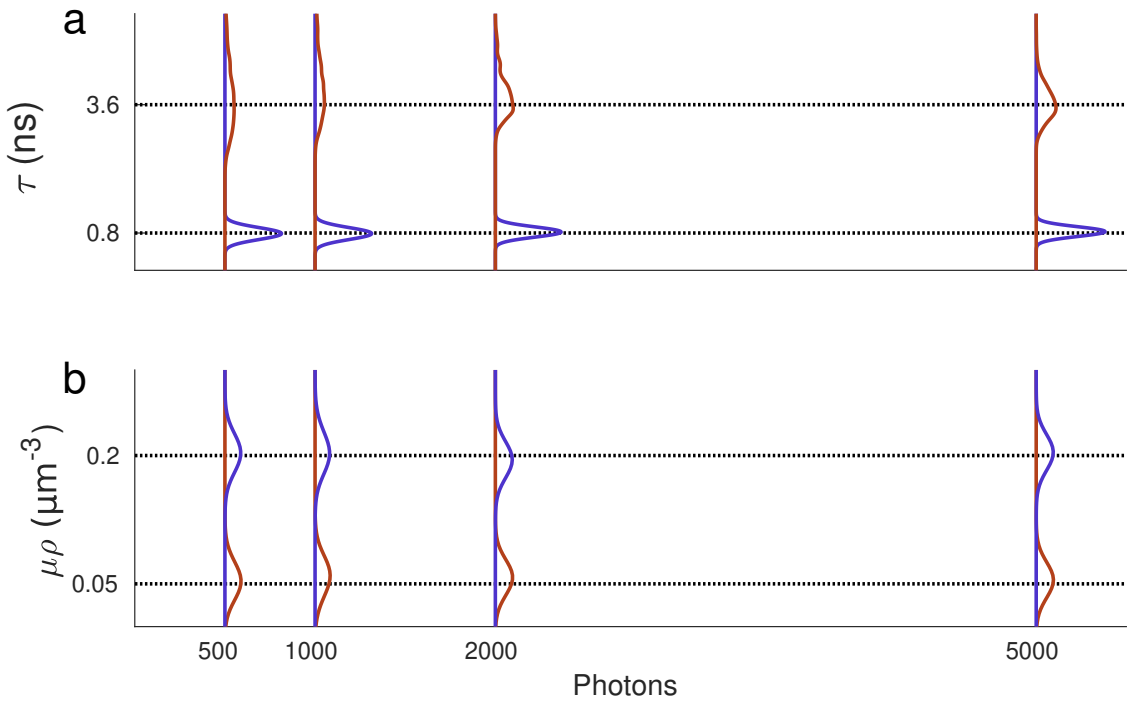

Fig. S1: Lifetime and $\mu\rho$ posteriors for data simulated with lifetimes of 0.8 and 3.6 ns, uniform $\mu\rho$ of 0.2 and 0.05 μm^{-3} and total photon counts of 500, 1,000 2,000 and 5,000. (a) Lifetime posteriors. (b) $\mu\rho$ posteriors. The blue and brown curves, respectively, represent posteriors of species with $\mu\rho$ of 0.2 and 0.05 μm^{-3}. The black dashed lines represent the ground truths.

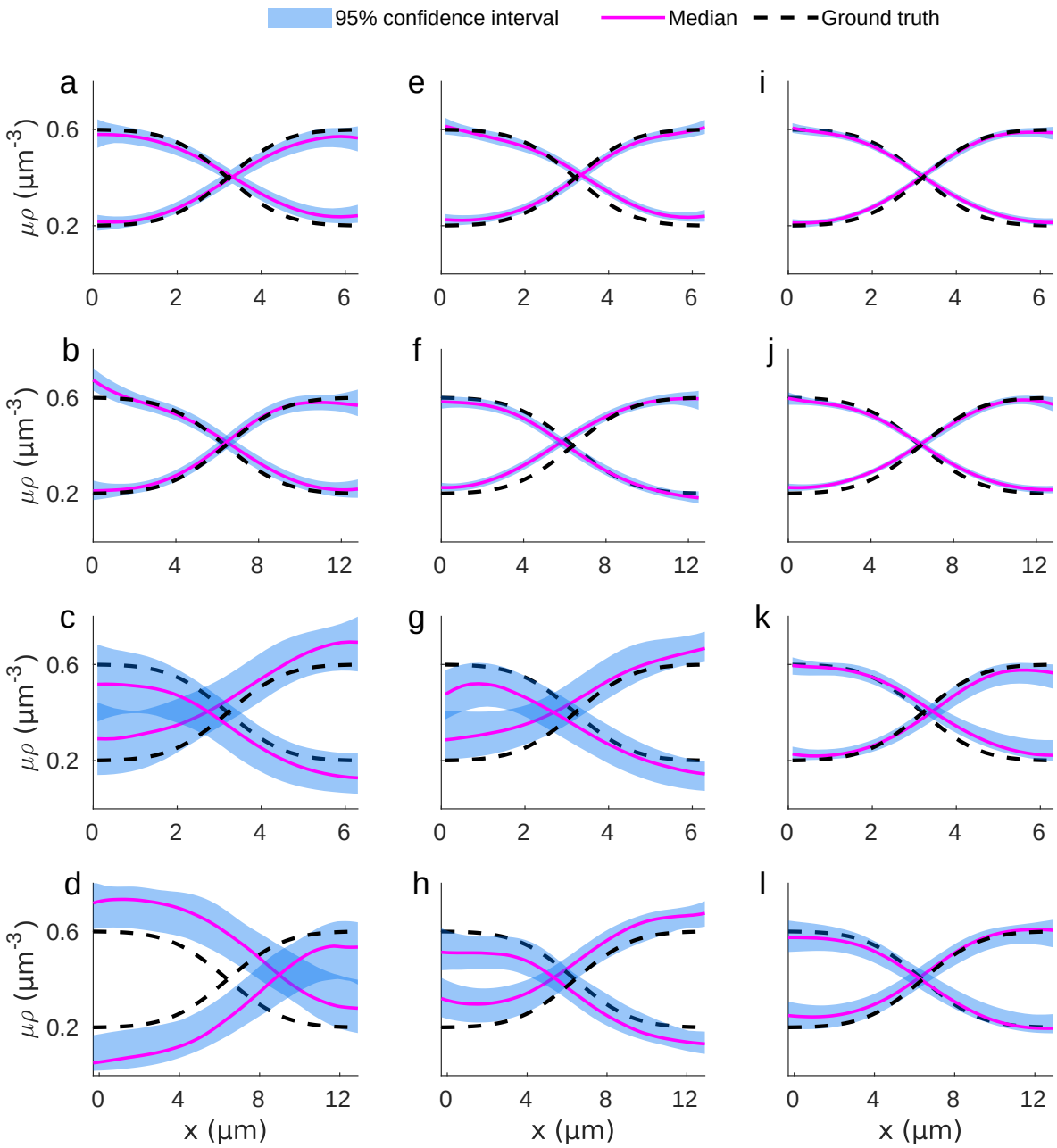

Fig. S2: Cross-sections of 3D non-uniform $\mu\rho$ profiles along x-axis. The first and second rows represent resulting $\mu\rho$ from data sets simulated with two species and a large gap in lifetimes (0.8, 3.6) ns. The third and fourth rows show resulting $\mu\rho$ from data sets with more similar lifetimes (2.5, 3.6) ns. (a-d), (e-h) and (i-l), respectively, depict results from simulated data with 50, 100 and 200 photons per pixel. The data used in the first and third rows were simulated using a pixel size of 0.2 μm and pixel size for the second and last rows was 0.4 μm. All the data were generated over a region of 32×32 pixels.

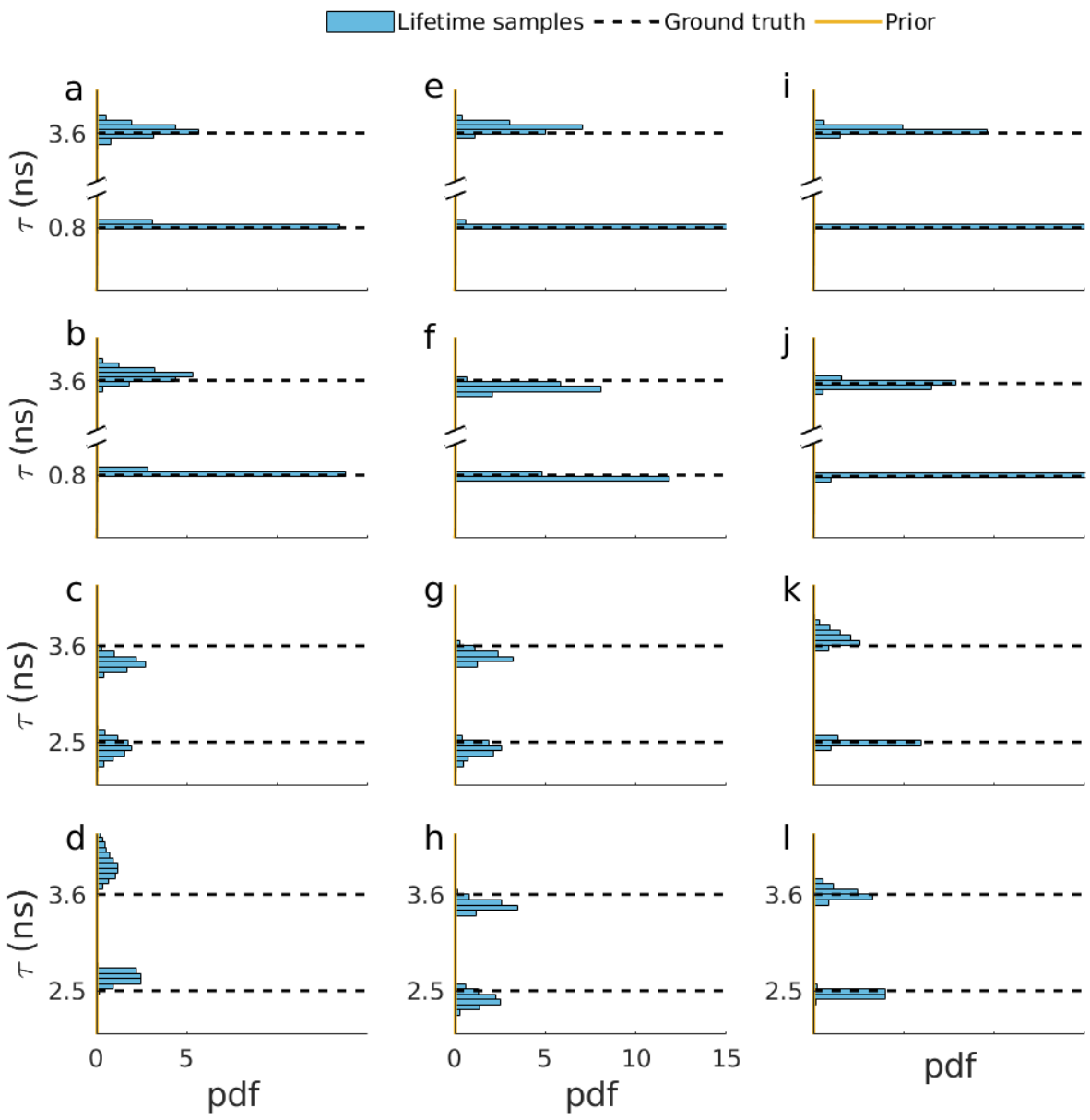

Fig. S3: Lifetimes of non-uniform profiles corresponding to the $\mu\rho$ profiles in Fig. S2. The black dashed lines and the solid yellow lines show the ground truth lifetimes and the used priors, respectively. The used priors are also depicted in Fig. S19. The same line-style convention is followed throughout the SI.

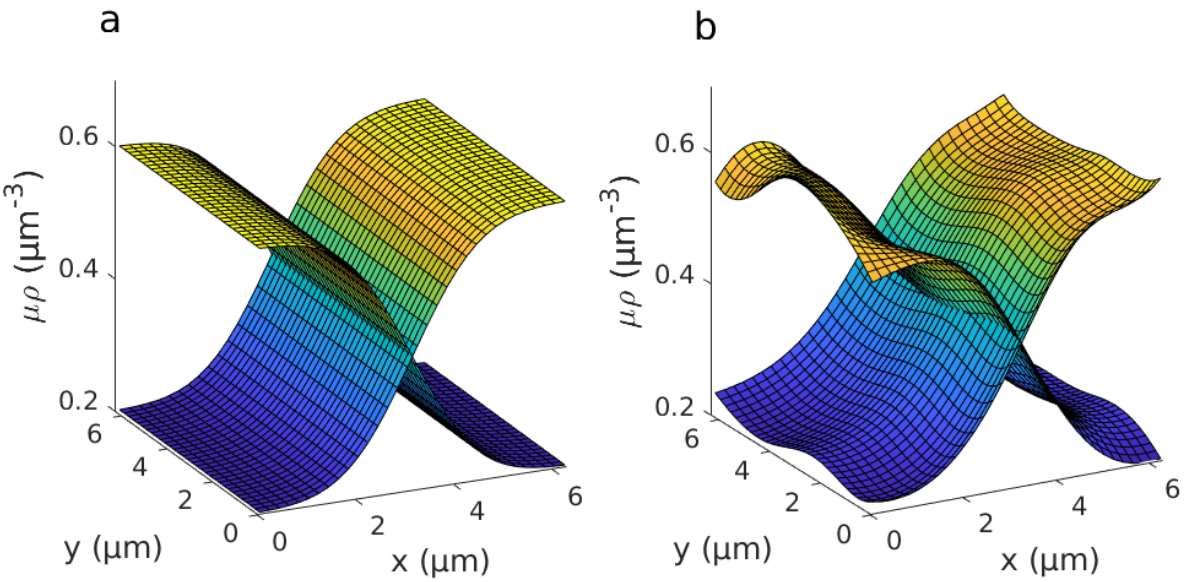

Fig. S4: An example of 3D profiles used in Fig. S2. 3D plots in panels a and b depict the ground truth and the median of $\mu\rho$ profiles sampled from the posterior for a data simulated with two species, lifetimes of 0.8 and 3.6 ns and 100 photons per pixel over 32×32 pixels. The cross sections shown in Fig. S2e was taken from the profiles shown here. The $\mu\rho$ profiles do not vary along the z-direction. Grids represent the pixel centers.

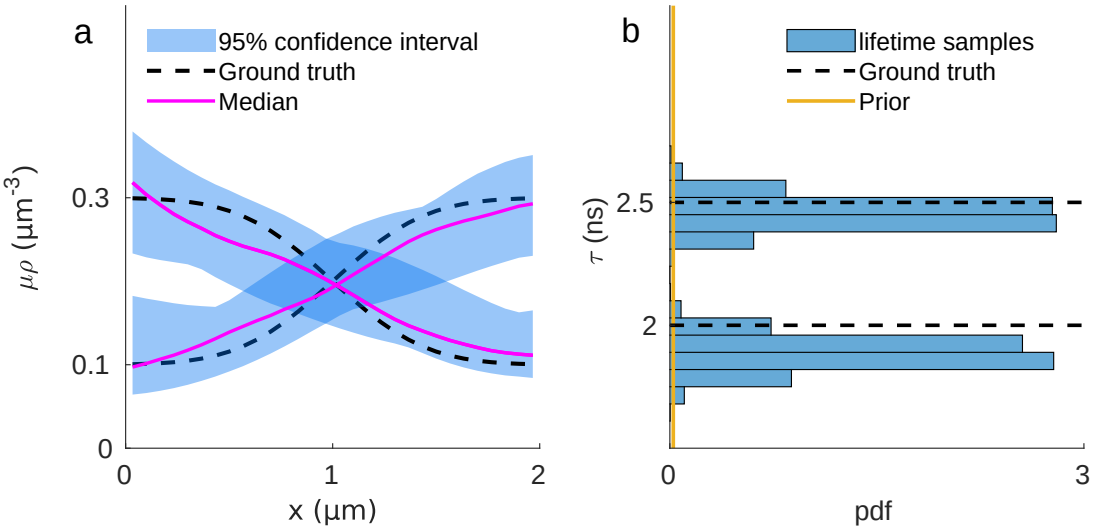

Fig. S5: Sub-nanosecond lifetime resolution. (a) Cross-sections of 3D $\mu\rho$ profiles for non-uniform profiles along x-axis simulated with lifetimes of (2, 2.5) ns, 200 photons per pixel over 10×10 pixels and pixel size of 0.2 μm. (b) Histogram of the sampled lifetimes.

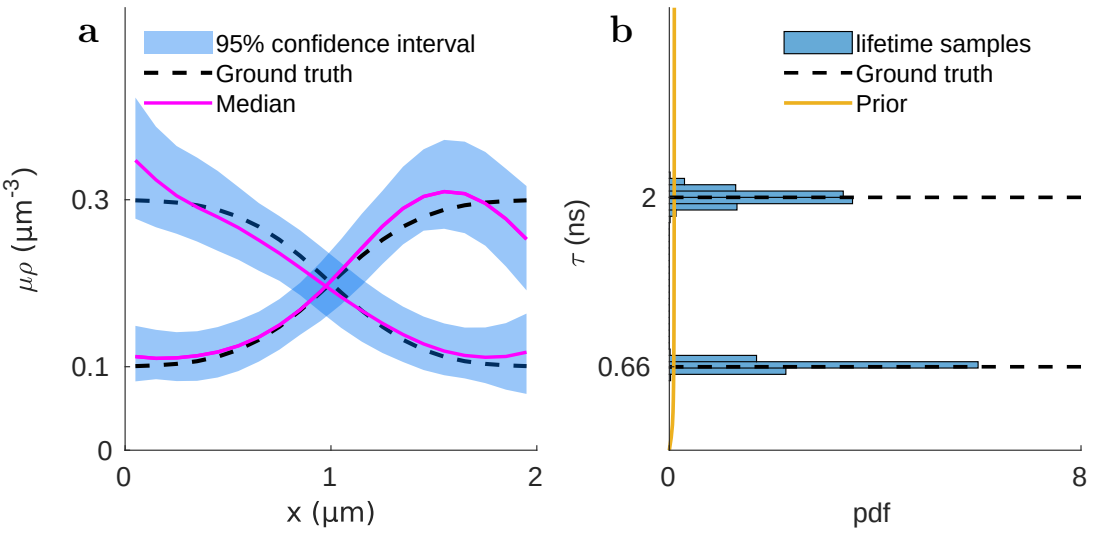

Fig. S6: Lifetime similar to σ_{IRF} (0.66 ns). (a) Cross-sections of 3D $\mu\rho$ profiles for non-uniform profiles along x-axis simulated with lifetimes of (2, 0.66) ns, 100 photons per pixel over 10×10 pixels and pixel size of 0.2 μm. (b) Histogram of the sampled lifetimes.

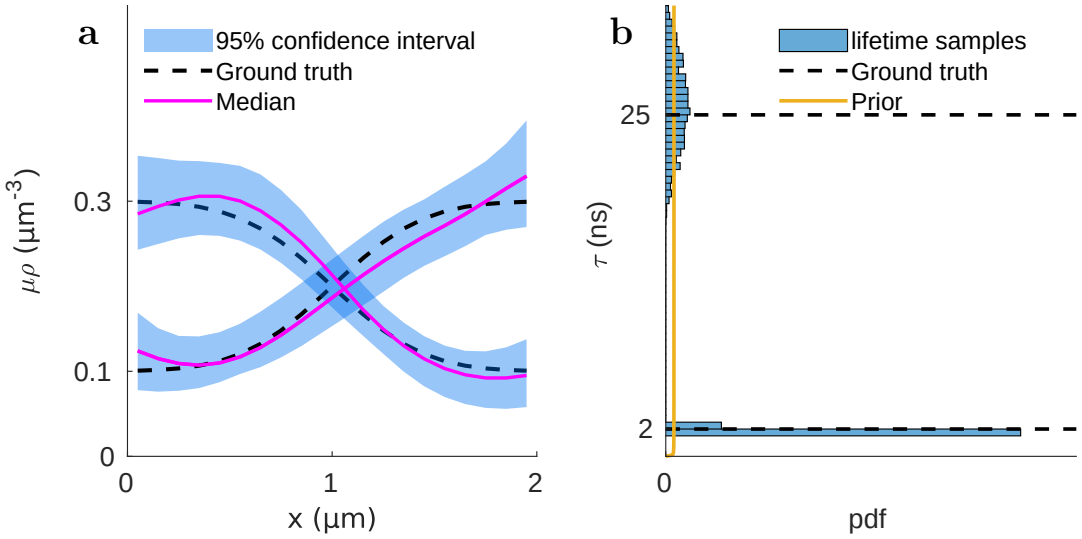

Fig. S7: Lifetime exceeding the inter-pulse time ($\sim$ 2 times larger than the inter-pulse time, $T = 12.8$ ns). (a) Cross-sections of 3D $\mu\rho$ profiles for non-uniform profiles along x-axis simulated with lifetimes of (2, 25) ns, 100 photons per pixel over 10×10 pixels and pixel size of $0.2\,\mu m$. (b) Histogram of the sampled lifetimes. While the lifetime larger than the inter-pulse period is detected and the ground truth falls within the range of the histogram, it is very wide implying a high uncertainty.

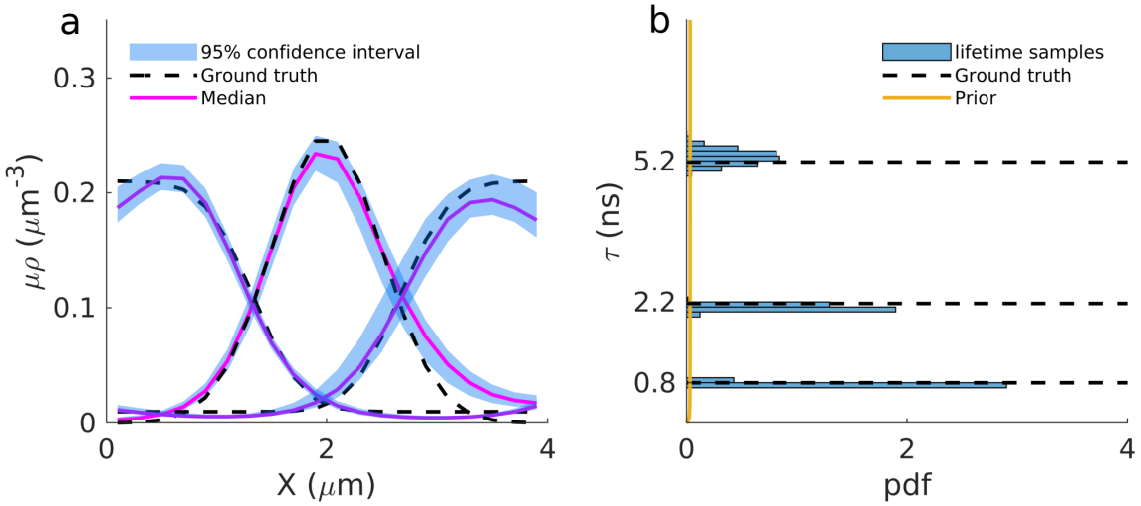

Fig. S8: Cross-section of 3D $\mu\rho$ profiles for simulated data with 3 species, lifetimes of 0.8, 2.2 and 5.2 ns, 100 photons per pixel over a region of 20×20 pixels and a pixel size of 0.2 μm. (a) The cross section of $\mu\rho$ profiles along the x-axis. (b) Histogram of the found lifetimes.

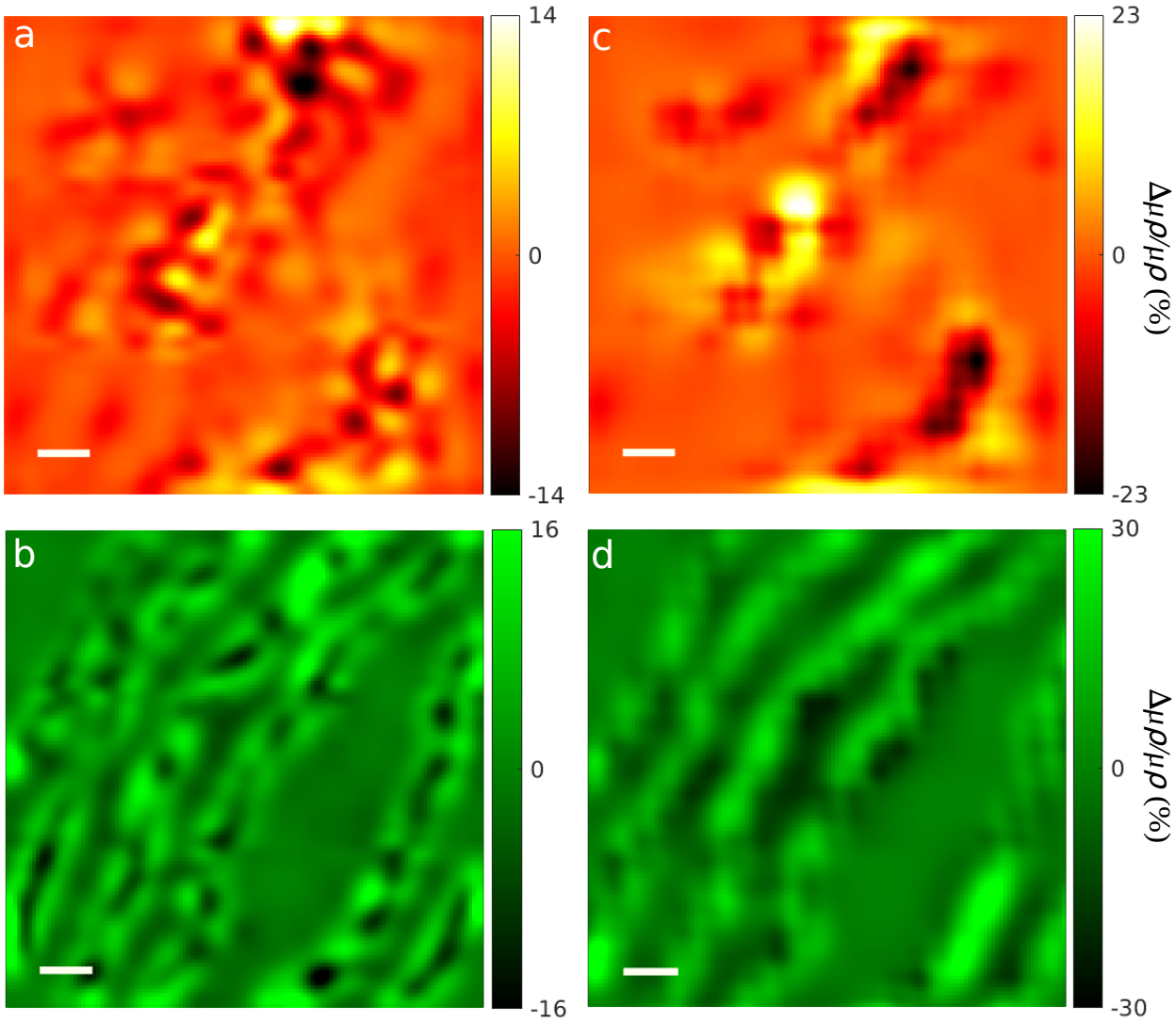

Fig. S9: Difference of the ground truth and learned $\mu\rho$ profiles shown in Fig. 3. (a,b) Relative differences of the ground truths and the learned $\mu\rho$ profiles using photons from all pixels. The relative differences was calculated by subtracting the resulting $\mu\rho$ profiles from the ground truth and dividing by the maximum value in the ground truth. The average of absolute differences are 1.8% and 3.7% in panels a and b, respectively. (c,d) Difference of the ground truths and the learned $\mu\rho$ profiles ignoring photons from pixels with even columns and rows. The average of absolute differences are 2.6% and 7% in panels c and d, respectively. Scale bars are 1 µm.

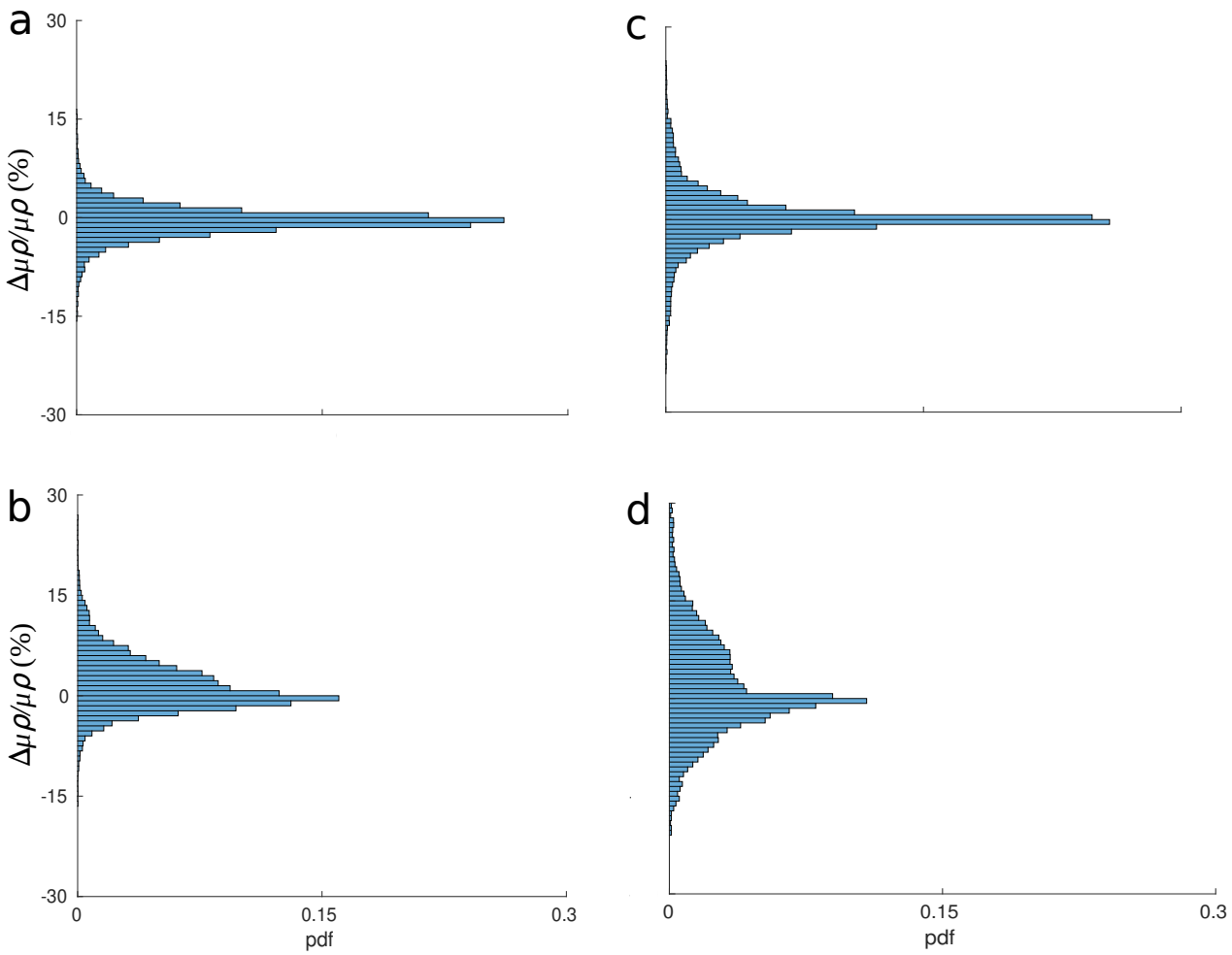

Fig. S10: Histograms of the relative differences shown in Fig. S9

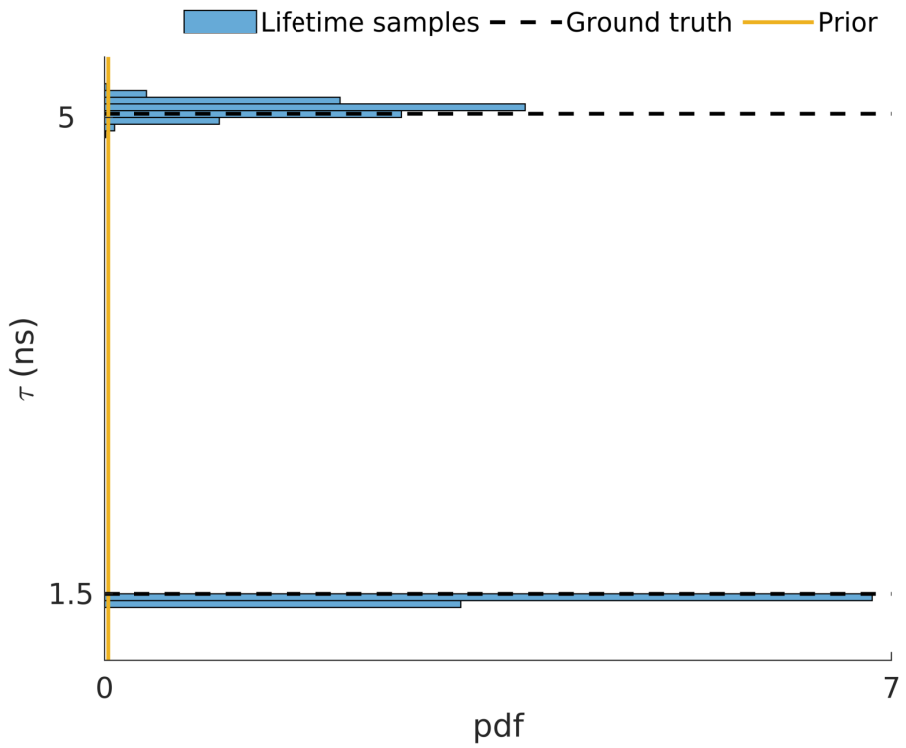

Fig. S11: Histogram of lifetime samples for simulated data with two species and heterogeneous $\mu\rho$ profiles shown in Fig. 3.

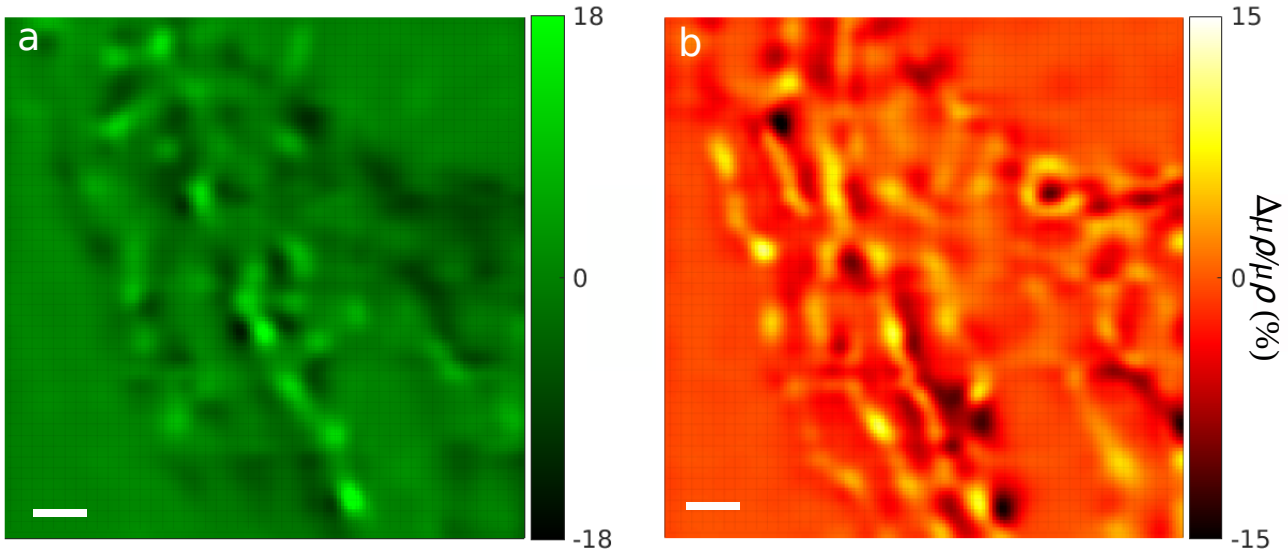

Fig. S12: Differences of the learned and ground truth $\mu\rho$ profiles for *in vivo* data set depicted in Fig. 4. (a) Difference of the $\mu\rho$ profiles of lysosomes using all the photons. The average of the absolute differences is 2.3%. (b) Difference of the $\mu\rho$ profiles of mitochondria using photons from pixels with odd indices. The average of the absolute differences is 3.4%. Scale bars are 1 µm.

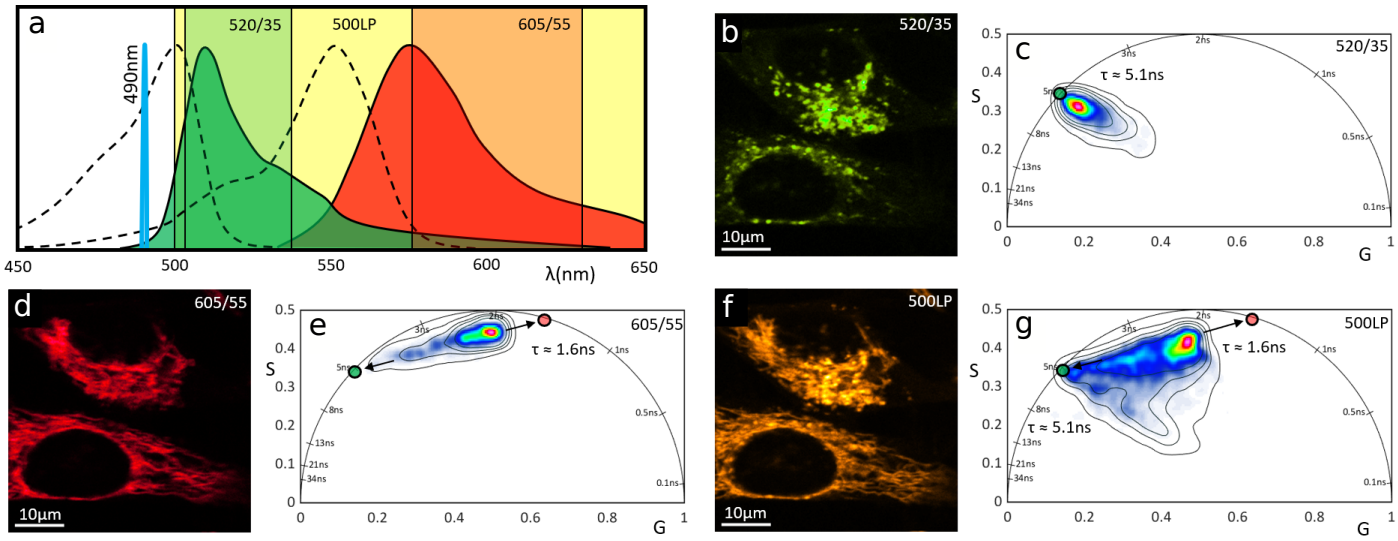

Fig. S13: Phasor approach is used to estimate the *in vivo* lifetimes of the probes. (a) Experimental design showing the excitation laser source, absorption (dotted) and emission (colored) curves of the two probes in the sample. The shaded areas correspond to the three filters used. (b,c) Intensity image and corresponding phasor plot when collecting in the 520/35 band. (d,e) Intensity image and corresponding phasor plot when collecting in the 605/55 band. (f,g) Intensity image and corresponding phasor plot when collecting with the 500LP filter. The fluorescence lifetime for LysoView 488 was obtained by calculating the lifetime corresponding to the extrapolation of the phasor cluster (panel c) to the universal circle, corresponding to 5.1 ns. This was done due to the fact that the phasor distribution appears elongated towards very short lifetimes due to the leakage of laser excitation through the emission filters. The lifetime for TMRM could not be obtained as a pure lifetime on this dataset due to the bleedthrough of LysoView 488 fluorescence in the 605/55 emission filter. For this reason, to estimate the lifetime of TMRM, we considered the extrapolation to the universal circle along the direction of maximum variance of the cluster (panel e), which corresponds to 1.6 ns.

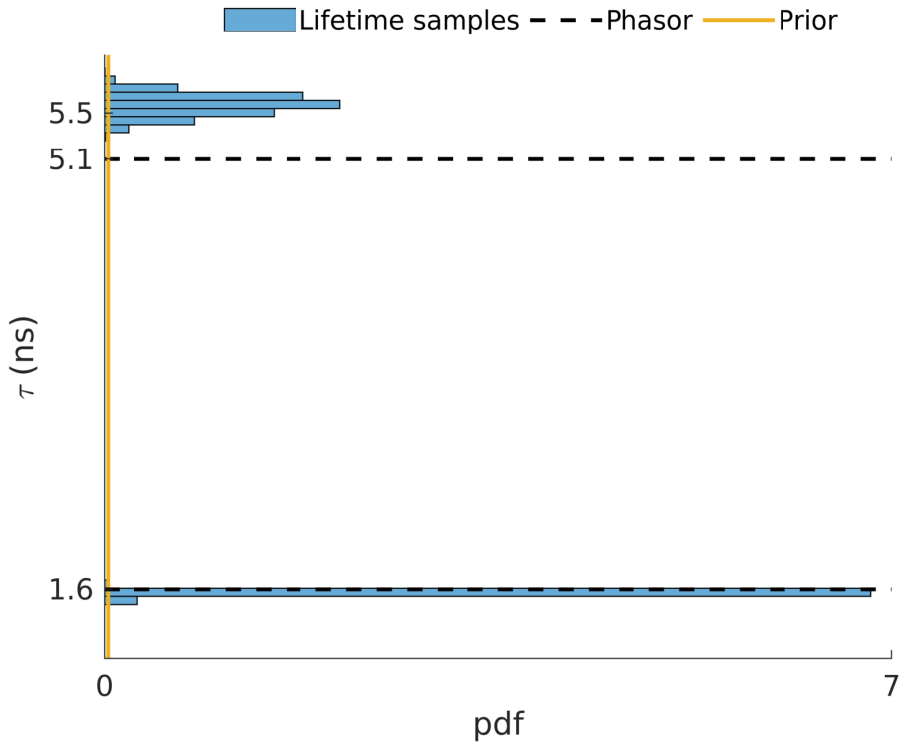

Fig. S14: Histogram of lifetime samples obtained form the mixture of two species *in vivo* data with intricate $\mu\rho$ profiles shown in Fig. 4.

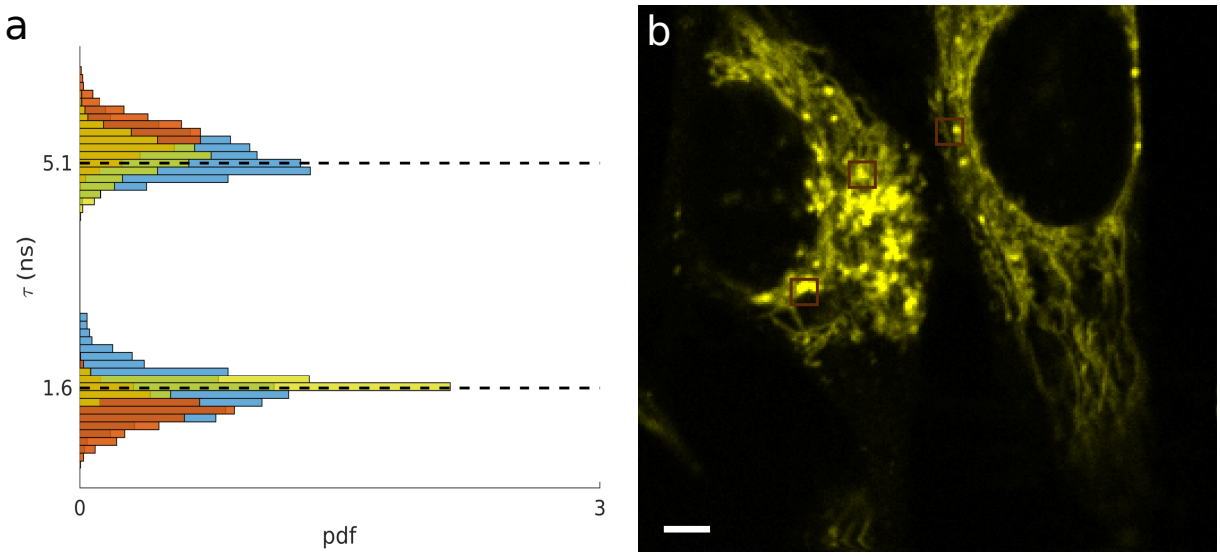

Fig. S15: Lifetimes at different locations. (a) Histograms of lifetime samples obtained form different subregions of the *in vivo* data in Fig. 4i. Lifetime samples with different colors are associated with different selected subregions. The dashed lines show lifetimes obtained by the phasor approach using all of the photons from the entire field of view. (b) The FLIM image data. The sub-regions used to obtain the lifetimes in panel a are included within the brown windows. Scale bar is 4 µm.

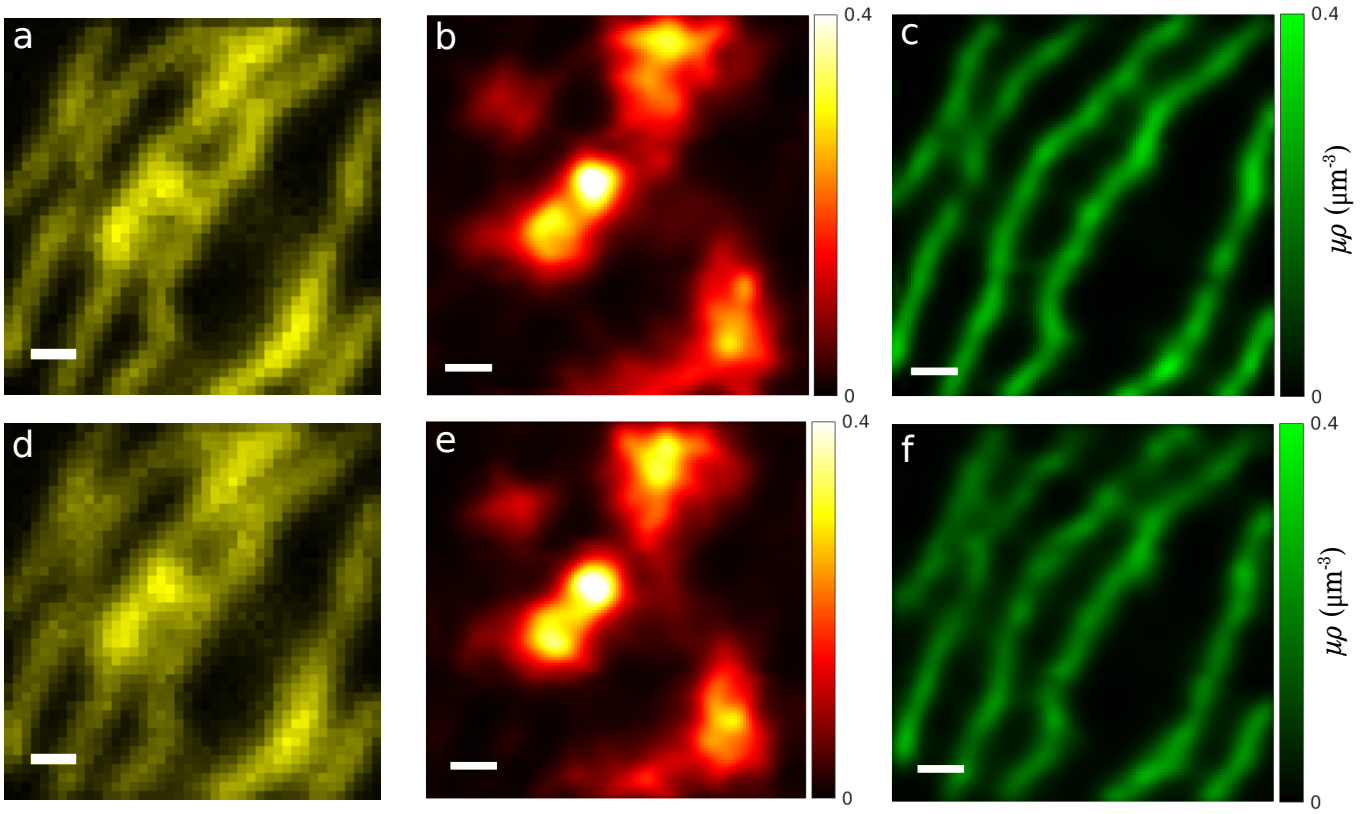

Fig. S 16: Scanning the sample in continuous trajectories versus discrete steps. (a) Data simulated by assuming an illumination spot moving with constant speed, scanning the mixture of simulated profiles shown in Figs. 3a-b. (b-c) Resulting $\mu\rho$ profiles by analyzing the data in (a) assuming an illumination spot moving in discrete steps. (d) Data simulated by assuming an illumination spot moving in discrete steps, scanning the mixture of simulated profiles shown in Figs. 3a-b. (e-f) Resulting $\mu\rho$ profiles by analysing the data in (d) assuming an illumination spot moving in discrete steps. All scale bars are 1 µm.

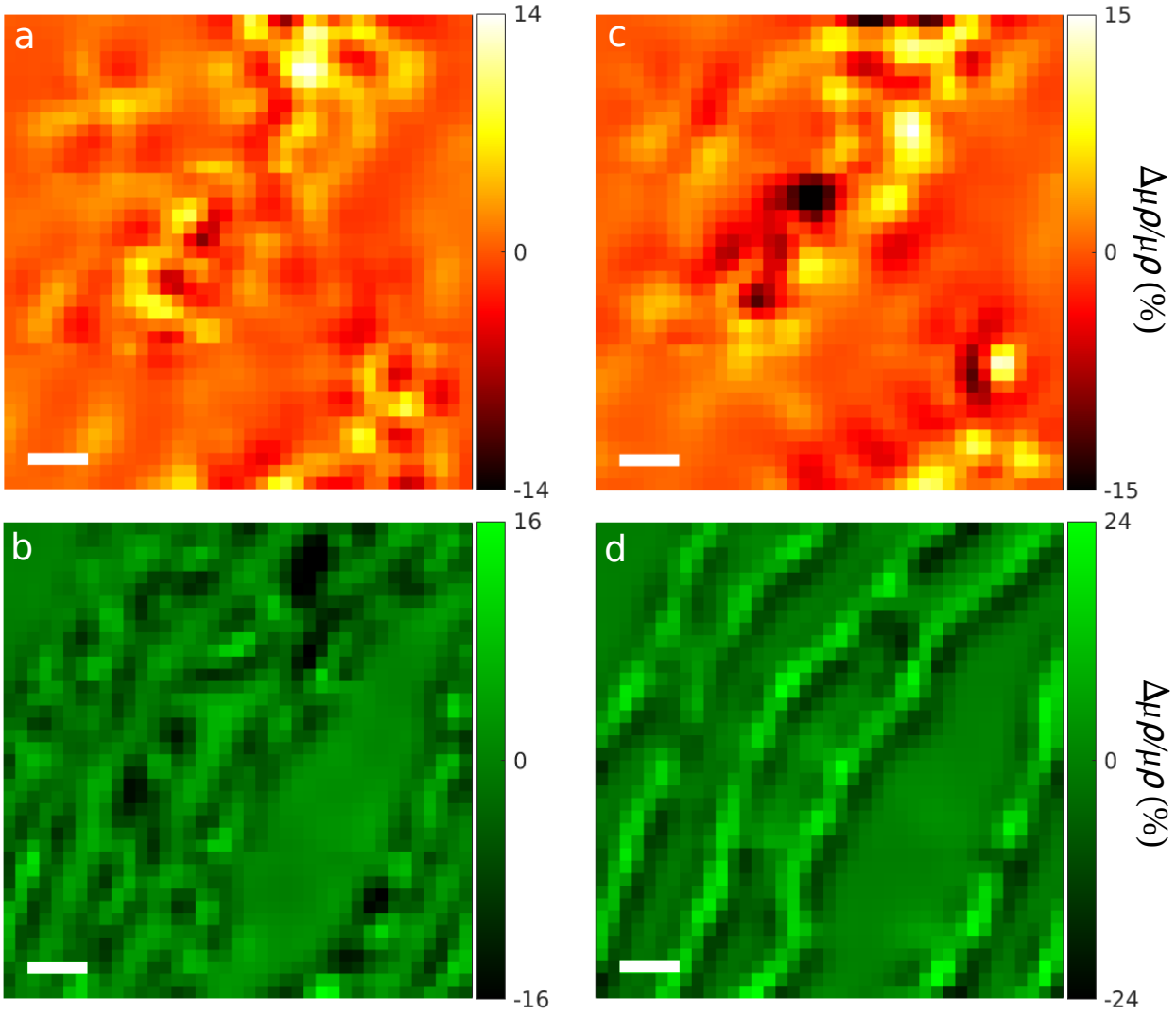

Fig. S17: Difference of the ground truth and the resulting $\mu\rho$ profiles obtained from data sets assuming illumination spots moving with a constant speed, Fig. S16a, and in discrete steps, Fig. S16d. (a-b) Difference of the ground truth, shown in Figs. 3a-b, and the resulting profiles from data simulated assuming an illumination spot moving in constant speed shown in Figs. 16b-c. The average of the absolute differences are 2% and 3.8% in panels a and b, respectively. (c-d) Difference of the ground truth, shown in Figs. 3a-b, and the resulting profiles from data simulated assuming an illumination spot moving in discrete steps shown in Figs. 16e-f. The average of absolute differences are 2.4% and 5.7% in panels c and d, respectively. All scale bars are 1 µm

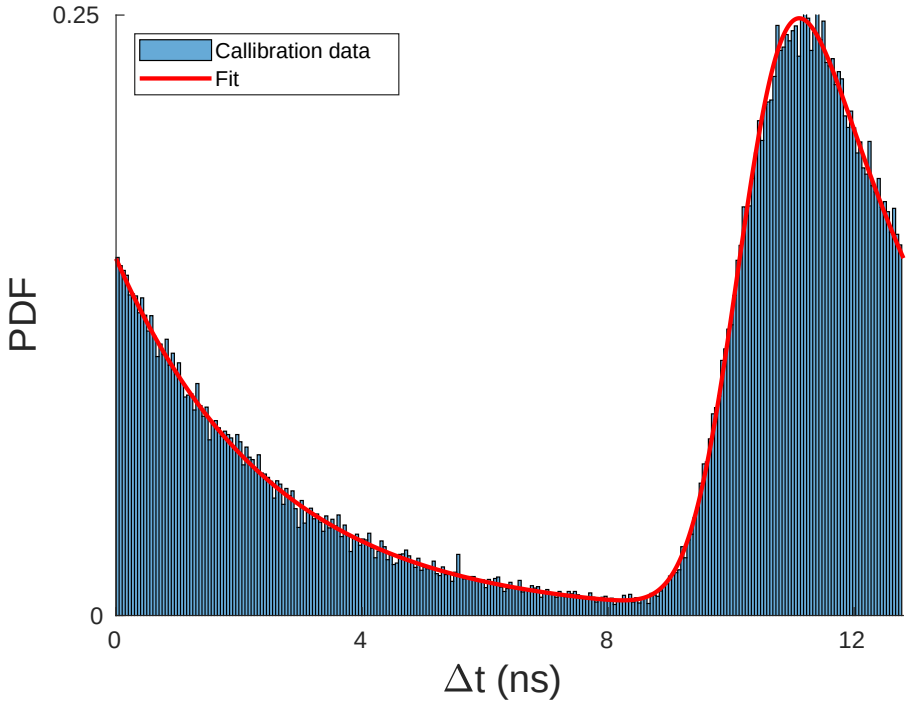

Fig. S18: The Gaussian IRF combined with an exponential decay profile fits (red curve) the calibration data (histogram) with a known lifetime. The calibration data is collected from an *in vitro* experiment using coumarin6 with lifetime of 2.5 ns. The offset (τ_{IRF}), the width (σ_{IRF}) of the Gaussian IRF and the inter-pulse time (T) are, respectively, 10.4, 0.66 and 12.8 ns.

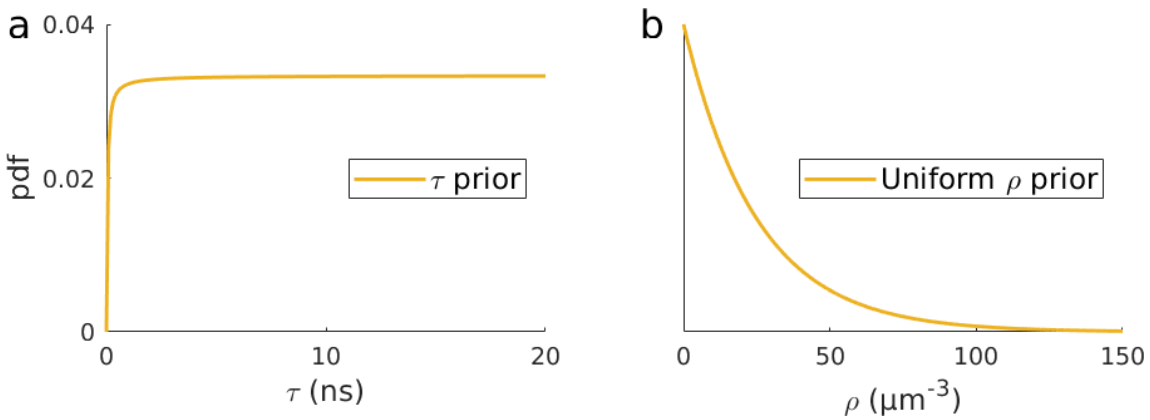

Fig. S19: Priors on (a) lifetimes and (b) uniform $\mu\rho$ profiles. The priors are shown in Figs. S3, S5, S7, S8, S11 and S14 but they appear to be very close to the axis due to scaling. The used priors are very broad and are effectively flat over the interval of the estimated parameters so that they have almost no impact on the final results.

Supplementary Note 1: Methods Description

Supplementary Note 1.1: Photon Arrival Time Likelihood

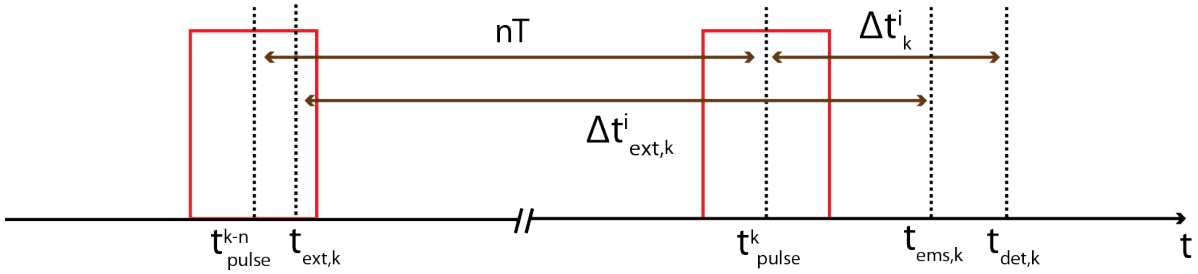

Fig. S20: Sequence of the laser pulses. The red boxes and the dashed lines, respectively, show the laser pulses and times of different events. $t^k_{\text{pulse}}, t^{k-n}_{\text{pulse}}, t_{\text{ext},k}, t_{\text{ems},k}$ and $t_{\text{det},k}$ are, respectively, times of the kth and $k-n$th pulses, the time that a molecule absorbs a photon and gets excited, the time that the excited molecule emits a photon and returns to the ground state, and the time that the emitted photon is detected. n counts the pulses before the kth pulse. T and Δt^i_k are, respectively, the time interval between two consecutive pulses and the arrival time of the photon detected after the kth pulse. Note that the detector reports the photon arrival time with respect to the immediate previous excitation pulse.

In this section, we derive the likelihood of photon arrival times by taking into account details of FLIM experiments. A photon arrival time (Δt^i_k) is recorded with respect to the immediate previous excitation pulse. However, a detected photon might, in general, come from a fluorophore that was excited in a pulse preceding to the immediate previous pulse. The total time from the excitation up to the times that a photon is detected by the detector is the sum of three periods: (1) the time that the excited fluorophore spends in the excited state ($\Delta t^i_{\text{ext},k}$); (2) the difference of the excitation time and the center of the excitation pulse; (3) the time that it takes to record the arrival time after the photon arrives to the detector [1]. The combination of the last two times is called the instrument response time (IRF; denoted by $\Delta t^i_{\text{IRF},k}$). Considering Fig. S20, the relation between these different time intervals and the arrival time can be written as follows

$$\Delta t^i_{\text{ext},k} = t_{\text{ems},k} - t_{\text{ext},k} \tag{S1}$$

$$\Delta t^i_{\text{IRF},k} = \left(t_{\text{ext},k} - t^{k-n}_{\text{pulse}}\right) + (t_{\text{det},k} - t_{\text{ems},k}) \tag{S2}$$

$$\Delta t^i_k = (t_{\text{ems},k} - t_{\text{ext},k}) + \left(t_{\text{ext},k} - t^{k-n}_{\text{pulse}}\right) + (t_{\text{det},k} - t_{\text{ems},k}) - nT \tag{S3}$$

$$\Delta t^i_k = \Delta t^i_{\text{ext},k} + \Delta t^i_{\text{IRF},k} - nT. \tag{S4}$$

where T is the inter-pulses time and n counts the preceding pulses. Since there is a discrete set of multiple options for n (instead of two options), it follows a categorical distribution. Additionally, the time spent in the excited state ($\Delta t^i_{\text{ext},k}$) and the IRF-time ($\Delta t^i_{\text{IRF},k}$), respectively, follow an exponential distribution and an approximate normal distribution, see Fig. S18.

Now that we have introduced different time intervals involved in a photon arrival time, we proceed to construct the photon arrival time likelihood. The arrival time, Δt^i_k, is given by the sum of the parameters $\Delta t^i_{\text{ext},k}, \Delta t^i_{\text{IRF},k}$ and nT, as given by eq. (S4). Thus the likelihood can be calculated by the convolution (shown by $*$) of the distributions of these parameters

$$P\left(\Delta t^i_k | \lambda_{s_k}\right) = \left[P\left(\Delta t^i_{\text{IRF},k} \middle| \tau_{\text{IRF}}, \sigma^2_{\text{IRF}}\right) * P\left(\Delta t^i_{\text{ext},k} | \lambda_{s_k}\right)\right] * P\left(n|N\right) \tag{S5}$$

where

$$P\left(n|N\right) = \text{Categorical}_{0:N}\left([A_0, ..., A_N]\right) \tag{S6}$$

$$P\left(\Delta t^i_{\text{IRF},k} \middle| \tau_{\text{IRF}}, \sigma^2_{\text{IRF}}\right) = \text{Normal}(\Delta t^i_{\text{IRF},k}; \tau_{\text{IRF}}, \sigma^2_{\text{IRF}}) \tag{S7}$$

$$P\left(\Delta t^i_{\text{ext},k} | \lambda_{s_k}\right) = \text{Exponential}(\Delta t^i_{\text{ext},k}; \lambda_{s_k}). \tag{S8}$$

$$\tag{S9}$$

A_n in eq. (S6) is the probability that a detected photon comes from nth pulse preceding the immediate previous pulse, given by

$$A_n = \int_{nT}^{(n+1)T} q(t)dt \tag{S10}$$

where $q(t)$ is introduced in eq. (S12). In general, n can be any number from zero up to the number of pulses at the time of detection of the kth photon. However, since the lifetimes are usually about a few nano-seconds and the time interval between two consecutive pulses is of the order of tens of nano-seconds, it is safe to assume $n \in [0, N]$, where N is an integer cutoff.

In order to derive the photon arrival time likelihood, we must calculate two convolutions. To calculate these convolutions, it is helpful to define an auxiliary parameter $\Delta t^i_{\star k} := \Delta t^i_{\text{ext},k} + \Delta t^i_{\text{IRF},k}$, which is a random variable that follows a distribution $q(\Delta t^i_{\star k})$ given by the following convolution (first convolution in eq. (S5))

$$\begin{aligned} q(\Delta t^i_{\star k}) &= \Bigg[\int_0^{\Delta t^i_{\star k}} \text{Normal}(\Delta t^i_{\star k} - \Delta t^i_{\text{ext},k}; \tau_{\text{IRF}}, \sigma^2_{\text{IRF}}) \\ &\quad \times \text{Exponential}(\Delta t^i_{\text{ext},k}; \lambda_{s_k})\, d\Delta t^i_{\text{ext},k} \Bigg] \qquad \text{(S11)} \\ &= \left[\text{erf}\left(\frac{\tau_{\text{IRF}} + \lambda_{s^i_k}\sigma^2_{\text{IRF}}}{\sigma_{\text{IRF}}\sqrt{2}} \right) - \text{erf}\left(\frac{\tau_{\text{IRF}} - \Delta t i_{\star k} + \lambda_{s^i_k}\sigma^2_{\text{IRF}}}{\sigma_{\text{IRF}}\sqrt{2}} \right) \right] \\ &\quad \times \frac{\lambda_{s^i_k}}{2} \exp\left(\frac{\lambda_{s^i_k}}{2} \left(2(\tau_{\text{IRF}} - \Delta t^i_{\star k}) + \lambda_{s^i_k}\sigma^2_{\text{IRF}} \right) \right) \qquad \text{(S12)} \end{aligned}$$

where erf(.) stands for the error function. We use the approximation

$$\text{erf}\left(\frac{\tau_{\text{IRF}} + \lambda_{s^i_k}\sigma^2_{\text{IRF}}}{\sigma_{\text{IRF}}\sqrt{2}} \right) - \text{erf}\left(\frac{\tau_{\text{IRF}} - \Delta t^i_{\star k} + \lambda_{s^i_k}\sigma^2_{\text{IRF}}}{\sigma_{\text{IRF}}\sqrt{2}} \right) \simeq \text{erfc}\left(\frac{\tau_{\text{IRF}} - \Delta t^i_{\star k} + \lambda_{s^i_k}\sigma^2_{\text{IRF}}}{\sigma_{\text{IRF}}\sqrt{2}} \right) \qquad \text{(S13)}$$

where $\text{erfc}(.) = 1 - \text{erf}(.)$. This approximation is valid as long as $\tau_{\text{IRF}} > 2.5\sigma_{\text{IRF}}$, which holds because there is no negative photon arrival times.

Now we are in a position to find the photon arrival time likelihood, which is given by the convolution of $q(\Delta t^i_{\star k})$ and $P(n|N)$ (second convolution in eq. (S5)). To calculate this convolution, it is helpful to use the substitution $\Delta t^i_{\star k} = \Delta t^i_k + nT$, see eq. (S4), and the likelihood can be obtained as follows

$$\begin{aligned} P\left(\Delta t^i_k | \lambda_{s_k}\right) &= \left[\sum_{n=0}^{N} q(\Delta t^i_k + nT) P(n|N) \right] \\ &= \Bigg[\sum_{n=0}^{N} \text{erfc}\left(\frac{\tau_{\text{IRF}} - \Delta t^i_k - nT + \lambda_{s^i_k}\sigma^2_{\text{IRF}}}{\sigma_{\text{IRF}}\sqrt{2}} \right) \\ &\quad \times \frac{\lambda_{s^i_k}}{2} \exp\left(\frac{\lambda_{s^i_k}}{2} \left(2(\tau_{\text{IRF}} - \Delta t^i_k - nT) + \lambda_{s^i_k}\sigma^2_{\text{IRF}} \right) \right) \Bigg]. \qquad \text{(S14)} \end{aligned}$$

The above expression gives the likelihood of photon arrival times of a photon detected in a non-empty pulse (W^i_k=1). However, there is no detected photon from an empty pulse (W^i_k=0) and hence no arrival time and the likelihood is simply one. Therefore, the arrival time likelihood for both empty and non-empty pulses is given by

$$P\left(\Delta t^i_k | \lambda_{s^i_k}, W^i_k\right) = P\left(\Delta t^i_k | \lambda_{s^i_k}\right)^{W^i_k}. \qquad \text{(S15)}$$

Supplementary Note 1.2: Data Simulation and Analysis

To generate data, it is presumed that a confocal microscope scans the sample sequentially over equally spaced spots, ξ^i. The probabilities of empty pulses (π_0) and probability of exciting a molecule of species m (π_m) are given in the Methods section of the main text. For simulating data, we calculate these probabilities for every individual pulse in the ith pixel using a cylindrical PSF with a Gaussian profile

$$\text{PSF}_{\xi^i}(\vec{X}) = \omega_z \exp\left(-2\frac{(x-\xi_x)^2 + (y-\xi_y)^2}{\omega^2_{xy}} \right) \qquad \text{(S16)}$$

where ω_{xy} and ω_z are respectively the full widths at half maximum (FWHM) in lateral and axial directions. Next, we employ these computed probabilities to generate empty and nonempty pulses, and also the indicator parameters s^i_k for photons from nonempty pulses sampled from the categorical distribution described in the Methods section. The time that

an excited fluorophore spends in the excited state is taken from an exponential distribution with a lifetime corresponding to its indicator s_k^i

$$\Delta t^i_{\text{ext},k} \sim \text{Exponential}(\tau_{s_k^i}). \tag{S17}$$

The IRF times are sampled from a normal distribution with the given offset, τ_{IRF}, and variance, σ^2_{IRF}

$$\Delta t^i_{\text{IRF},k} \sim \text{Normal}(\tau_{\text{IRF}}, \sigma^2_{\text{IRF}}). \tag{S18}$$

Now we can generate the photon arrival times as follows

$$\Delta t^i_k = \Delta t^i_{\text{ext},k} + \Delta t^i_{\text{IRF},k} - T \left\lfloor \frac{\Delta t^i_{\text{ext},k} + \Delta t^i_{\text{IRF},k}}{T} \right\rfloor \tag{S19}$$

where $\lfloor\ \rfloor$ gives the integer part of its content. The data are generated for either a given number of pulses or photons per pixel. A list of parameter values used in simulating different data sets are given in Table S2.

Now that we have described the data generation, we proceed by explaining the data analysis. In order to analyze both experimental and simulated data, we initialized the chains by taking random values from their corresponding priors. The parameters used for the prior and proposal distributions are listed in Table S1.

In order to speed up the algorithm to analyze large regions, namely simulated and experimental data with intricate profiles shown in Figs. 3-4, we split them to smaller subregions where each subregion were overlapped with its neighbors to avoid edge artefacts. Each subregion are then processed independently. In the end, the resulting $\mu\rho$ profiles from each subregion are stitched back together to obtain the entire $\mu\rho$ profile [2].

Supplementary Note 1.3: Characterizing the Instrument Response Function

Instrument response function is due to instrument imperfections, defined in eq. (S2) and Fig. S20. The IRF-time is a random variable and we assume that it follows a normal distribution (which we simply call IRF distribution) with an average of τ_{IRF} and variance of σ^2_{IRF} as given in eq. (S7). The resulting lifetimes and $\mu\rho$ profiles from our method are very sensitive to the parameters of the IRF distribution. Therefore, it is essential to accurately characterize these parameters. However, IRF-times are not directly observed and observations are photon arrival times. Photon arrival times are combinations of the lifetimes and the IRF-times that follow the distribution given by eq. (S14). Therefore, the IRF parameters can be calculated using calibration experiments where a trace of photon arrival times is acquired for a single fluorophore species with a known lifetime; see Fig. S18.

To characterize the IRF given calibration data, we developed a MCMC algorithm to learn the IRF mean (τ_{IRF}) and variance (σ^2_{IRF}) by sampling from the posterior of these parameters. The posterior is proportional to the likelihood given by eq. (S14) and the priors over the parameters of interest, namely τ_{IRF} and σ^2_{IRF}. We take uniform priors for both of these parameters and the resulting posterior is therefor proportional to the likelihood. We used the Metropolis Hasting (MH) algorithm [3, 4] to sample from this posterior where the acceptance ratio is given by

$$A = \frac{P\left(\overline{\overline{\Delta t}}|\tau^{prop}_{\text{IRF}}, \sigma^{prop}_{\text{IRF}}, \lambda\right)}{P\left(\overline{\overline{\Delta t}}|\tau^{old}_{\text{IRF}}, \sigma^{old}_{\text{IRF}}, \lambda\right)} \tag{S20}$$

where $P\left(\overline{\overline{\Delta t}}|\tau^{prop}_{\text{IRF}}, \sigma^{prop}_{\text{IRF}}, \lambda\right)$ is the likelihood of the set of given arrival times. Here, we dropped the the subscript s_k^i of λ since there is only one fluorophore species in the calibration data. After taking adequate samples, the mean τ_{IRF} and variance σ^2_{IRF} is calculated by taking averages of the drawn samples.

Supplementary Note 2: Description of the Inference Framework

Supplementary Note 2.1: Prior Distributions

In the Bayesian framework, priors are included on all the unknown parameters. Here, these parameters are: the inverse of lifetimes λ_m, the indicator parameters s_k^i, the concentrations ρ_m (we learn $\mu\rho$ rather than the concentration ρ), and the mean of GP prior ν_m. In the following, We describe the priors used in our inference framework.

Prior on inverse of lifetimes

Since the inverse of lifetimes are positive, we put a gamma prior on this parameter to restrict them to positive values

$$\lambda_m \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda), \quad m = 1, \ldots, M \tag{S21}$$

where α_λ and β_λ are shape and scale parameters, in turn.

Prior on the indicator parameters

The indicator parameter s_k^i is a discrete variable that accept values from 1 to M (total number of species). It therefore follows a categorical distribution as described in the Methods section

$$s_k^i \sim \text{Categorical}_{1:M}(\{\pi_m\}_m). \tag{S22}$$

This distribution is also conjugate to the likelihood and results in a closed form for the posterior (derived in the next section), which permits direct sampling of this parameter.

Prior on non-uniform $\mu\rho$

Non-uniform $\mu\rho$ profiles are comprised of a set of correlated random variables. Therefore, we put a Gaussian process (GP) prior on these profiles

$$\mu\rho_m \sim \mathbf{GP}(\nu_m, \boldsymbol{K}) \tag{S23}$$

where ν_m and $\boldsymbol{K}$, respectively, are the mean and covariance matrix. The covariance matrix of the GP prior is given by [5–7]

$$\boldsymbol{K}\left(\overline{X}, \overline{X'}\right) = \sigma_{\text{GP}}^2 \exp\left[-\frac{1}{2}\left(\frac{\overline{X} - \overline{X'}}{L}\right)^2\right] \tag{S24}$$

where σ_{GP} and L are positive parameters, and $\overline{X}$ is the set of x and y locations used to calculate correlations.

Prior on the mean of GP prior

We also learn the mean of GP prior ν_m and therefore we must put a hyper-prior on this parameter as well. We use a normal distribution with mean zero as the prior for ν_m

$$\nu_m \sim \text{Normal}(0, \sigma_\nu). \tag{S25}$$

Prior on uniform $\mu\rho$ profiles

Uniform $\mu\rho$ profiles do not vary across space and therefore can be learned as a scalar value. Since $\mu\rho$ is positive, we use a gamma prior which is flexible and also guarantees that $\mu\rho$ does not take negative values

$$\rho_m \sim \text{Gamma}(\alpha_\rho, \beta_\rho) \tag{S26}$$

where α_ρ and β_ρ are shape and scale parameters, respectively.

Summary of all equations

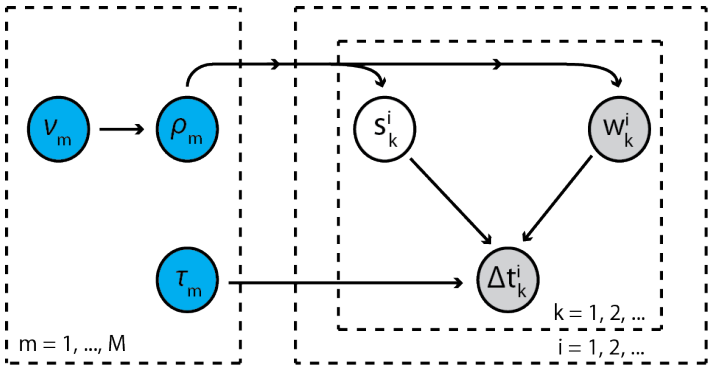

Fig. S21: Graphical model describes the conditional dependence between different parameters in our method. The gray, blue and blank circles, respectively, indicate observations, unknown parameters and intermediate parameters. k, i and m, respectively, count pulses, pixels and species. W_k^i and Δt_k^i reports if a pulse was empty or not and the arrival time of a photon from the kth pulse in ith pixel, in turn. s_k^i is the indicator parameter for a photon from the kth pulse in the ith pixel and can be an integer from 1 to M. ρ_m and τ_m are respectively the concentration and lifetime of the mth species. ν_m is the mean of GP prior on the mth concentration profile.

For concreteness, we summarize the entire set of distributions used in this formulation, including hyper-priors, priors and the likelihood, in the following.

Hyper-priors:

$$\nu_m \sim \text{Normal}(0, \sigma_\nu). \tag{S27}$$

Priors:

$$\lambda_m \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda), \tag{S28}$$
$$\mu\rho_m|\nu_m \sim \mathbf{GP}(\nu_m, \boldsymbol{K}), \tag{S29}$$
$$s_k^i|\{\pi_m^i\}_m \sim \text{Categorical}_{1:M}(\{\pi_m^i\}_m). \tag{S30}$$
$$\tag{S31}$$

Likelihoods:

$$\Delta t_k^i|\lambda_{s_k^i}, W_k^i \sim \left[\sum_{n=0}^{N} \frac{\lambda_{s_k^i}}{2} \exp\left(\frac{\lambda_{s_k^i}}{2}\left(2(\tau_{\text{IRF}} - \Delta t_k^i - nT) + \lambda_{s_k^i}\sigma_{\text{IRF}}^2\right)\right) \times \text{erfc}\left(\frac{\tau_{\text{IRF}} - \Delta t_k^i - nT + \lambda_{s_k^i}\sigma_{\text{IRF}}^2}{\sigma_{\text{IRF}}\sqrt{2}}\right)\right]^{W_k^i}, \tag{S32}$$
$$W_k^i|\pi_0^i \sim \text{Bernoulli}(1 - \pi_0^i). \tag{S33}$$

Supplementary Note 2.2: Description of the Sampling Framework

The joint posterior of all the parameters of interest is shown by $P(\{\mu\rho_m, \nu_m, \tau_m\}_m, \overline{\overline{s}}|\overline{\overline{W}}, \overline{\overline{\Delta t}})$ where the curly bracket with the subscript m indicates the set of the parameters inside the bracket for all the species indexed by m, and s with over-lines represent the set of all indicator parameters for every pulse (indexed by k) and pixel (indexed by i). This joint posterior is proportional to the product of the likelihood and the priors described before. In order to make inference about the system, we use the MCMC technique. Our MCMC technique implements the Gibbs sampling by sequentially taking samples from the posterior of every parameter conditioned on the rest of parameters, which is called either the target distribution or full conditional distribution [8, 9]. Here, the set of parameters of interest are $\{\mu\rho_m\}_m$, $\{\nu_m\}_m$, $\overline{\overline{s}}$ and $\{\lambda_m\}_m$ which are sampled as follows:
1) The set of $\mu\rho_m$ profiles are jointly sampled using the MH algorithm;
2) The set of hyper-parameters, ν_m, are jointly sampled using the MH algorithm;
3) The set of indicator parameters, s_k^i, are directly sampled;
4) The set of lifetimes, τ_m, are jointly sampled using the MH algorithm.
In the following, we introduce the target distribution for each of theses parameters and discuss our sampling scheme in further detail.

Supplementary Note 2.2.1: Sampling λ_m

The target distribution of $\{\lambda_m\}_m$ is given by (Fig. S21)

$$\{\lambda_m\}_m \sim P\left(\{\lambda_m\}_m|\overline{\overline{s}}, \overline{\overline{\Delta t}}, \overline{\overline{W}}, \{\mu\rho_m, \nu_m\}_m\right) \tag{S34}$$
$$\propto P\left(\overline{\overline{\Delta t}}|\overline{\overline{s}}, \overline{\overline{W}}, \{\lambda_m\}_m\right) P\left(\{\lambda_m\}_m\right) \tag{S35}$$
$$= \left[\prod_i \prod_k P\left(\Delta t_k^i|W_k^i, \lambda_{s_k^i}\right)\right] \prod_m P(\lambda_m) \tag{S36}$$
$$= \left[\prod_i \prod_k \left[\sum_{n=0}^{N} \frac{\lambda_{s_k^i}}{2} \exp\left(\frac{\lambda_{s_k^i}}{2}\left(2(\tau_{\text{IRF}} - \Delta t_k^i - nT) + \lambda_{s_k^i}\sigma_{\text{IRF}}^2\right)\right) \times \text{erfc}\left(\frac{\tau_{\text{IRF}} - \Delta t_k^i - nT + \lambda_{s_k^i}\sigma_{\text{IRF}}^2}{\sigma_{\text{IRF}}\sqrt{2}}\right)\right]^{W_k^i}\right] \left[\prod_m \text{Gamma}(\lambda_m; \alpha_\lambda, \beta_\lambda)\right] \tag{S37}$$

where the likelihood of the entire set of photon arrival times is simply the product of the likelihood of every individual photon arrival time, eq. (S14), due to their independence. The above expression does not have a closed form and therefore the MH algorithm is used to sample $\{\lambda_m\}_m$. We take the proposed lifetimes, λ_m^{prop}, from a gamma distribution

$$\lambda_m \sim \text{Gamma}\left(\alpha_\lambda^{\text{prop}}, \frac{\lambda_m^{\text{old}}}{\alpha_\lambda^{\text{prop}}}\right), \;\; m = 1, ..., M \tag{S38}$$

and the MH acceptance ratio is given by

$$A = \frac{P\left(\overline{\overline{\Delta t}}|\overline{\overline{s}}, \overline{\overline{W}}, \{\lambda_m^{\text{prop}}\}_m\right) P\left(\{\lambda_m^{\text{prop}}\}_m\right) \text{Gamma}\left(\{\lambda_m^{\text{old}}\}_m; \alpha_\lambda^{\text{prop}}, \frac{\{\lambda_m^{\text{prop}}\}_m}{\alpha_\lambda^{\text{prop}}}\right)}{P\left(\overline{\overline{\Delta t}}|\overline{\overline{s}}, \overline{\overline{W}}, \{\lambda_m^{\text{old}}\}_m\right) P\left(\{\lambda_m^{\text{old}}\}_m\right) \text{Gamma}\left(\{\lambda_m^{\text{prop}}\}_m; \alpha_\lambda^{\text{prop}}, \frac{\{\lambda_m^{\text{old}}\}_m}{\alpha_\lambda^{\text{prop}}}\right)}. \tag{S39}$$

Supplementary Note 2.2.2: Sampling s_k^i

The target distribution of s_k^i is given by (Fig. S21)

$$s_k^i \sim P\left(s_k^i|\Delta t_k^i, W_k^i, \{\lambda_m, \mu\rho_m, \nu_m\}_m\right) \tag{S40}$$

$$\propto P\left(\Delta t_k^i|W_k^i, \{\lambda_m\}_m, s_k^i\right) P\left(s_k^i|\{\mu\rho_m\}_m\right) \tag{S41}$$

$$\begin{aligned} &= \Bigg[\sum_{n=0}^{N} \frac{\lambda_{s_k^i}}{2} \exp\left(\frac{\lambda_{s_k^i}}{2}\left(2(\tau_{\text{IRF}} - \Delta t_k^i - nT) + \lambda_{s_k^i}\sigma_{\text{IRF}}^2\right)\right) \\ &\quad \times \text{erfc}\left(\frac{\tau_{\text{IRF}} - \Delta t_k^i - nT + \lambda_{s_k^i}\sigma_{\text{IRF}}^2}{\sigma_{\text{IRF}}\sqrt{2}}\right)\Bigg]^{W_k^i} \text{Categorical}(s_k^i; \{\rho_m\}_m) \end{aligned} \tag{S42}$$

$$\begin{aligned} &= \text{Categorical}_{1:M}\Bigg(\pi_1^i\Bigg[\frac{\lambda_1}{2}\sum_n \exp\left(\frac{\lambda_1}{2}\left(2(\tau_{\text{IRF}} - \Delta t_k^i - nT) + \lambda_1\sigma_{\text{IRF}}^2\right)\right) \\ &\qquad \times \text{erfc}\left(\frac{\tau_{\text{IRF}} - \Delta t_k^i - nT + \lambda_1\sigma_{\text{IRF}}^2}{\sigma_{\text{IRF}}\sqrt{2}}\right)\Bigg]^{W_k^i}, ..., \\ &\qquad \pi_M^i\Bigg[\frac{\lambda_M}{2}\sum_n \exp\left[\frac{\lambda_M}{2}\left(2(\tau_{\text{IRF}} - \Delta t_k^i - nT) + \lambda_M\sigma_{\text{IRF}}^2\right)\right] \\ &\qquad \times \text{erfc}\left(\frac{\tau_{\text{IRF}} - \Delta t_k^i - nT + \lambda_M\sigma_{\text{IRF}}^2}{\sigma_{\text{IRF}}\sqrt{2}}\right)\Bigg]^{W_k^i}\Bigg) \end{aligned} \tag{S43}$$

where π_m^i is given in the Methods section. Samples can be directly taken from the above distribution since it has a closed form.

Supplementary Note 2.2.3: Sampling non-uniform $\mu\rho_m$

The target distribution of $\{\mu\rho_m\}_m$ is given by (Fig. S21)

$$\{\mu\rho_m\}_m \sim P\left(\{\mu\rho_m\}_m|\overline{\overline{\Delta t}}, \overline{\overline{s}}, \overline{\overline{W}}, \{\lambda_m, \nu_m\}_m\right) \tag{S44}$$

$$\propto P\left(\overline{\overline{s}}|\{\mu\rho_m\}_m\right) P\left(\overline{\overline{W}}|\{\mu\rho_m\}_m\right) P\left(\{\mu\rho_m\}_m|\{\nu_m\}_m\right) \tag{S45}$$

where

$$\begin{aligned} P\left(\overline{\overline{s}}|\{\mu\rho_m\}_m\right) &= \left[\prod_i \prod_k \pi_{s_k^i}^i\right] \\ P\left(\overline{\overline{W}}\middle|\{\mu\rho_m\}_m\right) &= \left[\prod_i \prod_k \left(1 - \pi_0^i\right)^{W_k^i} \left(\pi_0^i\right)^{1-W_k^i}\right] \end{aligned} \tag{S46}$$

where π_0^i and π_m^i are introduced in the Methods section of the main text. We use GP priors on the species $\mu\rho$ profiles, given in (S23). In the following, we first give a short theoretical background for GPs and then describe our implementation of GPs to learn the $\mu\rho$ profiles.

The objective of GPs is making inference about a generic function, f, which is considered to be comprised of an infinite set of correlated random variables. The correlation between these random variables is described by the correlation function $\boldsymbol{K}$. In practice, dealing with infinite number of random variables is formidable and therefore a finite subset of test points, $\overline{X^t}$, is picked and we make inference about the function values at these set of test points, $\overline{f^t} = f\left(\overline{X^t}\right)$ [5–7]. The function f is usually hidden and observations are noisy measurements, $\overline{y}$, of this function at a set of observation points $\overline{X^o}$, where

$$\begin{aligned} \overline{y} &\sim P\left(\overline{y}|\overline{f^o}\right) \\ \overline{f^o} &= f\left(\overline{X^0}\right). \end{aligned} \tag{S47}$$

A system with Gaussian noise is described by a Gaussian likelihood which is conjugate to the GP prior and results in a closed form posterior that can be directly sampled [5–7]

$$P\left(\overline{y}|\overline{f^o}\right) = \text{Normal}\left(\overline{y}; \overline{f^o}, \boldsymbol{I}\sigma\right) \tag{S48}$$

$$P\left(\overline{f^t}, \overline{f^o}\right) = \text{Normal}\left(\begin{bmatrix} f^o \\ f^t \end{bmatrix}; \begin{bmatrix} \mu_0^o \\ \mu_0^t \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}^{o,o} & \boldsymbol{K}^{o,t} \\ \boldsymbol{K}^{t,o} & \boldsymbol{K}^{t,t} \end{bmatrix}\right) \tag{S49}$$

where $\boldsymbol{K}^{o,o}, \boldsymbol{K}^{t,t}$ and $\boldsymbol{K}^{t,o}$ are, respectively, covariance matrices, which is given by (S24), between subsets $\overline{f^o}$ and itself, $\overline{f^t}$ and itself, and $\overline{f^o}$ and $\overline{f^t}$. μ_0^o and μ_0^t are the mean values at observation points and test points, respectively, which are usually taken to be zero. $\boldsymbol{I}$ denotes the identity matrix with size the same as $\boldsymbol{K}$. Marginalization over $\overline{f^o}$ results in a multi-variate normal distribution with the following mean and covariance matrix

$$\hat{\mu} = \mu_0^t + \boldsymbol{K}^{t,o}\left(\boldsymbol{K}^{o,o} + \sigma\boldsymbol{I}\right)^{-1}\left(\overline{y} - \mu_0^o\right) \tag{S50}$$

$$\hat{\boldsymbol{K}} = \boldsymbol{K}^{t,t} - \boldsymbol{K}^{t,o}\left(\boldsymbol{K}^{o,o} + \sigma\boldsymbol{I}\right)^{-1}\boldsymbol{K}^{o,t}. \tag{S51}$$

The above framework allows implementation of a direct sampling to make inference about f, where f can be learned with a desired resolution by taking a more densely spaced set of test points. Note that the most intensive computational calculation in GPs is computing the matrix inverse and multiplications in the above formula, which is of order of $\mathcal{O}\left(n^3\right)$ for n observations. For big data sets with more than $100K$ data points, these calculations become intractable. One way to ameliorate this difficulty is introducing a set of inducing points $\overline{X_c}$ [10], where the number of inducing points is smaller than the set of data/observation points, and $\overline{f}^c = f\left(\overline{X_c}\right)$. The function values at the test points, $\overline{f^t}$, are deterministically related to the function values at the inducing points, $\overline{f^c}$, by

$$\overline{f^t} = \boldsymbol{K}^{t,c}(\boldsymbol{K}^{c,c})^{-1}\overline{f^c} \tag{S52}$$

where $\boldsymbol{K}^{t,c}$ is the covariance matrix between the sets of inducing points and test points. Once the function values at inducing points are learned, the function values at test points are updated using eq. (S52).

Now that we have discussed GP prior for conjugate likelihoods, we proceed by the more challenging but more common case of non-conjugate likelihoods. Any system with non-Gaussian noise is described by non-conjugate likelihoods to Gaussian processes and thus the posterior does not have a closed form. Therefore, the MH algorithm is employed to learn the function values at the inducing points where the acceptance ratio of a proposed jump is [11]

$$A = \frac{P\left(\overline{y}|\overline{f^c}_{j+1}\right)P\left(\overline{f^c}_{j+1}\right)}{P\left(\overline{y}|\overline{f^c}_{j}\right)P\left(\overline{f^c}_{j}\right)}\frac{Q\left(\overline{f^c}_{j}|\overline{f^c}_{j+1}\right)}{Q\left(\overline{f^c}_{j+1}|\overline{f^c}_{j}\right)} \tag{S53}$$

where j counts jumps and Q is the proposal distribution. A better approximate of the likelihood can be made by using the test points, which have a better resolution than the inducing points, to calculate the likelihood and hence [11]

$$A = \frac{P\left(\overline{y}|\overline{f^t}_{j+1}\right)P\left(\overline{f^c}_{j+1}\right)}{P\left(\overline{y}|\overline{f^t}_{j}\right)P\left(\overline{f^c}_{j}\right)}\frac{Q\left(\overline{f^c}_{j}|\overline{f^c}_{j+1}\right)}{Q\left(\overline{f^c}_{j+1}|\overline{f^c}_{j}\right)}. \tag{S54}$$

While GP priors allow negative values, molecular $\mu\rho$ profiles are positive quantities and thus we use the following substitution to assure that the proposed values are always non-negative

$$\mu\rho_m = \exp(\chi_m) \tag{S55}$$

where χ_m takes both negative and positive values. We learn χ_m employing the GP method and then find $\mu\rho$ using eq. (S55). The likelihood of $\{\chi_m\}_m$, $P\left(\overline{\overline{s}}, \overline{\overline{W}}|\{\chi_m\}_m\right)$, is obtained by replacing $\mu\rho_m$ with $\exp(\chi_m)$ in the calculations. The

resulting likelihood is non-conjugate to the GP prior and the posterior does not have a closed form. Therefore, we employ the MH algorithm described above to make inference about $\{\chi_m\}_m$. We use a multivariate normal distribution centered at zero and a GP with a scalar mean, ν_m, as the proposal and prior distributions, respectively. The covariance matrix of the multivariate normal proposal distribution is taken to be the same as the GP prior. The mean of the GP prior is learned independently. This procedure is described by the following steps:

Step 1

Select the set of inducing points to be the center of pixels and pick the set of test points according to the desired resolution. Initialize χ_m at inducing points to $\overline{\chi^c}_{m,j=1}$.

Step 2

Update χ_m at inducing points for all species by taking samples from the GP proposal distribution using the Cholesky decomposition [12] as follows

$$\overline{\chi^c}_{m,j+1} = \overline{\chi^c}_{m,j} + \boldsymbol{L}\overline{u} \tag{S56}$$

$$\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{K}^{c,c} \tag{S57}$$

where $\overline{u}$ is a set of random numbers drawn from a normal distribution.

Step 3

Update χ_m values at the test points, $\overline{\chi^t}_m$, given the set of inducing points $\chi^c_{m,j+1}$ using (see eq. (S52))

$$\chi^t_{m,j+1} = \boldsymbol{K}^{t,c}(\boldsymbol{K}^{c,c})^{-1}\chi^c_{m,j+1} \tag{S58}$$

Step 4

Accept the proposed $\chi^t_{m,j+1}$ by calculating the acceptance ratio (S54)

$$A = \frac{P\left(\overline{\overline{s}}, \overline{\overline{W}} | \{\overline{\chi^t}_{m,j+1}\}_m\right) \mathbf{GP}\left(\{\overline{\chi^t}_{m,j+1}\}_m; \{\nu_m\}_m, \boldsymbol{K}\right)}{P\left(\overline{\overline{s}}, \overline{\overline{W}} | \{\overline{\chi^t}_{m,j}\}_m\right) \mathbf{GP}\left(\{\overline{\chi^t}_{m,j}\}_m; \{\nu_m\}_m, \boldsymbol{K}\right)} \tag{S59}$$

where $\{\nu_m\}_m$ are the means of GP priors. Note that the proposal distributions are canceled. These steps are repeated in every iteration.

Supplementary Note 2.2.4: sampling ν_m

The target distribution of $\{\nu_m\}_m$ is described by (Fig. S21)

$$\begin{aligned} P\left(\{\nu_m\}_m | \{\chi_m\}_m\right) &\propto P\left(\{\chi_m\}_m | \{\nu_m\}_m\right) P\left(\{\nu_m\}_m\right) \\ &= \mathbf{GP}\left(\{\chi_m\}_m; \{\nu_m\}_m, \boldsymbol{K}\right) \prod_m \text{Normal}\left(\nu_m\, 0, \sigma_\nu^2\right) \end{aligned} \tag{S60}$$

where σ_ν^2 is the variance of the normal prior. The values of $\{\nu_m\}_m$ are modified using the MH method where the acceptance ratio is

$$A = \frac{\mathbf{GP}\left(\{\chi_m\}_m; \{\nu_{m,j+1}\}_m, \boldsymbol{K}\right) \prod_m \text{Normal}\left(\nu_{m,j+1}, 0, \sigma_\nu^2\right)}{\mathbf{GP}\left(\{\chi_m\}_m; \{\nu_{m,j}\}_m, \boldsymbol{K}\right) \prod_m \text{Normal}\left(\nu_{m,j}, 0, \sigma_\nu^2\right)} \tag{S61}$$

and the new values are proposed using a random walk.

Supplementary Note 2.2.5: Sampling uniform $\mu\rho_m$

To learn uniform $\mu\rho$ profiles, we employ a regular MH sampling technique, which is much less computationally intense than GP. We take the proposal distribution and the priors on $\{\mu\rho_m\}_m$ to be gamma distributions

$$\{\mu\rho_m\}_m \sim \text{Gamma}\left(\alpha_\rho^{\text{Prop}}, \frac{\{\mu\rho_m^{\text{old}}\}_m}{\alpha_\rho^{\text{Prop}}}\right), \quad m = 1, ..., M \tag{S62}$$

$$P\left(\{\mu\rho_m\}_m\right) = \text{Gamma}\left(\{\mu\rho_m\}_m; \alpha_\rho, \beta_\rho\right), \qquad m = 1, ..., M. \tag{S63}$$

The acceptance ratio of the proposed $\mu\rho^{\text{prop}}$, is given by

$$A = \frac{P\left(\overline{\overline{s}}, \overline{\overline{W}} | \{\mu\rho_m^{\text{prop}}\}_m\right) \text{Gamma}\left(\{\mu\rho_m^{\text{prop}}\}_m; \alpha_\rho, \beta_\rho\right) \text{Gamma}\left(\{\mu\rho_m^{\text{old}}\}_m; \alpha_\rho^{\text{prop}}, \frac{\{\mu\rho_m^{\text{prop}}\}_m}{\alpha_\rho^{\text{prop}}}\right)}{P\left(\overline{\overline{s}}, \overline{\overline{W}} | \{\mu\rho_m^{\text{old}}\}_m\right) \text{Gamma}\left(\{\mu\rho_m^{\text{old}}\}_m; \alpha_\rho, \beta_\rho\right) \text{Gamma}\left(\{\mu\rho_m^{\text{prop}}\}_m; \alpha_\rho^{\text{prop}}, \frac{\{\mu\rho_m^{\text{old}}\}_m}{\alpha_\rho^{\text{prop}}}\right)} \tag{S64}$$

where the likelihood is given by (S46).

Supplementary Note 3: Additional Results using Smoothly Varying Profiles

To further assess our method's performance–with respect to photon counts, pixel size, lifetimes, and number of lifetime components–we generated data sets with smoothly varying $\mu\rho$ profiles where we also varied: 1) photon counts per pixel; 2) pixel size; 3) number of lifetime components; and 4) lifetimes; Fig. S2-S8.

To examine the performance of our algorithm with respect to pixel size and photon counts, simulated data sets were generated with two $\mu\rho$ profiles that vary between 0.2 and 0.6 µm^{-3} over a region of 32×32 pixels; pixel sizes of 0.2 or 0.4 µm; lifetimes of (0.8 3.6) or (2.5 3.6) ns; and 50, 100 and 200 photons per pixel, Figs. S2-4. To assess the lifetime resolution achieved by our method, a data set with two lifetimes of (2, 2.5) ns, which differ by only 0.5 ns, and 200 photons per pixel was simulated over a region of 10×10 pixels giving 20K photon counts in total, Fig. S5. To evaluate the performance of our algorithm in learning lifetimes similar to the IRF width and exceeding the inter-pulse time, we generated data sets similar to the previous ones except we used lifetimes of (2, 25) and (0.66, 2) ns, Figs. S6-7. We also simulated a data set with three $\mu\rho$ profiles smoothly varying between 0 and 0.2 µm^{-3} over 20×20 pixel; lifetimes of 0.8, 2.2 and 5.2 ns; 100 photon count per pixel, and pixel size of 0.2 µm, Fig. S8.

To process these data sets, we used a 2D grid of test points (see Supplementary Note 2.2) with a grid size of 0.06 µm which gives resolutions of 1/3 and 1/6 of the pixel sizes for the small and large pixels in Fig. S2, respectively. Furthermore, for all data sets we assumed molecules can be excited up to $N = 5$ previous pulses except for the data set with a lifetime larger than the inter-pulse time where we used $N = 20$, Fig. S7. Slices of the resulting $\mu\rho$ profiles from our algorithm are depicted in Fig. S2 where the blue regions, solid pink curves and dashed black curves, respectively, represent the posterior's 95% confidence interval (i.e., credible interval), the median and the ground truth of $\mu\rho$.

We proceed by discussing the performance of our method with respect to lifetimes. The first and second two rows in Figs. S2 and 3, respectively, depict the results for lifetimes of (0.8, 3.6) ns and (2.5, 3.6) ns. For the simpler case with pair of more dissimilar lifetimes, our method learns both $\mu\rho$ and lifetimes, Figs. S2-S3. For the case of more challenging pair of lifetimes, (2.5, 3.6) ns, the ground truths of $\mu\rho$ reside within the 95% confidence regions. However, the size of 95% confidence intervals are larger and the median of the learned $\mu\rho$ slightly deviates from the ground truths by less than 18 percent for data sets with 50 and 100 photons per pixel; Figs. S2c,g-h. The deviation for Fig. S2d is larger than 18 percent. The peaks of histograms for the lifetime samples also deviates from the ground truths by about 0.3 ns, Fig. S3. c-d,g-h. This is due to: 1) the number of available photons per unit of area is small; 2) the lifetimes are very similar and difficult to discriminate. We further assessed the performance of our algorithm in resolving lifetimes that differ by only 0.5 ns (the lifetime difference in this case is 25%) using more photons per pixel over a smaller region, Fig. S5. The algorithm learns both concentration profiles and lifetimes with $\sim 5\%$ error, Fig. S5. We also evaluated the performance of our algorithm in learning lifetimes over a wide range of values, namely lifetimes similar to the IRF width and larger than the inter-pulse interval, Figs. S6-S7. To do so, we generated data sets with two lifetimes and smoothly varying concentration profiles. In both data sets, one of the lifetimes used was 2 ns while the other lifetimes were either 0.66 ns (similar to the IRF) or 25 ns ($\sim$ 2 times larger than the inter-pulse time). Our algorithm learns both $\mu\rho$ profiles and lifetimes for these data sets, see Figs. S6-S7, where the difference of the mean of the sampled lifetimes and the ground truth for the lifetime larger than the inter-pulse time is $\sim 3.7\%$ with a standard deviation of 3.45 ns, Table S3. Although the reported uncertainty is high the reported mean is very close to the ground truth (3.7% difference). Therefore, while our method learns lifetimes over a wide range, the reported uncertainties increase for larger lifetimes, Table S3.

Next, we examine the results with respect to photon counts and pixel sizes. The first, second and third columns in Figs. S2-S3 show results using simulated data sets with 50, 100 and 200 photons per pixel, in turn. Larger photon counts result in smaller 95% confidence regions for the learned $\mu\rho$, Fig. S2, and sharper lifetime histograms, Fig. S3. The first and third rows show results from data sets simulated with a pixel size of 0.2 µm and the pixel size for the data sets used in the second and fourth rows was 0.4 µm. The width of 95% confidence regions increase for cases with the larger pixel size due to the decrease in the number of collected photon per unit of area.

To show an example of the entire simulated smoothly varying $\mu\rho$ profiles, a 3D plot of the generated ground truths with lifetimes of 0.8 and 3.6 ns are presented in Fig. S4a. Fig. S4b shows a 3D plot of the median of sampled $\mu\rho$ profiles for the simulated data set synthesized using concentration profiles in the panel a. The grids show pixels with a size of 0.2 µm.

We also assessed our method's performance employing synthesized data with 3 lifetime components, shown in Fig. S8. Slices of the learned $\mu\rho$ profiles are depicted in Fig. S8a. The ground truths of $\mu\rho$ reside within the 95% confidence area of the learned $\mu\rho$; Fig. S8a. Moreover, our method infers the corresponding lifetimes of the concentration profiles where the histogram of the samples are shown in Fig. S8b.

Supplementary Note 4: Table of Parameter Values used for Data Analysis

Parameter	ω_{xy}	ω_z	σ_{IRF}	τ_{IRF}	T	α_λ	β_λ	α_λ^{prop}	σ_{GP}	L	α_ρ	β_ρ	α_ρ^{prop}	σ_ν
Unit	μm	μm	ns	μm	ns	-	ns	-	μm	μm	-	μm^3	-	μm
Fig. 2	0.15	0.24	0.66	10.4	12.8	1	100	5,000	-	-	1.01	150	20,000	50
Fig. 3	0.3	1	0.66	10.4	12.8	1	100	3,000	1	0.4	-	-	-	50
Fig. 4	0.3	1	0.66	10.4	12.8	1	100	3,000	1	0.4	-	-	-	50
SI Fig. 1	0.15	0.24	0.66	10.4	12.8	1	100	5,000	-	-	1.01	150	20,000	50
SI Fig. 2-4	0.3	1	0.66	10.4	12.8	1	10	2,000	1	3	-	-	-	50
SI Fig. 5	0.3	1	0.66	10.4	12.8	1	10	3,000	1	0.8	-	-	-	50
SI Fig. 6	0.3	1	0.66	10.4	12.8	1	50	5,000	1	0.8	-	-	-	50
SI Fig. 7	0.3	1	0.66	10.4	12.8	1	10	2,000	1	0.8	-	-	-	50
SI Fig. 8	0.3	1	0.66	10.4	12.8	1	10	2,000	1	0.75	-	-	-	50

Table S1: Table of the the parameter values used for the analysis of the data.

Supplementary Note 5: Table of Parameter Values used in Data Simulation

Parameter Unit	ω_{xy} μm	ω_z μm	σ_{IRF} ns	τ_{IRF} μm	T ns	δt ns	μ ns^{-1}	τ ns	ρ μm^{-3}
Fig. 2a-b	0.15	0.24	0.66	10.4	12.8	0.1	10^{-3}	2.5, 3.6	0.5K, 2K
Fig. 3	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	1.5, 5	Profiles
SI Fig. 1	0.15	0.24	0.66	10.4	12.8	0.1	10^{-3}	0.8, 3.6	0.5K, 2K
SI Fig. 2-3a-b,e-f,i-j	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	0.8, 3.6	Profiles
SI Fig. 2-3c-d,g-h,k-l	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	2.5, 3.6	Profiles
SI Fig. 4	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	0.8, 3.6	Profiles
SI Fig. 5	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	2,2.5	Profiles
SI Fig. 6	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	0.66, 2	Profiles
SI Fig. 7	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	2, 25	Profiles
SI Fig. 8	0.3	1	0.66	10.4	12.8	0.1	10^{-3}	0.8, 2.2, 5.2	Profiles

Table S2: Table of parameter values used in generation of the simulated data.

Supplementary Note 6: Table of Results

	$\mu\rho_1$	std	$\mu\rho_2$	std	τ_1	std	τ_2	std	τ_3	std
Fig. 2a-b (500 photon)	0.205	0.024	0.049	0.024	3.79	1.13	2.32	0.48	-	-
Fig. 2a-b (1K photon)	0.207	0.018	0.042	0.017	3.85	1.02	2.59	0.17	-	-
Fig. 2a-b (2K photon)	0.206	0.026	0.057	0.025	3.78	0.75	2.48	0.19	-	-
Fig. 2a-b (5K photon)	0.203	0.017	0.047	0.015	3.56	0.45	2.51	0.08	-	-
Fig. 2c-d (500 photon)	0.188	0.016	0.053	0.013	3.48	0.25	2.40	0.57	-	-
Fig. 2c-d (1K photon)	0.195	0.010	0.049	0.010	3.37	0.20	2.37	0.46	-	-
Fig. 2c-d (2K photon)	0.19	0.010	0.065	0.007	3.54	0.23	2.57	0.39	-	-
Fig. 2c-d (5K photon)	0.197	0.006	0.046	0.005	3.70	0.12	2.60	0.41	-	-
SI Fig. 1 (500 photon)	0.202	0.011	0.054	0.01	3.90	1.02	0.79	0.078	-	-
SI Fig. 1 (1K photon)	0.203	0.015	0.058	0.013	4.05	1.22	0.86	0.08	-	-
SI Fig. 1 (2K photon)	0.197	0.09	0.056	0.008	3.80	0.60	0.83	0.05	-	-
SI Fig. 1 (5K photon)	0.203	0.006	0.054	0.006	3.60	0.36	0.83	0.03	-	-
SI Fig. 3a	-	-	-	-	3.62	0.07	0.82	0.018	-	-
SI Fig. 3b	-	-	-	-	3.66	0.07	0.82	0.017	-	-
SI Fig. 3c	-	-	-	-	3.41	0.07	2.43	0.097	-	-
SI Fig. 3d	-	-	-	-	3.97	0.19	2.65	0.06	-	-
SI Fig. 3e	-	-	-	-	3.65	0.05	0.81	0.012	-	-
SI Fig. 3f	-	-	-	-	3.46	0.05	0.78	0.014	-	-
SI Fig. 3g	-	-	-	-	3.48	0.06	2.42	0.075	-	-
SI Fig. 3h	-	-	-	-	3.48	0.06	2.40	0.072	-	-
SI Fig. 3i	-	-	-	-	3.62	0.03	0.80	0.009	-	-
SI Fig. 3j	-	-	-	-	3.46	0.04	0.77	0.011	-	-
SI Fig. 3k	-	-	-	-	3.69	0.08	2.50	0.029	-	-
SI Fig. 3l	-	-	-	-	3.61	0.06	2.46	0.028	-	-
SI Fig. 5b	-	-	-	-	2.45	0.06	1.89	0.06	-	-
SI Fig. 6b	-	-	-	-	2.00	0.04	0.67	0.026		
SI Fig. 7b	-	-	-	-	1.92	0.06	25.92	3.45	-	-
SI Fig. 8b	-	-	-	-	5.28	0.015	2.12	0.05	0.79	0.012
SI Fig. 10	-	-	-	-	5.11	0.05	1.46	0.012	-	-
SI Fig. 13	-	-	-	-	5.58	0.24	1.57	0.04	-	-

Table S3: Table of mean values and standard deviations of the last fifth of the chains of samples for scalar variables.

Supplementary Note 7: Table of Notations

Indices	Description
m	Index indicating species, it takes a value from $1, ..., M$
k	Index referring to a pulse or a photon detected after this pulse
l_m	Index of a molecule of species m
i	Index of pixels
n	Index counting the preceding pulses to a pulse where a photon is detected
Spatial Coordinates	**Description**
$\vec{X}$	Molecule locations with elements (x, y, z)
ξ^i	Center of the ith confocal spot/pixel with elements $(\xi^i_x, \xi^i_y, \xi^i_z)$
Auxiliary Parameters	**Description**
π^i_m	Probability of detecting a photon from species m in the ith pixel
π^i_*	Probability of excitation of more than a single fluorophore during a pulse in the ith pixel
p_{0m,l_m}	Probability of the lth molecule of type m located at $\vec{X}_{l_m}$ not excited
P_{0m}	Probability of no molecule from species m being excited
$\Delta t^i_{\text{ext},k}$	Time that an excited fluorophore spends in the excited state
$\Delta t^i_{\text{IRF},k}$	IRF time of a photon detected after the kth pulse in the ith pixel
$\Delta t^i_{\star k}$	The arrival time of a photon detected after the kth pulse, $\Delta t^i_{\star k} = \Delta t^i_{\text{ems,k}} + \Delta t^i_{\text{IRF,k}}$
χ_m	Logarithm of concentration of the mth species, $\chi_m = \log(\rho_m)$
Random Variables	**Description**
ρ_m	Absolute concentration of species m
$\{\rho_m\}_m$	Set of all absolute concentrations
τ_{m}	Average lifetime of excited state of species m
$\{\tau_m\}_m$	Set of all lifetimes
λ_m	Inverse of the lifetime, $1/\tau_m$
$\{\lambda_m\}_m$	Set of inverse of lifetimes for all species
ν_m	Mean of GP prior on χ_m
$\{\nu_m\}_m$	Set of means of all GP priors on $\{\chi_m\}_m$
s^i_k	Indicator parameter indicating a photon from pulse k in the ith pixel is associated to what species
$\overline{s}^i$	The collection of the indicator parameters for pulses in the ith pixel
$\overline{\overline{s}}$	The collection of the indicator parameters for all pulses and pixels
Data	**Description**
Δt^i_k	Arrival time of a photon detected from the ith pixel and kth pulse
$\overline{\Delta t}^i$	The collection of photon arrival times from the ith pixel
$\overline{\overline{\Delta t}}$	The collection of all photon arrival times during course of the experiment
W^i_k	Binary label indicating if the kth pulse from the ith pixel was empty or not
$\overline{W}^i$	The collection of binary labels for all the excitation pulses in the ith pixel
$\overline{\overline{W}}$	The collection of binary labels for all the excitation pulses during course of the experiment
Experimental Parameters	**Description**
δt	Average duration of laser pulses
$\hat{\mu}_m$	Excitation rate of species m
μ_m	Average number of excited fluorophores of type m per excitation pulse, $\mu_m = \hat{\mu}_m \delta t$
τ_{IRF}	Offset of instrument response function
σ^2_{IRF}	Variance of instrument response function
T	Time interval between two consecutive laser pulses
$\text{PSF}_{\xi^i}(\vec{X})$	Excitation and detection point spread function (PSF)

Priors Hyper-parameters	Description
α_λ	Shape parameter of gamma prior on inverse of lifetimes
β_λ	Scale parameter of gamma prior on inverse of lifetimes
α_ρ	Shape parameter of gamma prior on uniform concentrations
β_ρ	Scale parameter of gamma prior on uniform concentrations
$\mathbf{K}$	Covariance matrix of GP priors
σ_{GP}	Parameter of covariance matrix of GP priors
L	Parameter of covariance matrix of GP priors
ν_m	Mean of GP prior on χ_m
σ_ν^2	Variance of Normal prior on ν_m
Proposal Parameters	Description
$\alpha_\lambda^{\mathrm{prop}}$	Parameter of gamma proposal distribution for inverse of lifetimes
$\alpha_\rho^{\mathrm{prop}}$	Parameter of gamma proposal distribution for uniform concentrations

Table S4: Table of notations used in the mathematical formulation.

Supplementary Note 8: Table of Abbreviations

Abbreviation	Description
FLIM	Fluorescence Lifetime Imaging Microscopy
FRET	Fluorescence Resonance Energy Transfer
NADH	Nicotinamide Adenine Dinucleotide + H
CW	Continuous Wave
BNP	Bayesian NonParametric
IRF	Instrument Response Function
PSF	Point Spread Function
K	Kilo
GP	Gaussian Process
MCMC	Markov Chain Monte Carlo
MH	Metropolis-Hastings
FWHM	Full Width at Half Maximum

Table S5: Table of abbreviations used in this manuscripts.

Supplementary Note 9: Table of Distribution

Distribution	Notation	Density Function	Mean	Variance
Normal	Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Exponential	Exponential(λ)	$\lambda e^{-\lambda x}$	λ	λ^2
GP	GP$(\nu, \boldsymbol{K})$	$\frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{K})}} e^{-(X-\nu)^T \mathbf{K}^{-1}(X-\nu)}$	ν	$\mathbf{K}$
Gamma	Gamma(α, β)	$\frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-\frac{x}{\beta}}$	$\alpha\beta$	$\alpha\beta^2$
Bernoulli	Bernoulli(q)	$(q-1)\delta_0(x) + q\delta_1(x)$	q	q(1-q)
Poisson	Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ

Table S6: Table of used Distributions in the model formulation.

References

[1] X Michalet, OHW Siegmund, JV Vallerga, P Jelinsky, JE Millaud, and S Weiss. Detectors for single-molecule fluorescence imaging and spectroscopy. *Journal of modern optics*, 54(2-3):239–281, 2007.

[2] Mohamadreza Fazel, Michael J Wester, Hanieh Mazloom-Farsibaf, Marjolein BM Meddens, Alexandra S Eklund, Thomas Schlichthaerle, Florian Schueder, Ralf Jungmann, and Keith A Lidke. Bayesian multiple emitter fitting using reversible jump markov chain monte carlo. *Scientific reports*, 9(1):1–10, 2019.

[3] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[4] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

[5] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[6] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[7] J Shepard Bryan IV, Ioannis Sgouralis, and Steve Pressé. Inferring effective forces for langevin dynamics using gaussian processes. *The journal of chemical physics*, 152(12):124106, 2020.

[8] Sina Jazani, Ioannis Sgouralis, Omer M Shafraz, Marcia Levitus, Sanjeevi Sivasankar, and Steve Pressé. An alternative framework for fluorescence correlation spectroscopy. *Nature communications*, 10(1):1–10, 2019.

[9] Udo Von Toussaint. Bayesian inference in physics. *Reviews of modern physics*, 83(3):943, 2011.

[10] Joaquin Quiñonero-Candela and Carl Edward Rasmussen. A unifying view of sparse approximate gaussian process regression. *Journal of machine learning research*, 6(Dec):1939–1959, 2005.

[11] Michalis K Titsias, Neil Lawrence, and Magnus Rattray. Markov chain monte carlo algorithms for gaussian processes. *Inference and estimation in probabilistic time-series models*, 9, 2008.

[12] James E Gentle. *Computational statistics*, volume 308. Springer, 2009.