Single Photon smFRET. I. Theory and Conceptual Basis

Ayush Saurabh[1,2], Mohamadreza Fazel[1,2], Matthew Safar[1,3], Ioannis Sgouralis[4], and Steve Pressé[1,2,5]

[1]Center for Biological Physics, Arizona State University, Tempe, AZ, USA
[2]Department of Physics, Arizona State University, Tempe, AZ, USA
[3]Department of Mathematics and Statistical Science, Arizona State University, Tempe, AZ, USA
[4]Department of Mathematics, University of Tennessee Knoxville, Knoxville, TN, USA
[5]School of Molecular Sciences, Arizona State University, Tempe, AZ, USA

November 8, 2022

Contents

1 Introduction **3**

2 Forward Model **7**
2.1 Conventions . . . 7
2.2 smFRET Data . . . 7
2.3 Likelihood . . . 8
2.3.1 Absence of observations . . . 12
2.3.2 Introducing observations . . . 12
2.3.3 Recursion formulas . . . 14
2.3.4 Reduced propagators . . . 16
2.3.5 Likelihood for the HMM with second order structure . . . 16
2.3.6 Continuum limit . . . 17
2.3.7 Final likelihood . . . 18
2.3.8 Effect of binning single photon smFRET data . . . 20
2.4 Detection Effects . . . 24

2.4.1 Crosstalk and detection efficiency . . . 25
2.4.2 Effects of detector dead time . . . 28
2.4.3 Adding the detection IRF . . . 30
2.5 Illumination Features . . . 30
2.5.1 Pulsed illumination . . . 30
2.6 Background Emissions . . . 34
2.7 Fluorophore Characteristics: Quantum Yield, Blinking, Photobleaching, and Direct Acceptor Excitation . . . 36
2.8 Synthetic Data Generation . . . 37
2.8.1 Gillespie and detector artefacts . . . 37

3 Inverse Strategy **39**
3.1 Parametric Sampler: BNP-FRET with Fixed Number of System States . . . 41
3.2 Nonparametrics: Predicting the Number of System States . . . 43
3.2.1 Bernoulli process for continuous illumination . . . 43
3.2.2 iHMM methods for pulsed illumination . . . 45
3.3 Likelihood Computation in Practice . . . 47

4 Results **48**
4.1 Parametric Examples . . . 48
4.1.1 Photon budget and excitation rate . . . 48
4.1.2 An example with crosstalk . . . 52
4.1.3 An example with background emissions . . . 53
4.1.4 An example with IRF . . . 54
4.2 A Nonparametric Example . . . 56

5 Discussion **58**

6 Code availability **60**

7 Acknowledgments **60**

Bibliography **61**

Abstract

We present a unified conceptual framework and the associated software package for single molecule Förster Resonance Energy Transfer (smFRET) analysis from single photon arrivals leveraging Bayesian nonparametrics, BNP-FRET. This unified framework addresses the following key physical complexities of a single photon smFRET experiment, including: 1) fluorophore photophysics; 2) continuous time kinetics of the labeled system with large timescale separations between photophysical phenomena such as excited photophysical state lifetimes and events such as transition between system states; 3) unavoidable detector artefacts; 4) background emissions; 5) unknown number of system states; and 6) both continuous and pulsed illumination. These physical features necessarily demand a novel framework that extends beyond existing tools. In particular, the theory naturally brings us to a hidden Markov model (HMM) with a

second order structure and Bayesian nonparametrics (BNP) on account of items 1, 2 and 5 on the list. In the second and third companion manuscripts, we discuss the direct effects of these key complexities on the inference of parameters for continuous and pulsed illumination, respectively.

Why It Matters

smFRET is a widely used technique for studying kinetics of molecular complexes. However, until now, smFRET data analysis methods required specifying *a priori* the dimensionality of the underlying physical model (the exact number of kinetic parameters). Such approaches are inherently limiting given the typically unknown number of physical configurations a molecular complex may assume. The methods presented here eliminate this requirement and allow estimating the physical model itself along with kinetic parameters, while incorporating all sources of noise in the data.

1 Introduction

Förster Resonance Energy Transfer (FRET) has served as a spectroscopic ruler to study motion at the nanometer scale [1–4], and has revealed insight into intra- and intermolecular dynamics of proteins [5–11], nucleic acids [12], and their interactions [13, 14]. In particular, single molecule FRET (smFRET) experiments have been used to determine the pore size and opening mechanism of ion channels sensitive to mechanical stress in the membrane [15], the intermediate stages of protein folding [16, 17], and the chromatin interactions modulated by the helper protein HP1α involved in allowing genetic transcription for tightly packed chromatin [18].

A typical FRET experiment involves labeling molecules of interest with donor and acceptor dyes such that the donor may transfer energy to the acceptor via dipole-dipole interaction when separated by distances of 2-10 nm [19]. This interaction weakens rapidly with increasing separation R and goes as R^{-6} [20, 21].

To induce FRET during experiments, the donor is illuminated by a continuous or pulsating light source for the desired time period or until the dyes photobleach. Upon excitation, the donor may emit a photon itself or transfer its energy nonradiatively to the acceptor which eventually relaxes to emit a photon of a different color [20, 21]. As such, the data collected consists of photon arrival times (for single photon experiments) or, otherwise, brightness values in addition to photon colors collected in different detection channels.

The distance dependence in the rate of energy transfer between donor and acceptor is key in using smFRET as a molecular ruler. Furthermore, this distance dependence directly manifests itself in the form of higher fraction of photons detected in the acceptor channel when the dyes are closer together (as demonstrated in Fig. 1). This fraction is commonly referred to as the FRET efficiency,

$$\varepsilon_{FRET} = \frac{n_A}{n_A + n_D} = \frac{1}{1 + (R/R_0)^6},$$

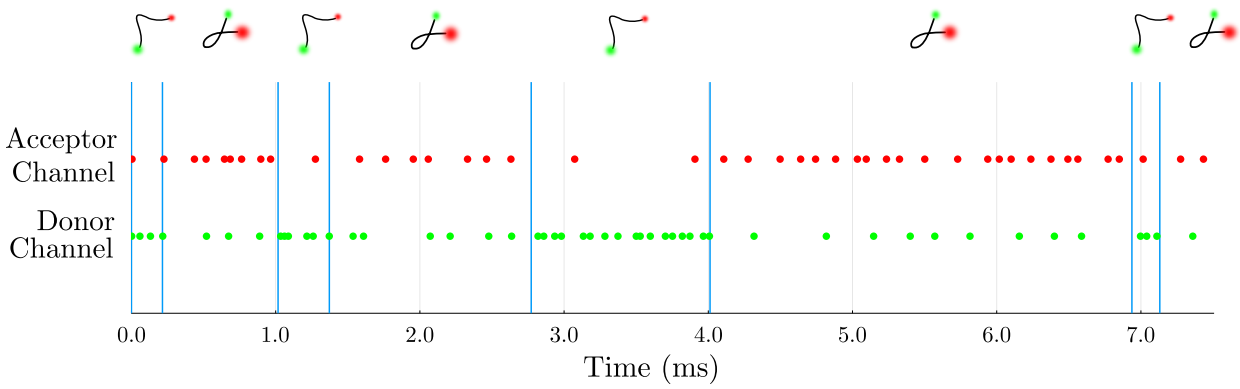

Figure 1: **A cartoon figure illustrating smFRET data**. For the experiments considered here, the kinetics along the reaction coordinate defined along the donor-acceptor distance are monitored using single photon arrival data. In the figure above, photon arrivals are represented by green dots for photons arriving into the donor channel and red dots for photons arriving in the acceptor channel. For the case where donor and acceptor label one molecule, a molecule's transitions between system states (coinciding with conformations) is reflected by the distance between labels measured by variations in detected photon arrival times and colors.

where n_D and n_A are the number of donor and acceptor photons detected in a given time period, respectively. Additionally, R_0 is the characteristic separation that corresponds to a FRET efficiency of 0.5 or 50% of the emitted photons emanating from the acceptor.

Now, the aim of smFRET is to capture on-the-fly changes in donor-acceptor distance. However, this is often confounded by several sources of stochasticity which unavoidably obscure direct interpretation. These include: 1) the stochasticity inherent to photon arrival times; 2) a detector's probabilistic response to an incoming photon [22]; 3) background emissions [2]; and 4) fluorescent labels' stochastic photophysical properties [2]. Taken together, these problems necessarily contribute to uncertainty in the number of distinct system states visited by a labeled system over an experiment's course [23–25].

Here, we delve into greater detail into items 2 and 4. In particular, item 2 pertains to questions of crosstalk, detector efficiency, dead time, dark current, and instrument response function (IRF) introducing uncertainty in excited photophysical state lifetime assessments [22, 26, 27].

Item 4 refers to a collection of effects including limited quantum yield and variable brightness due to blinking of dyes caused by nonradiative pathways [28, 29], photobleaching or permanent deactivation of the dyes [2, 28, 29], spectral overlap between the donor and acceptor dyes which may result in direct excitation of the acceptors or leaking of photons into the incorrect channel [2, 26], or a donor-acceptor pair's relative misalignment or positioning resulting in false signals and inaccurate characterization of the separation between labeled molecules [2, 30].

Though the goal has always remained to analyze the rawest form of data, the reality of these noise properties has traditionally led to the development of approximate binned photon analyses even when data is collected at the level of single photons across two detectors. Binning is either achieved by directly summing photon arrivals over a time period when

using single photon detectors [23, 31] or by integrating intensity over a few pixels when using widefield detectors [32].

While binned data analyses can be used to determine the number and connectivity of system states [33]–by computing average FRET efficiencies over bin time windows and using them in turn to construct FRET efficiency histograms [23, 25, 31, 34–36]–they come at the cost of averaging kinetics that may exist below a time bin not otherwise easily accessible [32, 37, 38]. They also eliminate information afforded by, say, the excited photophysical state lifetime in the case of pulsed illumination.

While histogram analyses are suited to infer static molecular properties, kinetics over binned time traces have also been extracted by supplementing these techniques with a hidden Markov model (HMM) treatment [23, 25, 34–36, 39].

Using HMMs, binned analysis techniques immediately face the difficulty of an unknown number of system states visited. Therefore, they require the number of system states as an input to deduce the putative kinetics between the candidate system states.

What is more, the binned analysis' accuracy is determined by the bin sizes where large bins may result in averaging of the kinetics. Moreover, increasing bin size may lead to estimation of an excess number of system states. This artifact arises when a system appears to artificially spend more time in the system states below the bin size [38]. To address these challenges, we must infer continuous time trajectories below the bin size through, for example, the use of Markov jump processes [32] while retaining a binned, *i.e.*, discrete measurement model.

When single photon data is available we may avoid the binning issues inherent to HMM analysis [32, 40, 41]. Doing so, also allows us to directly leverage the noise properties of detectors for single photon arrivals (*e.g.*, IRF) well calibrated at the single photon level. Moreover, we can now also incorporate information available through photophysical state lifetimes when using pulsed illumination otherwise eliminated in binning data. Incorporating all of this additional information, naturally, comes with added computational cost [37] whose burden a successful method should mitigate.

Often to help reduce computational costs, further approximations on the system kinetics are invoked such as assuming system kinetics to be much slower than FRET label excitation and relaxation rates. This approximation helps decouple photophysical and system (molecular) kinetics [16, 37, 42, 43].

What is more, as they exist, the rigor of direct photon arrival analysis methods are further compromised to help reduce computational cost by treating detector features and background as preprocessing steps [16, 37, 42, 43]. In doing so, simultaneous and self-consistent inference of kinetics and other molecular features becomes unattainable. Finally, all methods, whether relying on the analysis of binned photons or single photon arrival, suffer from the "model selection problem". That is, the problem associated with identifying the number of system states warranted by the data. More precisely, the problem associated with propagating the uncertainty introduced by items 1-4 into a probability over the models (*i.e.*, system states). Existing methods for system state identification only provide partial reprieve.

For example, while FRET histograms identify peaks to intuit the number of system states, these peaks may provide unreliable estimates for a number of reasons: 1) fast transitions between system states may result in a blurring of otherwise distinct peaks [1] or, counter-intuitively, introduce more peaks [25, 38]; 2) system states may differ primarily in kinetics but

not FRET efficiency [40]; 3) detector properties and background may introduce additional features in the histograms.

To address the model selection problem, overfitting penalization criteria (such as the Bayesian information criterion or BIC) [23, 44] or variational Bayesian [24] approaches have been employed.

Often, these model selection methods assume implicit properties of the system. For example, the BIC requires the assumption of weak independence between measurements (*i.e.*, ideally independent identically distributed measurements and thus no Markov kinetics in state space) and a unique likelihood maximum, both of which are violated in smFRET data [24]. Furthermore, BIC and other such methods provide point estimates rather than full probabilities over system states ignoring uncertainty from items 1-4 propagated over models [45].

As such, we need to learn distributions over system states and kinetics warranted by the data and whose breadth is dictated by the sources of uncertainty discussed above. More specifically, to address model selection and build joint distributions over system states and their kinetics, we treat the number of system states as a random variable just as the current community treats smFRET kinetic rates as random variables [25, 40, 41]. Our objective is therefore to obtain distributions over all unknowns (including system states and kinetics) while accounting for items 1-4. Furthermore, this must be achieved in a computationally efficient way avoiding, altogether, the draconian assumptions of existing in single photon analysis methods. In other words, we want to do more (by learning joint distributions over the number of system states alongside everything else) and we want it to cost less.

If we insist on learning distributions over unknowns, then it is convenient to operate within a Bayesian paradigm. Also, if the model (*i.e.*, the number of system states) is unknown, then we must further generalize to the Bayesian nonparametric (BNP) paradigm [25, 41, 46–53]. BNPs directly address the model selection problem concurrently and self-consistently while learning the associated model's parameters and output full distributions over the number of system states and the other parameters.

In this series of three companion manuscripts, we present a complete description of single photon smFRET analysis within the BNP paradigm addressing noise sources discussed above (items 1-4). In addition, we develop specialized computational schemes for both continuous and pulsed illumination for it to "cost less".

Indeed, mitigating computational cost becomes critical especially with the added complexity of working within the BNP paradigm. This, in itself, warrants a detailed treatment of continuous and pulsed illumination analyses in two companion manuscripts.

To complement this theoretical framework, we also provide to the community a suite of programs called BNP-FRET written in the compiled language Julia for high performance. These freely available programs allow for comprehensive analysis of single photon smFRET time traces on immobilized molecules obtained with a wide variety of experimental setups.

In what follows, we will first present a forward model. Next, we will build an inverse strategy to learn full posteriors within the BNP paradigm. Finally, multiple examples are presented by applying the method to simulated data sets across different parameter regimes. Experimental data are treated in the two subsequent companion manuscripts.

2 Forward Model

2.1 Conventions

To be consistent throughout our three part manuscript, we precisely define some terms as follows

1. a macromolecular complex under study is always referred to as a *system*,
2. the configurations through which a system transitions are termed *system states*, typically labeled using σ,
3. FRET dyes undergo quantum mechanical transitions between *photophysical states*, typically labeled using ψ,
4. a system-FRET combination is always referred to as a *composite*,
5. a composite undergoes transitions among its *superstates*, typically labeled using ϕ,
6. all transition rates are typically labeled using λ,
7. the symbol N is generally used to represent the total number of discretized time windows, typically labeled with n, and
8. the symbol w_n is generally used to represent the observations in the n-th time window.

2.2 smFRET Data

Here, we briefly describe the data collected from typical smFRET experiments analyzed by BNP-FRET. In such experiments, donor and acceptor dyes labeling a system can be excited using either continuous illumination, or pulsed illumination where short laser pulses arrive at regular time intervals. Moreover, acceptors can also be excited by nonradiative transfer of energy from an excited donor to a nearby acceptor. Upon relaxation, both donor and acceptor can emit photons collected by single photon detectors. These detectors record the set of photon arrival times and detection channels. We denote the arrival times by

$$\{T_{start}, T_1, T_2, T_3, \ldots, T_K, T_{end}\},$$

and detection channels with

$$\{c_1, c_2, c_3, \ldots, c_K\}$$

for a total number of K photons. In the equations above, T_{start} and T_{end} are experiment's start and end times. Further, we emphasize here that the strategy used to index the detected photons above is independent of the illumination setup used.

Throughout the experiment, photon detection rates from the donor and acceptor dyes vary as the distance between them changes, due to the system kinetics. In cases where the distances form an approximately finite set, we treat the system as exploring a discrete system state space. The acquired FRET traces can then be analyzed to estimate the transition rates

between these system states assuming a known model (*i.e.*, known number of system states). We will lift this assumption of knowing the model *a priori* in Sec. 3.2.

Cases where the system state space is continuous fall outside the scope of the current work and require extensions of Ref. [54] and Ref. [55] currently in progress.

In the following subsections, we present a physical model (forward model) describing the evolution of an immobilized system labeled with a FRET pair. We use this model to derive, step-by-step, the collected data's likelihood given a choice of model parameters. Furthermore, given the mathematical nature of what is to follow, we will accompany major parts of our derivations with a pedagogical example of a molecule labeled with a FRET pair undergoing transitions between just two system states to demonstrate each new concept in example boxes.

2.3 Likelihood

To derive the likelihood, we begin by considering the stochastic evolution of an idealized system, transitioning through a discrete set of total M_σ system states, $\{\sigma_1, \ldots, \sigma_{M_\sigma}\}$, labeled with a FRET pair having M_ψ discrete photophysical states, $\{\psi_1, \ldots, \psi_{M_\psi}\}$, representing the fluorophores in their ground, excited, triplet, blinking, photobleached, or other quantum mechanical states. The combined system-FRET composite now undergoes transitions between $M_\phi = M_\sigma \times M_\psi$ superstates, $\{\phi_1, \ldots, \phi_{M_\phi}\}$, corresponding to all possible ordered pairs (σ_j, ψ_k) of the system and photophysical states. To be precise, we define $\phi_i \equiv (\sigma_j, \psi_k)$, where $i = (j-1)M_\psi + k$.

Assuming Markovianity (memorylessness) of transitions among superstates, the probability of finding the composite in a specific superstate at a given instant evolves according to the master equation [40]

$$\frac{d\boldsymbol{\rho}(t)}{dt} = \boldsymbol{\rho}(t)\mathbf{G}, \tag{1}$$

where the row vector $\boldsymbol{\rho}(t)$ of length M_ϕ has elements coinciding with probabilities for finding the system-FRET composite in a given superstate at time t. More explicitly, defining the photophysical portion of the probability vector $\boldsymbol{\rho}(t)$ corresponding to system state σ_i as

$$\boldsymbol{\rho}_{\sigma_i}(t) = \begin{bmatrix} \rho_{\sigma_i,\psi_1}(t) & \rho_{\sigma_i,\psi_2}(t) & \ldots & \rho_{\sigma_i,\psi_{M_\psi}}(t) \end{bmatrix},$$

we can write $\boldsymbol{\rho}(t)$ as

$$\boldsymbol{\rho}(t) = \begin{bmatrix} \boldsymbol{\rho}_{\sigma_1}(t) & \boldsymbol{\rho}_{\sigma_2}(t) & \ldots & \boldsymbol{\rho}_{\sigma_{M_\sigma}}(t) \end{bmatrix}.$$

Furtheremore, in the master equation above, $\mathbf{G}$ is the generator matrix of size $M_\phi \times M_\phi$ populated by all transition rates $\lambda_{\phi_i \to \phi_j}$ between superstates.

Each diagonal element of the generator matrix $\mathbf{G}$ corresponds to self-transitions and is equal to the negative sum of the remaining transition rates within the corresponding row. That is, $\lambda_{\phi_i \to \phi_i} = -\sum_{j \neq i} \lambda_{\phi_i \to \phi_j}$. This results in zero row-sums, assuring that $\boldsymbol{\rho}(t)$ remains normalized at all times as described later in more detail (see Eq. 5). Furthermore, for simplicity, we assume no simultaneous transitions among system states and photophysical states as such events are rare (although the incorporation of these events in the model

may be accommodated by expanding the superstate space). This assumption results in $\lambda_{(\psi_i,\sigma_j)\to(\psi_l,\sigma_m)} = 0$ for simultaneous $i \neq l$ and $l \neq m$, which allows us to simplify the notation further. That is, $\lambda_{(\psi_i,\sigma_j)\to(\psi_i,\sigma_k)} \equiv \lambda_{\sigma_j\to\sigma_k}$ (for any i) and $\lambda_{(\psi_i,\sigma_j)\to(\psi_k,\sigma_j)} \equiv \lambda_{\sigma_j,\psi_i\to\psi_k}$ (for any j). This leads to the following form for the generator matrix containing blocks of exclusively photophysical and exclusively system transition rates, respectively

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}^{\psi}_{\sigma_1} - \sum\limits_{j\neq 1} \lambda_{\sigma_1\to\sigma_j}\mathbb{I} & \lambda_{\sigma_1\to\sigma_2}\mathbb{I} & \dots & \lambda_{\sigma_1\to\sigma_{M_\sigma}}\mathbb{I} \\ \lambda_{\sigma_2\to\sigma_1}\mathbb{I} & \mathbf{G}^{\psi}_{\sigma_2} - \sum\limits_{j\neq 2} \lambda_{\sigma_2\to\sigma_j}\mathbb{I} & \dots & \lambda_{\sigma_2\to\sigma_{M_\sigma}}\mathbb{I} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{\sigma_{M_\sigma}\to\sigma_1}\mathbb{I} & \lambda_{\sigma_{M_\sigma}\to\sigma_2}\mathbb{I} & \dots & \mathbf{G}^{\psi}_{\sigma_{M_\sigma}} - \sum\limits_{j\neq M_\sigma} \lambda_{\sigma_{M_\sigma}\to\sigma_j}\mathbb{I} \end{bmatrix}, \tag{2}$$

where the matrices on the diagonal $\mathbf{G}^{\psi}_{\sigma_i}$ are the photophysical parts of the generator matrix for a system found in the σ_i system state. Additionally, $\mathbb{I}$ is the identity matrix of size M_ψ.

For later convenience, we also organize the system transition rates $\lambda_{\sigma_i\to\sigma_j}$ in Eq. 2 as a matrix

$$\mathbf{G}_\sigma = \begin{bmatrix} * & \lambda_{\sigma_1\to\sigma_2} & \lambda_{\sigma_1\to\sigma_3} & \dots & \lambda_{\sigma_1\to\sigma_{M_\sigma}} \\ \lambda_{\sigma_2\to\sigma_1} & * & \lambda_{\sigma_2\to\sigma_3} & \dots & \lambda_{\sigma_2\to\sigma_{M_\sigma}} \\ \lambda_{\sigma_3\to\sigma_1} & \lambda_{\sigma_3\to\sigma_2} & * & \dots & \lambda_{\sigma_3\to\sigma_{M_\sigma}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{\sigma_{M_\sigma}\to\sigma_1} & \lambda_{\sigma_{M_\sigma}\to\sigma_2} & \lambda_{\sigma_{M_\sigma}\to\sigma_3} & \dots & * \end{bmatrix}, \tag{3}$$

which we call system generator matrix.

Moreover, the explicit forms of $\mathbf{G}^{\psi}_{\sigma_i}$ in Eq. 2 depend on the photophysical transitions allowed in the model. For instance, if the FRET pair is allowed to go from its ground photophysical state (ψ_1) to the excited donor (ψ_2) or excited acceptor (ψ_3) states only, the matrix is given as

$$\mathbf{G}^{\psi}_{\sigma_i} = \begin{bmatrix} * & \lambda_{\sigma_i,\psi_1\to\psi_2} & \lambda_{\sigma_i,\psi_1\to\psi_3} \\ \lambda_{\sigma_i,\psi_2\to\psi_1} & * & \lambda_{\sigma_i,\psi_2\to\psi_3} \\ \lambda_{\sigma_i,\psi_3\to\psi_1} & 0 & * \end{bmatrix} = \begin{bmatrix} * & \lambda_{ex} & \lambda_{direct} \\ \lambda_d & * & \lambda^{FRET}_{\sigma_i} \\ \lambda_a & 0 & * \end{bmatrix}, \tag{4}$$

where the $*$ along the diagonal represents the negative row-sum of the remaining elements, λ_{ex} is the excitation rate, λ_d and λ_a are the donor and acceptor relaxation rates, respectively, and λ_{direct} is direct excitation of the acceptor by a laser, and $\lambda^{FRET}_{\sigma_i}$ is the donor to acceptor FRET transition rate when the system is in its i-th system state. We note that only FRET transitions depend on the system states (identified by dye-dye separations) and correspond to FRET efficiencies given by

$$\varepsilon^{FRET}_{\sigma_i} = \frac{\lambda^{FRET}_{\sigma_i}}{\lambda^{FRET}_{\sigma_i} + \lambda_d},$$

where the ratio on the right hand side represents the fraction of FRET transitions among all competing transitions out of an excited donor, that is, the fraction of emitted acceptor photons among total emitted photons.

With the generator matrix at hand, we now look for solutions to the master equation of Eq. 1. Due to its linearity, the master equation accommodates the following analytical

solution:

$$\boldsymbol{\rho}(t) = \boldsymbol{\rho}(t_0)\ \mathbf{exp}\left((t-t_0)\mathbf{G}\right) \equiv \boldsymbol{\rho}(t_0)\ \boldsymbol{\Pi}(t-t_0), \tag{5}$$

illustrating how the probability vector $\boldsymbol{\rho}(t)$ arises from the propagation of the initial probability vector at time t_0 by the exponential of the generator matrix (the propagator matrix $\boldsymbol{\Pi}(t-t_0)$). The exponential maps the transition rates $\lambda_{\phi_i\to\phi_j}$ in the generator matrix to their corresponding transition probabilities $\pi_{\phi_i\to\phi_j}$ populating the propagator matrix. The zero row-sums of the generator matrix guarantee that the resulting propagator matrix is stochastic (*i.e.*, has rows of probabilities that sum to unity, $\sum_j \pi_{\phi_i\to\phi_j} = 1$).

Example I: State Space and Generator Matrix

For a molecule undergoing transitions between its two conformations, we have $M_\sigma = 2$ system states given as $\{\sigma_1, \sigma_2\}$. The photophysical states of the FRET pair labeling this molecule are defined according to whether the donor or acceptor are excited. Denoting the ground state by G and excited state by E, we can write all photophysical states of the FRET pair as $\{\psi_1 = (G,G), \psi_2 = (E,G), \psi_3 = (G,E)\}$, where the first element in the ordered pair represents the donor state. Further, here, we assume no simultaneous excitation of the donor and acceptor owing to its rarity.

Next, we construct the superstate space with $M_\phi = 6$ ordered pairs $\{\phi_1 = (\psi_1,\sigma_1), \phi_2 = (\psi_2,\sigma_1), \phi_3 = (\psi_3,\sigma_1), \phi_4 = (\psi_1,\sigma_2), \phi_5 = (\psi_2,\sigma_2), \phi_6 = (\psi_3,\sigma_2)\}$. Finally, the full generator matrix for this setup reads

$$\begin{aligned}
\mathbf{G} &= \begin{bmatrix} \mathbf{G}^{\psi}_{\sigma_1} - \lambda_{\sigma_1\to\sigma_2}\mathbb{I} & \lambda_{\sigma_1\to\sigma_2}\mathbb{I} \\ \lambda_{\sigma_2\to\sigma_1}\mathbb{I} & \mathbf{G}^{\psi}_{\sigma_2} - \lambda_{\sigma_2\to\sigma_1}\mathbb{I} \end{bmatrix} \\
&= \begin{bmatrix} * & \lambda_{ex} & \lambda_{direct} & \lambda_{\sigma_1\to\sigma_2} & 0 & 0 \\ \lambda_d & * & \lambda^{FRET}_{\sigma_1} & 0 & \lambda_{\sigma_1\to\sigma_2} & 0 \\ \lambda_a & 0 & * & 0 & 0 & \lambda_{\sigma_1\to\sigma_2} \\ \lambda_{\sigma_2\to\sigma_1} & 0 & 0 & * & \lambda_{ex} & \lambda_{direct} \\ 0 & \lambda_{\sigma_2\to\sigma_1} & 0 & \lambda_d & * & \lambda^{FRET}_{\sigma_2} \\ 0 & 0 & \lambda_{\sigma_2\to\sigma_1} & \lambda_a & 0 & * \end{bmatrix}.
\end{aligned}$$

Both here, and in similar example boxes that follow, we choose values for rates commonly encountered in experiments [17]. We consider a laser exciting a donor at rate $\lambda_{ex} = 10\,\mathrm{ms}^{-1}$. Next, we suppose that the molecule switches between system states σ_1 and σ_2 at rates $\lambda_{\sigma_1\to\sigma_2} = 2.0\,\mathrm{ms}^{-1}$ and $\lambda_{\sigma_2\to\sigma_1} = 1\,\mathrm{ms}^{-1}$.

Furthermore, assuming typical lifetimes of 3.6 ns and 3.5 ns for the donor and acceptor dyes [17], their relaxation rates are, respectively, $\lambda_d = 1/3.6\ \mathrm{ns}^{-1}$ and $\lambda_a = 1/3.5\ \mathrm{ns}^{-1}$. We also assume that there is no direct excitation of the acceptor and thus $\lambda_{direct} = 0$. Next, we choose FRET efficiencies of 0.2 and 0.9 for the two system states resulting in $\lambda^{FRET}_{\sigma_1} = \lambda_d/4 = 0.06\ \mathrm{ns}^{-1}$ and $\lambda^{FRET}_{\sigma_2} = 9\lambda_d = 2.43\ \mathrm{ns}^{-1}$.

Finally, these values lead to the following generator matrix (in ms^{-1} units)

$$\mathbf{G} = \begin{bmatrix} -12 & 10.0 & 0 & 2.0 & 0.0 & 0.0 \\ 277000 & -347002 & 70000 & 0.0 & 2.0 & 0.0 \\ 285000 & 0.0 & -285002 & 0.0 & 0.0 & 2.0 \\ 1.0 & 0.0 & 0.0 & -11 & 10.0 & 0 \\ 0.0 & 1.0 & 0.0 & 277000 & -2777001 & 2500000 \\ 0.0 & 0.0 & 1.0 & 285000 & 0.0 & -285001 \end{bmatrix}.$$

After describing the generator matrix and deriving the solution to the master equation, we continue by explaining how to incorporate observations into a likelihood.

In the absence of observations, any transition among the set of superstates are unconstrained. However, when monitoring the system using suitable detectors, observations rule out specific transitions at the observation time. For example, ignoring background for now, the detection of a photon from a FRET label identifies a transition from an excited photophysical state to a lower energy photophysical state of that label. On the other hand, no photon detected during a time period indicates the absence of radiative transitions or the failure of detectors to register such transition. Consequently, even periods without photon detections are informative in the presence of a detector. In other words, observations from a single photon smFRET experiment are continuous in that they are defined at every point in time.

Additionally, since smFRET traces report radiative transitions of the FRET labels at photon arrival times, uncertainty remains about the occurrences of unmonitored transitions (*e.g.*, between system states). Put differently, smFRET traces (observations) only partially specify superstates at any given time.

Now, to compute the likelihood for such smFRET traces, we must sum those probabilities over all trajectories across superstates (superstate trajectories) consistent with a given set of observations. Assuming the system ends in superstate ϕ_i at T_{end}, this sum over all possible trajectories can be very generally given by the element of the propagated vector $\boldsymbol{\rho}(T_{end})$ corresponding to superstate ϕ_i. Therefore, a general likelihood may be written as

$$L = p(\phi_i) = [\boldsymbol{\rho}(T_{end})]_i. \tag{6}$$

However, as the final superstate at time T_{end} is usually unknown, we must therefore marginalize (sum) over the final superstate to obtain the following likelihood

$$L = \sum_{i=1}^{M_\phi} p(\phi_i) = \boldsymbol{\rho}(T_{end})\, \boldsymbol{\rho}_{norm}^T, \tag{7}$$

where all elements of the vector $\boldsymbol{\rho}_{norm}$ are set to 1 as a means to sum the probabilities in vector $\boldsymbol{\rho}(T_{end})$. In the following sections, we describe how to obtain concrete forms for these general likelihoods.

2.3.1 Absence of observations

For pedagogical reasons, it is helpful to first look at the trivial case where a system-FRET composite evolves but no observations are made (due to a lack, say, of detection channels). In this case, all allowed superstate trajectories are possible between the start time of the experiment, T_{start}, and end, T_{end}. This is because the superstate cannot be specified or otherwise restricted at any given time by observations previously explained. Consequently, the probability vector $\boldsymbol{\rho}(t)$ remains normalized throughout the experiment as no superstate trajectory is excluded. As such, the likelihood is given by summing over probabilities associated to the entire set of trajectories, that is,

$$L = p((T_1, e_1), \ldots, (T_K, e_K)|\mathbf{G}) = \boldsymbol{\rho}(T_{start})\boldsymbol{\Pi}(T_{end} - T_{start})\, \boldsymbol{\rho}_{norm}^T = \boldsymbol{\rho}(T_{end})\, \boldsymbol{\rho}_{norm}^T = 1, \quad (8)$$

where $\{e_1, \ldots, e_K\}$ are the emission times of all emitted photons, not recorded due to lack of detection channels and thus not appearing on the right hand side of the expression.

In what follows, we describe how the probability vector $\boldsymbol{\rho}(t)$ does not remain normalized as it evolves to $\boldsymbol{\rho}(T_{end})$ when detectors partially collapse knowledge of the occupied superstate during the experiment. This results in a likelihood smaller than one. We do so for the conceptually simpler case of continuous illumination for now.

2.3.2 Introducing observations

To compute the likelihood when single photon detectors are present, we start by defining a measurement model where the observation at a given time is dictated by ongoing transitions and detector features (*e.g.*, crosstalk, detector efficiency). As we will see in more detail later, if we describe the evolution of a system by defining its states at a discrete time points and these states are not directly observed, and thus hidden, then this measurement model adopts the form of a hidden Markov Model (HMM). Here, Markovianity arises when a given hidden state *only* depends on its immediate preceding hidden state. In such HMMs, an observation at a given time is directly derived from the concurrent hidden state.

As an example of an HMM, for binned smFRET traces, an observation is often approximated to depend only on the current hidden state. However, contrary to such a naive HMM, an observation in a single photon setup in a given time period depends on the current superstate and the immediate previous superstate. This naturally enforces a second order structure on the HMM where each observed random variable depends on two superstates as we demonstrate shortly. A similar HMM structure was noted previously to model a fluorophore's photo switching behavior in Ref. [56].

Now, in order to address this observation model, we first divide the experiment's time duration into N windows of equal size, $\epsilon = (T_{end} - T_{start})/N$. We will eventually take the continuum limit $\epsilon \to 0$ to recover the original system as described by the master equation. We also sum over all possible transitions between superstates within each window. These windows are marked by the times (see Fig. 2 (a))

$$\{t_0, t_1, t_2, \ldots, t_N\} \ ,$$

where the n-th window is given by (t_{n-1}, t_n) with $t_0 = T_{start}$ and $t_N = T_{end}$. Corresponding to each time window, we have observations

$$\boldsymbol{w} = \{w_1, w_2, w_3, \ldots, w_N\},$$

where $w_n = \varnothing$ if no photons are detected and $w_n = \{(T_n^{(1)}, c_n^{(1)}), (T_n^{(2)}, c_n^{(2)}), \ldots\}$ otherwise, with the j-th photon in a window being recorded by the channel $c_n^{(j)}$ at time $T_n^{(j)}$. Note here that observations in a time window, being a continuous quantity, allow for multiple photon arrivals or none at all.

As mentioned earlier, each of these observations originate from the evolution of the superstate. Therefore, we define superstates occupied at the beginning of each window as

$$\{a_1, a_2, a_3, \ldots, a_{N-1}, a_N, a_{N+1}\},$$

where a_n is the superstate at the beginning of the n-th time window as shown in Fig. 2(a). The framework described here can be employed to compute the likelihood. However, the second order structure of the HMM leads to complications in these calculations. In the rest of this section, we first illustrate the mentioned complication using a simple example and then describe a solution to this issue.

Example II: Naive Likelihood Computation

Here, we calculate the likelihood for our two state system described earlier. For simplicity alone, we attempt the likelihood calculation for a time period spanning the first two time windows ($N = 2$) in Fig. 2(a). Within this period the system-FRET composite evolves from superstate a_1 to a_3 giving rise to observations $w_{1:2}$. The likelihood for such a setup is typically obtained using a recursive strategy by marginalizing over superstates $a_{1:3}$ (summing over all possible superstate trajectories)

$$\begin{aligned} L = p(w_{1:2}|\mathbf{G}, \boldsymbol{\rho}_{start}) &= \sum_{a_{1:3}} p(w_{1:2}, a_{1:3}|\mathbf{G}, \boldsymbol{\rho}_{start}) \\ &= \sum_{a_{1:3}} p(w_2|a_{1:3}, w_1\mathbf{G}, \boldsymbol{\rho}_{start}) p(w_1, a_{1:3}|\mathbf{G}, \boldsymbol{\rho}_{start}) \\ &= \sum_{a_{1:3}} p(w_2|a_{2:3}, \mathbf{G}) p(w_1|a_{1:2}, \mathbf{G}) p(a_{1:3}|\mathbf{G}, \boldsymbol{\rho}_{start}). \end{aligned}$$

Here, we have applied the chain rule of probabilities in each step. Moreover, in the last step, we have only retained the parameters that are directly connected to the random variable on the left in each term, as shown by arrows in Fig. 2(a).

Now, for our two system state example, a_n can be any of the six superstates $\phi_{1:6}$ ($M_\phi = 6$) given earlier. As such, the sum above contains $M_\phi^{N+1} = 6^3$ terms for such a simple example. For a large number of time windows, computing this sum becomes prohibitively expensive. Therefore, it is common to use a recursive approach to find the likelihood, only requiring $M_\phi^2(N+1)$ operations, as we will describe in the next section. However, due to our HMM's second order structure, the two first terms (involving observations) in the above sum are conditioned on a mutual superstate a_2 which forbids recursive calculations.

After describing the issue in computing the likelihood due to the second order structure of our HMM, we now describe a solution to this problem. As such, to simplify the

likelihood calculation, we temporarily introduce superstates b_n at the end of n-th window separated from superstate a_{n+1} at the beginning of $(n+1)$-th window by a short time τ as shown in Fig. 2(b) during which no observations are recorded (inactive detectors). This procedure allows us to conveniently remove dependency of consecutive observations on a mutual superstate. That is, consecutive observations w_n and w_{n+1} now do not depend on a common superstate a_{n+1}, but rather on separated (a_n, b_n) pairs; see Fig. 2(b). The sequence of superstates now looks like (see Fig. 2b)

$$\{a_1, b_1, a_2, b_2, a_3, b_3, \ldots, a_{N-1}, b_{N-1}, a_N, b_N, a_{N+1}\}, \tag{9}$$

which now permits a recursive strategy for likelihood calculation as described in the next section. Furthermore, we will eventually take the $\tau \to 0$ limit to obtain the likelihood of the original HMM with the second order structure.

2.3.3 Recursion formulas

We now have the means to compute the terminal probability vector $\boldsymbol{\rho}_{end} = \boldsymbol{\rho}(T_{end})$ by evolving the initial vector $\boldsymbol{\rho}_{start} = \boldsymbol{\rho}(T_{start})$. This is most conveniently achieved by recursively marginalizing (summing) over all superstates in Eq. 9 backwards in time, starting from the last superstate a_{N+1} as follows

$$\begin{aligned} L = p(w_{1:N}|\mathbf{G}, \boldsymbol{\rho}_{start}) &= \sum_{a_{N+1}} p(w_{1:N}, a_{N+1}|\mathbf{G}, \boldsymbol{\rho}_{start}) \\ &= \sum_{a_{N+1}} \mathcal{A}_{N+1}(a_{N+1}) = \boldsymbol{\mathcal{A}}_{N+1}\, \boldsymbol{\rho}_{norm}^T, \end{aligned} \tag{10}$$

where $\mathcal{A}_{N+1}(a_{N+1})$ are elements of the vector $\boldsymbol{\mathcal{A}}_{N+1}$ of length M_ϕ, commonly known as a filter [57]. Moving backwards in time, the filter at the beginning of the nth time window, $\mathcal{A}_n(a_{n+1})$, is related to the filter at the end of the nth window, $\mathcal{B}_n(b_n)$, due to Markovianity, as follows

$$\begin{aligned} \mathcal{A}_{n+1}(a_{n+1}) &= p(w_{1:n}, a_{n+1}|\mathbf{G}, \boldsymbol{\rho}_{start}) \\ &= \sum_{b_n} p(a_{n+1}|b_n, \mathbf{G})\, \mathcal{B}_n(b_n), \end{aligned} \tag{11}$$

or in matrix notation as

$$\boldsymbol{\mathcal{A}}_{n+1} = \boldsymbol{\mathcal{B}}_n \tilde{\boldsymbol{\Pi}}_n,$$

where $p(a_{n+1}|b_n)$ are the elements of the transition probability matrix $\tilde{\boldsymbol{\Pi}}_n$ described in the next section. Again due to Markovianity, the filter at the end of the nth window, $\mathcal{B}_n(b_n)$, is related to the filter at the beginning of the same time window, $\mathcal{A}_n(a_n)$, as

$$\begin{aligned} \mathcal{B}_n(b_n) &= p(w_{1:n-1}, b_n|\mathbf{G}, \boldsymbol{\rho}_{start}) \\ &= \sum_{a_n} p(w_n|a_n, b_n, \mathbf{G})\, p(b_n|a_n, \mathbf{G})\, \mathcal{A}_n(a_n), \end{aligned} \tag{12}$$

or in matrix notation as

$$\boldsymbol{\mathcal{B}}_n = \boldsymbol{\mathcal{A}}_n \boldsymbol{\Pi}_n^{(r)},$$

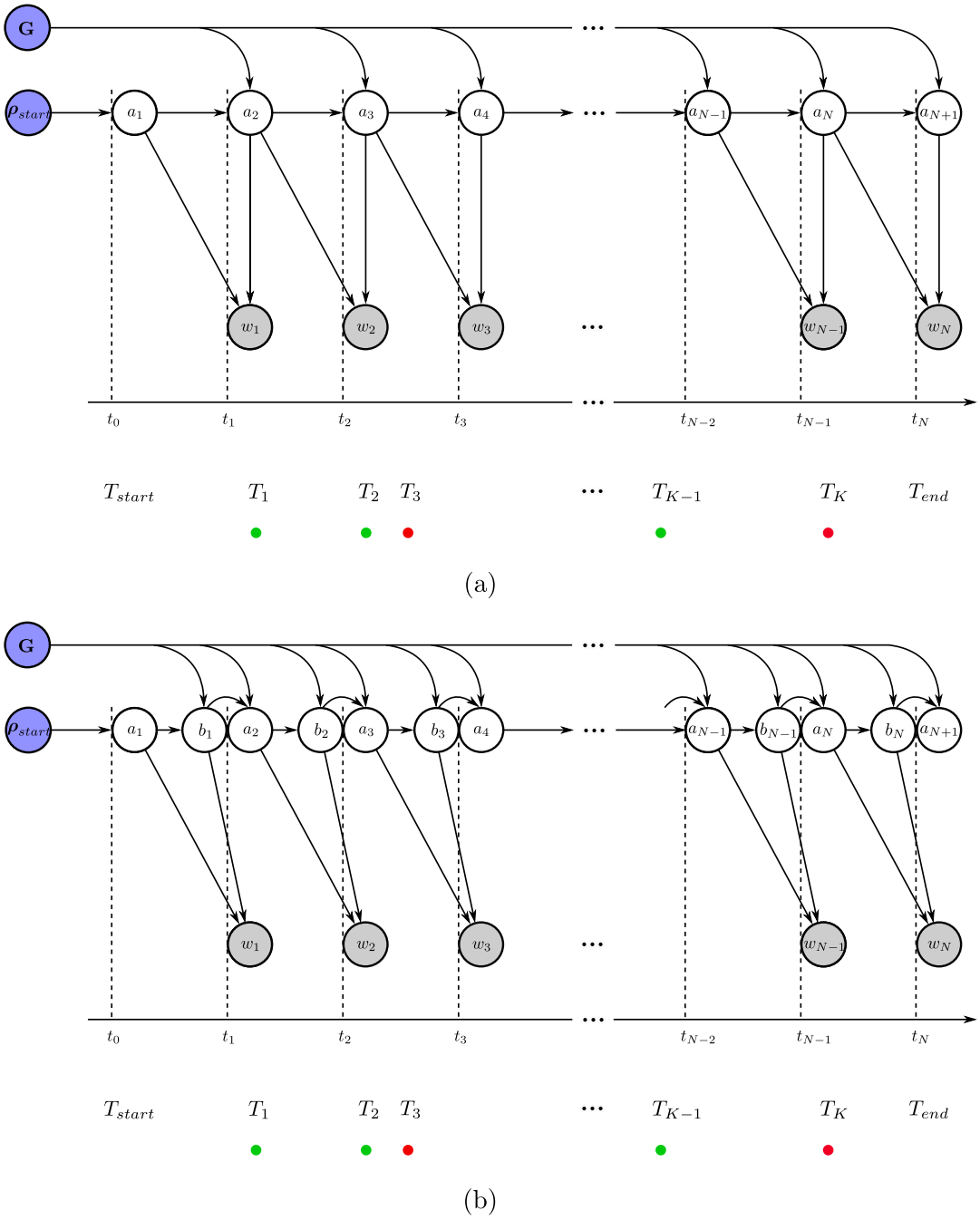

Figure 2: **Graphical models depicting the random variables and parameters involved in the generation of photon arrival data for smFRET experiments**. Circles shaded in blue represent parameters of interest we wish to deduce, namely transition rates and probabilities. The circles shaded in gray correspond to observations. The unshaded circles represent the superstates. The arrows reflect conditional dependence among these variables and colored dots represent photon arrivals. Going from panel (a) to (b), we convert the original HMM with a second order structure to a naive HMM where each observation only depends on one state.

where the terms $p(w_n|a_n, b_n, \mathbf{G})\, p(b_n|a_n, \mathbf{G})$ populate the transition probability matrix $\mathbf{\Pi}_n^{(r)}$ described in the next section. Here, we use the superscript (r) to denote that elements of this matrix include observation probabilities. We note here that the last filter in the recursion formula, $\boldsymbol{\mathcal{A}}_1$, is equal to starting probability vector $\boldsymbol{\rho}_{start}$ itself.

2.3.4 Reduced propagators

To derive the different terms in the recursive filter formulas, we first note that the transition probabilities $p(a_n|b_{n-1}, \mathbf{G})$ and $p(b_n|a_n, \mathbf{G})$ do not involve observations. As such, we can use the full propagator as follows

$$p(b_n|a_n, \mathbf{G}) = (\mathbf{\Pi})_{a_n \to b_n} = (\mathbf{exp}((\epsilon - \tau)\mathbf{G}))_{a_n \to b_n}$$

and

$$p(a_n|b_{n-1}, \mathbf{G}) = (\tilde{\mathbf{\Pi}})_{b_{n-1} \to a_n} = (\mathbf{exp}(\tau \mathbf{G}))_{b_{n-1} \to a_n}\,,$$

respectively. On the other hand, the term $p(w_n|a_n, b_n, \mathbf{G})$ includes observations which results in modification to the propagator by ruling out a subset of transitions. For instance, observation of a photon momentarily eliminates all nonradiative transitions. The modifications now required can be structured into a matrix $\mathbf{D}_n$ of the same size as the propagator with elements $(\mathbf{D}_n)_{a_n \to b_n} = p(w_n|a_n, b_n, \mathbf{G})$. We term all such matrices *detection* matrices. The product $p(w_n|a_n, b_n, \mathbf{G})\; p(b_n|a_n, \mathbf{G})$ in Eq. 12 can now be written as

$$(\mathbf{\Pi}_n^{(r)})_{a_n \to b_n} = (\mathbf{\Pi})_{a_n \to b_n} \times (\mathbf{D}_n)_{a_n \to b_n}\,,$$

relating the modified propagator (termed reduced propagator and distinguished by the superscript (r) hereafter) $(\mathbf{\Pi}_n^{(r)})_{a_n \to b_n}$ in the presence of observations to the full propagator (no observations). Plugging in the matrices introduced above into the recursive filter formulas (Eqs. 11-12), we obtain in matrix notation

$$\begin{aligned} \boldsymbol{\mathcal{A}}_{n+1} &= \boldsymbol{\mathcal{B}}_n \tilde{\mathbf{\Pi}} = \boldsymbol{\mathcal{B}}_n\, \mathbf{exp}((\epsilon - \tau)\mathbf{G}) \\ \boldsymbol{\mathcal{B}}_n &= \boldsymbol{\mathcal{A}}_n \mathbf{\Pi}_n^{(r)} = \boldsymbol{\mathcal{A}}_n\, \left(\mathbf{exp}(\epsilon\, \mathbf{G}) \odot \mathbf{D}_n\right), \end{aligned} \tag{13}$$

where the symbol $\odot$ represents element-by-element product of matrices. Here, however, the detection matrices cannot yet be computed analytically as the observations w_n allow for an arbitrary number of transitions within the finite time window (t_{n-1}, t_n). However, they become manageable in the limit that the time windows become vanishingly small, as we will demonstrate later.

2.3.5 Likelihood for the HMM with second order structure

Now, inserting the matrix expressions for filters of Eq. 13 into the recursive formula likelihood Eq. 10, we arrive at

$$\begin{aligned} L &= \sum_{a_{N+1}} \boldsymbol{\rho}_{start}\; \mathbf{\Pi}_1^{(r)}\, \tilde{\mathbf{\Pi}}\, \mathbf{\Pi}_2^{(r)}\, \tilde{\mathbf{\Pi}}\, \mathbf{\Pi}_3^{(r)}\, \tilde{\mathbf{\Pi}} \ldots \mathbf{\Pi}_{N-1}^{(r)}\, \tilde{\mathbf{\Pi}}\, \mathbf{\Pi}_N^{(r)} \\ &= \boldsymbol{\rho}_{start}\; \mathbf{\Pi}_1^{(r)}\, \tilde{\mathbf{\Pi}}\, \mathbf{\Pi}_2^{(r)}\, \tilde{\mathbf{\Pi}}\, \mathbf{\Pi}_3^{(r)}\, \tilde{\mathbf{\Pi}} \ldots \mathbf{\Pi}_{N-1}^{(r)}\, \tilde{\mathbf{\Pi}}\, \mathbf{\Pi}_N^{(r)}\; \boldsymbol{\rho}_{norm}^T = \boldsymbol{\rho}(T_{end})\boldsymbol{\rho}_{norm}^T, \end{aligned}$$

where, in the second step, we added a row vector of ones, $\boldsymbol{\rho}_{norm}$ at the end to sum over all elements. Here, the superscript T denotes matrix transpose. As we now see, the structure of the likelihood above amounts to propagation of the initial probability vector $\boldsymbol{\rho}_{start}$ to the final probability vector $\boldsymbol{\rho}(T_{end})$ via multiple propagators corresponding to N time windows.

Now, under the limit $\tau \to 0$, we have

$$\tilde{\mathbf{\Pi}} = \lim_{\tau \to 0} \mathbf{exp}(\tau \mathbf{G}) = \mathbb{I},$$
$$\mathbf{\Pi} = \lim_{\tau \to 0} \mathbf{exp}((\epsilon - \tau)\mathbf{G}) = \mathbf{exp}(\epsilon \mathbf{G}),$$

where $\mathbb{I}$ is the identity matrix. In this limit, we recover the likelihood for the HMM with a second order structure as

$$L = \boldsymbol{\rho}_{start}\, \mathbf{\Pi}_1^{(r)}\, \mathbf{\Pi}_2^{(r)}\, \mathbf{\Pi}_3^{(r)}\, \ldots\, \mathbf{\Pi}_{N-1}^{(r)}\, \mathbf{\Pi}_N^{(r)} \boldsymbol{\rho}_{norm}^T\,. \tag{14}$$

We note here that the final probability vector $\boldsymbol{\rho}(T_{end})$ is not normalized to one upon propagation due to the presence of reduced propagators corresponding to observations. More precisely, the reduced propagators restrict the superstates evolution to only a subset of trajectories over a time window ϵ in agreement with the observation over this window. This, in turn, results in a probability vector whose elements sum to less than one. That is,

$$\boldsymbol{\rho}_{start}\, \mathbf{\Pi}_1^{(r)}\, \mathbf{\Pi}_2^{(r)}\, \mathbf{\Pi}_3^{(r)}\, \ldots\, \mathbf{\Pi}_{N-1}^{(r)}\, \mathbf{\Pi}_N^{(r)} \boldsymbol{\rho}_{norm}^T = \boldsymbol{\rho}(T_{end})\boldsymbol{\rho}_{norm} < 1\,. \tag{15}$$

2.3.6 Continuum limit

Up until now, the finite size of the time window ϵ allowed for an arbitrary number of transitions per time window (t_{n-1}, t_n), which hinders the computation of an exact form for the detection matrices. Here, we take the continuum limit, as the time windows become vanishingly small (that is $\epsilon \to 0$ as $N \to \infty$). Thus, no more than one transition is permitted per window. This allows us to fully specify the detection matrices $\mathbf{D}_n$.

To derive the detection matrices, we first assume ideal detectors with 100% efficiency and include detector effects in the subsequent sections (see Sec. 2.4). In such cases, the absence of photon detections during a time window, while detectors are active, indicates that only nonradiative transitions took place. Thus, only nonradiative transitions have nonzero probabilities in the detection matrices. As such, for evolution from superstate a_n to b_n, the elements of the nonradiative detection matrix, $\mathbf{D}^{non}$, are given by

$$(\mathbf{D}^{non})_{a_n \to b_n} = \begin{cases} 1 & \text{Nonradiative transitions} \\ 0 & \text{Radiative transitions} \end{cases}. \tag{16}$$

On the other hand, when the k-th photon is recorded in a time window, only elements corresponding to radiative transitions are nonzero in the detection matrix denoted by $\mathbf{D}_k^{rad}$ as

$$(\mathbf{D}_k^{rad})_{a_n \to b_n} = \begin{cases} 0 & \text{All transitions except for the } k\text{-th photon emission} \\ 1 & k\text{-th photon emission} \end{cases}. \tag{17}$$

Here, we note that the radiative detection matrices have zeros along their diagonals, since self-transitions are nonradiative.

We can now define the reduced propagators corresponding to the nonradiative and radiative detection matrices, $\mathbf{D}^{non}$ and $\mathbf{D}_k^{rad}$, using the Taylor approximation $\lim_{\epsilon\to 0}\mathbf{\Pi}_n = \mathbb{I} + \epsilon\mathbf{G} + \mathcal{O}(\epsilon^2)$ as

$$\mathbf{\Pi}^{(r)non} = (\mathbb{I} + \epsilon\mathbf{G} + \mathcal{O}(\epsilon^2)) \odot \mathbf{D}^{non} = \mathbf{exp}(\epsilon\mathbf{G}^{non}) + \mathcal{O}(\epsilon^2)\,, \text{ and} \tag{18}$$

$$\mathbf{\Pi}_k^{(r)rad} = (\mathbb{I} + \epsilon\mathbf{G} + \mathcal{O}(\epsilon^2)) \odot \mathbf{D}_k^{rad} = \epsilon\mathbf{G}_k^{rad} + \mathcal{O}(\epsilon^2). \tag{19}$$

In the equations above, $\mathbf{G}^{non} = \mathbf{G} \odot \mathbf{D}^{non}$ and $\mathbf{G}_k^{rad} = \mathbf{G} \odot \mathbf{D}_k^{rad}$, where the symbol $\odot$ represents an element-by-element product of the matrices. Furthermore, the product between the identity matrix and $\mathbf{D}_k^{rad}$ above vanishes in the radiative propagator due to zeros along the diagonals of $\mathbf{D}_k^{rad}$.

Example III: Detection Matrices

For our example with two system states described earlier, the detection matrices of Eqs. 16-17 take simple forms. The radiative detection matrix has the same size as the generator matrix with nonzero elements wherever there is a rate associated to a radiative transition

$$\mathbf{D}_{d/a}^{rad} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/0 & 0 & 0 & 0 & 0 & 0 \\ 0/1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/0 & 0 & 0 \\ 0 & 0 & 0 & 0/1 & 0 & 0 \end{bmatrix},$$

where the subscripts d and a, respectively, denote photon detection in donor and acceptor channels. Similarly, the nonradiative detection matrix is obtained by setting all elements of the generator matrix related to radiative transitions to zero and the remaining to one as

$$\mathbf{D}^{non} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

2.3.7 Final likelihood

With the asymptotic forms of the reduced propagators in Eq. 14 now defined in the last subsection, we have all the ingredients needed to arrive at the final form of the likelihood.

To do so, we begin by considering the period right after the detection of the $(k-1)$th photon until the detection of the kth photon. For this time period, the nonradiative propagators in Eq. 14 can now be easily merged into a single propagator $\mathbf{\Pi}_k^{non} = \mathbf{exp}((T_k - T_{k-1})\mathbf{G}^{non})$,

as the commutative arguments of the exponentials can be readily added. Furthermore, at the end of this interphoton period, the radiative propagator $\mathbf{\Pi}_k^{(r)rad}$ marks the arrival of the k-th photon. The product of these two propagators

$$\mathbf{\Pi}_k^{non}\mathbf{\Pi}_k^{(r)rad} = \epsilon\mathbf{\Pi}_k^{non}\mathbf{G}_k^{rad} + \mathcal{O}(\epsilon^2) = \epsilon\,\mathbf{exp}((T_k - T_{k-1})\mathbf{G}^{non})\,\mathbf{G}_k^{rad} + \mathcal{O}(\epsilon^2), \tag{20}$$

now governs the stochastic evolution of the system-FRET composite during that interphoton period.

Inserting Eq.20 for each interphoton period into the likelihood for the HMM with second order structure in Eq. 14, we finally arrive at our desired likelihood

$$L = \epsilon^K \boldsymbol{\rho}_{start}\ \mathbf{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \mathbf{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \mathbf{\Pi}_{K-1}^{non}\mathbf{G}_{K-1}^{rad}\ \mathbf{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \mathbf{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{norm}^T + \mathcal{O}(\epsilon^{K+1}). \tag{21}$$

This likelihood has the same structure as shown by Gopich and Szabo in Ref. [40].

Example IV: Propagator and Likelihood

Here, we consider a simple FRET trace where two photons are detected at times 0.05 ms and 0.15 ms in the donor and acceptor channels, respectively. To demonstrate the ideas developed so far, we calculate the likelihood of these observations as (see Eq. 21)

$$L = \epsilon^2 \boldsymbol{\rho}_{start}\mathbf{\Pi}_1^{non}\mathbf{G}_1^{rad}\mathbf{\Pi}_2^{non}\mathbf{G}_2^{rad}\boldsymbol{\rho}_{norm}^T.$$

To do so, we first need to calculate $\mathbf{\Pi}_1^{non}$ using the nonradiative detection ($\mathbf{D}^{non}$) and generator ($\mathbf{G}$) matrices found in the previous example boxes

$$\mathbf{\Pi}_1^{non} = \mathbf{exp}(0.05\ (\mathbf{G} \odot \mathbf{D}^{non})) = \begin{bmatrix} 0.55 & 0 & 0 & 0.06 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.03 & 0 & 0 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and similarly

$$\mathbf{\Pi}_2^{non} = \mathbf{exp}((0.15 - 0.05)\ (\mathbf{G} \odot \mathbf{D}^{non})) = \begin{bmatrix} 0.30 & 0 & 0 & 0.06 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.03 & 0 & 0 & 0.33 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Next, we proceed to calculate $\mathbf{G}_1^{rad}$ and $\mathbf{G}_2^{rad}$. Remembering that the first photon was detected in the donor channel, we have (in ms^{-1} units)

$$\mathbf{G}_1^{rad} = \mathbf{G} \odot \mathbf{D}_d^{rad} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 277000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 277000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Similarly, since the second photon was detected in the acceptor channel, we can write (in ms^{-1} units)

$$\mathbf{G}_2^{rad} = \mathbf{G} \odot \mathbf{D}_d^{rad} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 285000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 285000 & 0 & 0 \end{bmatrix}.$$

We also assume that the system is initially in the superstate ϕ_1 giving $\boldsymbol{\rho}_{start} = [1, 0, 0, 0, 0, 0]$. Finally, putting everything together, we can find the likelihood as $L = 3.06\epsilon^2$ where ϵ is a constant and does not contribute to parameter estimations as we will show later.

2.3.8 Effect of binning single photon smFRET data

When considering binned FRET data, the time period of an experiment ($T_{end} - T_{start}$) is typically divided into a finite number ($= N$) of equally sized ($= \epsilon$) time windows (bins), and the photon counts (intensities) in each bin are recorded in the detection channels. This is in constrast to single photon analysis where individual photon arrival times are recorded. To arrive at the likelihood for such binned data, we start with the single photon likelihood derived in Eq. 15 where ϵ is not infinitesimally small, that is,

$$L = \boldsymbol{\rho}_{start}\, \mathbf{\Pi}_1^{(r)}\, \mathbf{\Pi}_2^{(r)}\, \mathbf{\Pi}_3^{(r)}\, \ldots\, \mathbf{\Pi}_{N-1}^{(r)}\, \mathbf{\Pi}_N^{(r)} \boldsymbol{\rho}_{norm}^T\,, \tag{22}$$

where

$$(\mathbf{\Pi}_n^{(r)})_{a_n \to a_{n+1}} = (\mathbf{\Pi})_{a_n \to a_{n+1}} \times (\mathbf{D}_n)_{a_n \to a_{n+1}}\,,$$

or in the matrix notation

$$\mathbf{\Pi}_n^{(r)} = \mathbf{\Pi} \odot \mathbf{D}_n = \mathbf{exp}(\epsilon \mathbf{G}) \odot \mathbf{D}_n, \tag{23}$$

where $\mathbf{D}_n$ is the detection matrix introduced in Sec. 2.3.4.

Next, we must sum over all superstate trajectories that may give rise to the recorded photon counts (observations) in each bin. However, such a sum is challenging to compute

analytically and has been attempted in [38]. Here, we will only show likelihood computation under commonly applied approximations/assumptions when analyzing binned smFRET data, which are: 1) bin size ϵ is much smaller than typical times spent in a system state or, in other words, for a system transition rate $\lambda_{\sigma_i \to \sigma_j}$, we have $\epsilon \lambda_{\sigma_i \to \sigma_j} \ll 1$; and 2) excitation rate λ_{ex} is much slower than dye relaxation and FRET rates, or in other words, interphoton periods are much larger than the excited state lifetimes.

The first assumption is based on realistic situations where system kinetics (at seconds timescale) are many orders of magnitude slower than the photophysical transitions (at nanoseconds timescale). This timescale separation allows us to simplify the propagator calculation in Eq. 23. To see that, we first separate the system transition rates from photophysical transition rates in the generator matrix as

$$\mathbf{G} = \mathbf{G}_\sigma \otimes \mathbb{I} + \mathbf{G}_\psi, \tag{24}$$

where $\otimes$ denotes a tensor product, $\mathbf{G}_\sigma$ is the portion of generator matrix $\mathbf{G}$ containing only system transition rates previously defined in Eq. 3, and $\mathbf{G}_\psi$ is the portion containing only photophysical transition rates, that is,

$$\mathbf{G}_\sigma \otimes \mathbb{I} = \begin{bmatrix} -\sum\limits_{j \neq 1} \lambda_{\sigma_1 \to \sigma_j} \mathbb{I} & \lambda_{\sigma_1 \to \sigma_2} \mathbb{I} & \dots & \lambda_{\sigma_1 \to \sigma_{M_\sigma}} \mathbb{I} \\ \lambda_{\sigma_2 \to \sigma_1} \mathbb{I} & -\sum\limits_{j \neq 2} \lambda_{\sigma_2 \to \sigma_j} \mathbb{I} & \dots & \lambda_{\sigma_2 \to \sigma_{M_\sigma}} \mathbb{I} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{\sigma_{M_\sigma} \to \sigma_1} \mathbb{I} & \lambda_{\sigma_{M_\sigma} \to \sigma_2} \mathbb{I} & \dots & -\sum\limits_{j \neq M_\sigma} \lambda_{\sigma_{M_\sigma} \to \sigma_j} \mathbb{I} \end{bmatrix}, \tag{25}$$

and

$$\mathbf{G}_\psi = \begin{bmatrix} \mathbf{G}^\psi_{\sigma_1} & 0 & \dots & 0 \\ 0 & \mathbf{G}^\psi_{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{G}^\psi_{\sigma_{M_\sigma}} \end{bmatrix}, \tag{26}$$

where $\mathbf{G}^\psi_{\sigma_i}$ is the photophysical generator matrix corresponding to system state σ_i given in Eq. 4.

Now plugging Eq. 24 into the full propagator $\mathbf{\Pi} = \mathbf{exp}(\epsilon \mathbf{G})$ and applying the famous Zassenhaus formula for matrix exponentials, we get

$$\begin{aligned} \mathbf{\Pi} &= \mathbf{exp}(\epsilon(\mathbf{G}_\sigma \otimes \mathbb{I} + \mathbf{G}_\psi)) \\ &= \mathbf{exp}(\epsilon(\mathbf{G}_\sigma \otimes \mathbb{I}))\, \mathbf{exp}(\epsilon \mathbf{G}_\psi)\, \mathbf{exp}(-\frac{\epsilon^2}{2}[\mathbf{G}_\sigma \otimes \mathbb{I}, \mathbf{G}_\psi])\, \mathbf{exp}(\mathcal{O}(\epsilon^3)) \end{aligned} \tag{27}$$

where the square brackets represent the commutator of the constituting matrices and the last term represents the remaining exponentials involving higher order commutators. Fur-

thermore, the commutator $[\mathbf{G}_\sigma \otimes \mathbb{I}, \mathbf{G}_\psi]$ results in a very sparse matrix given by

$$[\mathbf{G}_\sigma \otimes \mathbb{I}, \mathbf{G}_\psi] = \begin{bmatrix} 0 & \lambda_{\sigma_1\to\sigma_2}(\mathbf{G}^\psi_{\sigma_2} - \mathbf{G}^\psi_{\sigma_1}) & \dots & \lambda_{\sigma_1\to\sigma_{M_\sigma}}(\mathbf{G}^\psi_{\sigma_{M_\sigma}} - \mathbf{G}^\psi_{\sigma_1}) \\ \lambda_{\sigma_2\to\sigma_1}(\mathbf{G}^\psi_{\sigma_1} - \mathbf{G}^\psi_{\sigma_2}) & 0 & \dots & \lambda_{\sigma_2\to\sigma_{M_\sigma}}(\mathbf{G}^\psi_{\sigma_{M_\sigma}} - \mathbf{G}^\psi_{\sigma_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{\sigma_{M_\sigma}\to\sigma_1}(\mathbf{G}^\psi_{\sigma_1} - \mathbf{G}^\psi_{\sigma_{M_\sigma}}) & \lambda_{\sigma_{M_\sigma}\to\sigma_2}(\mathbf{G}^\psi_{\sigma_2} - \mathbf{G}^\psi_{\sigma_{M_\sigma}}) & \dots & 0 \end{bmatrix}, \tag{28}$$

where

$$\lambda_{\sigma_i\to\sigma_j}(\mathbf{G}^\psi_{\sigma_j} - \mathbf{G}^\psi_{\sigma_i}) = \lambda_{\sigma_i\to\sigma_j} \begin{bmatrix} 0 & 0 & 0 \\ 0 & -(\lambda^{FRET}_{\sigma_j} - \lambda^{FRET}_{\sigma_i}) & (\lambda^{FRET}_{\sigma_j} - \lambda^{FRET}_{\sigma_i}) \\ 0 & 0 & 0 \end{bmatrix}.$$

Now, the propagator calculation in Eq. 27 simplifies if the commutator $\epsilon^2[\mathbf{G}_\sigma \otimes \mathbb{I}, \mathbf{G}_\psi] \to 0$, implying that either the bin size ϵ is very small such that $\epsilon\lambda_{\sigma_i\to\sigma_j} \ll 1$ (our first assumption) or FRET rates/efficiencies are almost indistinguishable ($\epsilon(\lambda^{FRET}_{\sigma_j} - \lambda^{FRET}_{\sigma_i}) \approx 0$). Under such conditions, the system state can be assumed to stay constant during a bin, with system transitions only occurring at the ends of bin periods. Furthermore, the full propagator $\mathbf{\Pi}$ in Eq. 27 can now be approximated as

$$\mathbf{\Pi} = \mathbf{exp}(\epsilon\mathbf{G}) \approx \mathbf{exp}(\epsilon(\mathbf{G}_\sigma \otimes \mathbb{I}))\,\mathbf{exp}(\epsilon\mathbf{G}_\psi) = (\mathbf{\Pi}_\sigma \otimes \mathbb{I})\,\mathbf{\Pi}_\psi, \tag{29}$$

where the last equality follows from the block diagonal form of $\mathbf{G}_\sigma \otimes \mathbb{I}$ given in Eq. 25 and $\mathbf{\Pi}_\sigma = \mathbf{exp}(\epsilon\mathbf{G}_\sigma)$ is the system transition probability matrix (propagator) given as

$$\mathbf{\Pi}_\sigma = \begin{bmatrix} \pi_{\sigma_1\to\sigma_1} & \pi_{\sigma_1\to\sigma_2} & \dots & \pi_{\sigma_1\to\sigma_{M_\sigma}} \\ \pi_{\sigma_2\to\sigma_1} & \pi_{\sigma_2\to\sigma_2} & \dots & \pi_{\sigma_2\to\sigma_{M_\sigma}} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{\sigma_{M_\sigma}\to\sigma_1} & \pi_{\sigma_{M_\sigma}\to\sigma_2} & \dots & \pi_{\sigma_{M_\sigma}\to\sigma_{M_\sigma}} \end{bmatrix}. \tag{30}$$

Moreover, $\mathbf{\Pi}_\psi = \mathbf{exp}(\epsilon\mathbf{G}_\psi)$ is the photophysical transition probability matrix (propagator) as

$$\mathbf{\Pi}_\psi = \begin{bmatrix} \mathbf{\Pi}^\psi_{\sigma_1} & 0 & \dots & 0 \\ 0 & \mathbf{\Pi}^\psi_{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{\Pi}^\psi_{\sigma_{M_\sigma}} \end{bmatrix}, \tag{31}$$

where the elements are given as $\mathbf{\Pi}^\psi_{\sigma_i} = \mathbf{exp}(\epsilon\mathbf{G}^\psi_{\sigma_i})$. Furthermore, because of the block diagonal structure of $\mathbf{\Pi}_\psi$, the matrix multiplication in Eq. 29 results in

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{\sigma_1\to\sigma_1}\mathbf{\Pi}^{\psi}_{\sigma_1} & \pi_{\sigma_1\to\sigma_2}\mathbf{\Pi}^{\psi}_{\sigma_1} & \dots & \pi_{\sigma_1\to\sigma_{M_\sigma}}\mathbf{\Pi}^{\psi}_{\sigma_1} \\ \pi_{\sigma_2\to\sigma_1}\mathbf{\Pi}^{\psi}_{\sigma_2} & \pi_{\sigma_2\to\sigma_2}\mathbf{\Pi}^{\psi}_{\sigma_2} & \dots & \pi_{\sigma_2\to\sigma_{M_\sigma}}\mathbf{\Pi}^{\psi}_{\sigma_2} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{\sigma_{M_\sigma}\to\sigma_1}\mathbf{\Pi}^{\psi}_{\sigma_{M_\sigma}} & \pi_{\sigma_{M_\sigma}\to\sigma_2}\mathbf{\Pi}^{\psi}_{\sigma_{M_\sigma}} & \dots & \pi_{\sigma_{M_\sigma}\to\sigma_{M_\sigma}}\mathbf{\Pi}^{\psi}_{\sigma_{M_\sigma}} \end{bmatrix}. \tag{32}$$

After deriving the full propagator $\mathbf{\Pi}$ for the time period ϵ (bin) under our first assumption, we now proceed to incorporate observations during this period via detection matrices $\mathbf{D}_n$ to compute the reduced propagator of Eq. 23. To do so, we now apply our second assumption of relatively slower excitation rate λ_{ex}. This assumption implies that interphoton periods are dominated by the time spent in the ground state of the FRET pair and are distributed according to a single exponential distribution, $\mathbf{Exponential}(\lambda_{ex})$. Consequently, the total photon counts per bin follow a Poisson distribution, $\mathbf{Poisson}(\epsilon\lambda_{ex})$, independent of the photophysical portion of the photophysical trajectory taken from superstate a_n to a_{n+1}.

Now, the first and the second assumptions imply that the observation during the n-th bin only depends on the system state s_n (or the associated FRET rate $\lambda^{FRET}_{s_n}$). As such we can approximate the detection matrix elements as

$$(\mathbf{D}_n)_{a_n\to a_{n+1}} = p(w_n|a_n, a_{n+1}) \approx p(w_n|s_n). \tag{33}$$

Using these approximations, the reduced propagator in Eq. 23 can now be written as

$$\begin{aligned} \mathbf{\Pi}^{(r)}_n &= \mathbf{\Pi} \odot \mathbf{D}_n \\ &\approx \begin{bmatrix} \pi_{\sigma_1\to\sigma_1}\, p(w_n|s_n=\sigma_1)\,\mathbf{\Pi}^{\psi}_{\sigma_1} & \dots & \pi_{\sigma_1\to\sigma_{M_\sigma}}\, p(w_n|s_n=\sigma_1)\,\mathbf{\Pi}^{\psi}_{\sigma_1} \\ \pi_{\sigma_2\to\sigma_1}\, p(w_n|s_n=\sigma_2)\,\mathbf{\Pi}^{\psi}_{\sigma_2} & \dots & \pi_{\sigma_2\to\sigma_{M_\sigma}}\, p(w_n|s_n=\sigma_2)\,\mathbf{\Pi}^{\psi}_{\sigma_2} \\ \vdots & \ddots & \vdots \\ \pi_{\sigma_{M_\sigma}\to\sigma_1}\, p(w_n|s_n=\sigma_{M_\sigma})\,\mathbf{\Pi}^{\psi}_{\sigma_{M_\sigma}} & \dots & \pi_{\sigma_{M_\sigma}\to\sigma_{M_\sigma}}\, p(w_n|s_n=\sigma_{M_\sigma})\,\mathbf{\Pi}^{\psi}_{\sigma_{M_\sigma}} \end{bmatrix}. \end{aligned} \tag{34}$$

Next, to compute the likelihood for the n-th bin, we need to sum over all possible superstate trajectories within this bin as

$$\begin{aligned} L_n &= \boldsymbol{\rho}_n\, \mathbf{\Pi}^{(r)}_n\, \boldsymbol{\rho}^T_{norm} \\ &\approx \sum_{i=1}^{M_\sigma}\sum_{j=1}^{M_\sigma} p(w_n|s_n=\sigma_i)\,\pi_{\sigma_i\to\sigma_j}\,\left(\boldsymbol{\rho}_{\sigma_i,n}\mathbf{\Pi}^{\psi}_{\sigma_i}\boldsymbol{\rho}^T_{norm}\right), \end{aligned} \tag{35}$$

where $\boldsymbol{\rho}_n$ is a normalized row vector populated by probabilities of finding the system-FRET composite in the possible superstates at the beginning of the n-th bin. Furthermore, we have written portions of $\boldsymbol{\rho}_n$ corresponding to system state σ_i as $\boldsymbol{\rho}_{\sigma_i,n}$. To be more explicit, we have

$$\boldsymbol{\rho}_n = \begin{bmatrix}\boldsymbol{\rho}_{\sigma_1,n} & \boldsymbol{\rho}_{\sigma_2,n} & \dots & \boldsymbol{\rho}_{\sigma_{M_\sigma},n}\end{bmatrix},$$

following the convention in Sec. 2.3 and using n to now represent time t_n.

Moreover, since each row of $\mathbf{\Pi}^{\psi}_{\sigma_i} = \mathbf{exp}(\epsilon \mathbf{G}^{\psi}_{\sigma_i})$ sums to one, we have $\mathbf{\Pi}^{\psi}_{\sigma_i} \boldsymbol{\rho}^T_{norm} = \boldsymbol{\rho}^T_{norm}$, which simplifies the bin likelihood of Eq. 35 to

$$\begin{aligned} L_n &\approx \sum_{i=1}^{M_\sigma} \sum_{j=1}^{M_\sigma} p(w_n|s_n = \sigma_i)\, \pi_{\sigma_i \to \sigma_j} \left(\boldsymbol{\rho}_{\sigma_i,n} \boldsymbol{\rho}^T_{norm}\right) \\ &= \sum_{i=1}^{M_\sigma} \sum_{j=1}^{M_\sigma} p(w_n|s_n = \sigma_i)\, \pi_{\sigma_i \to \sigma_j}\, \rho_{\sigma_i,n}, \end{aligned} \tag{36}$$

where we have defined $\rho_{\sigma_i,n} \equiv \boldsymbol{\rho}_{\sigma_i,n} \boldsymbol{\rho}^T_{norm} = \sum_j \rho_{\sigma_i,\psi_j}$ as the probability of the system to occupy system state σ_i. We can also write the previous equation in the matrix form as

$$L_n \approx \boldsymbol{\rho}^{\sigma}_n \left(\mathbf{\Pi}_\sigma \odot \mathbf{D}^{\sigma}_n\right) \boldsymbol{\rho}^T_{norm}, \tag{37}$$

where $\boldsymbol{\rho}^{\sigma}_n$ is a row vector of length M_σ (number of system states) populated by $\rho_{\sigma_i,n}$ for each system state, and $\mathbf{D}^{\sigma}_n$, in the same spirit as $\mathbf{D}_n$, is a detection matrix of dimensions $M_\sigma \times M_\sigma$ populated by observation probability $p(w_n|s_n = \sigma_i)$ in each row corresponding to system state σ_i. Furthermore, defining $\mathbf{\Pi}^{\sigma}_n \equiv (\mathbf{\Pi}_\sigma \odot \mathbf{D}^{\sigma}_n)$, we note here that $\mathbf{\Pi}^{\sigma}_n$ propagates probabilities during the n-th bin in a similar manner as the reduced propagators $\mathbf{\Pi}^{(r)}_n$ of Eq. 23.

Therefore, we can now multiply these new propagators for each bin to approximate the likelihood of Eq. 22 as

$$L \approx \boldsymbol{\rho}^{\sigma}_{start}\, \mathbf{\Pi}^{\sigma}_1\, \mathbf{\Pi}^{\sigma}_2\, \mathbf{\Pi}^{\sigma}_3\, \ldots\, \mathbf{\Pi}^{\sigma}_N\, \boldsymbol{\rho}^T_{norm}. \tag{38}$$

where $\boldsymbol{\rho}^{\sigma}_{start}$ is a row vector, similar to $\boldsymbol{\rho}^{\sigma}_n$, populated by probabilities of being in a given system state at the beginning of an experiment.

To conclude, our two assumptions regarding system kinetics and excitation rate allow us to significantly reduce the dimensions of the propagators. This, in turn, leads to much lowered expense for likelihood computation. However, cheaper computation comes at the expense of requiring large number of photon detections or excitation rate per bin to accurately determine FRET efficiencies (identify system states) since we marginalize over photophysics in each bin. Such high excitation rates lead to faster photobleaching and increased phototoxicity, and thereby much shorter experiment durations. As we will see in Sec. 2.5.1, this problem can be mitigated by using pulsed illumination, where the likelihood takes a similar form as Eq. 38, but FRET efficiencies can be accurately estimated from the measured microtimes.

2.4 Detection Effects

In the previous section, we assumed idealized detectors to illustrate basic ideas on detection matrices. However, realistic FRET experiments must typically account for detector nonidealities. For instance, many emitted photons may simply go undetected when the detection efficiency of single photon detectors, *i.e.*, the probability of an incident photon being successfully registered, is less than one due to inherent nonlinearities associated with the

electronics [22] or the use of filters in cases of polarized fluorescent emission [58, 59]. Additionally, donor photons may be detected in the channel reserved for acceptor photons or vice-versa due to emission spectrum overlap [60]. This phenomenon, commonly known as crosstalk, crossover, or bleedthrough, can significantly affect the determination of quantities such as transition rates and FRET efficiencies as we demonstrate later in our results. Other effects adding noise to fluorescence signal include dark current (false signal in the absence of incident photons), dead time (the time a detector takes to relax back into its active mode after a photon detection), and timing jitter or IRF (stochastic delay in the output signal after a detector receives a photon) [22]. In this section, we describe the incorporation of all such effects into our model except dark current and background emissions which require more careful treatment and will be discussed in Sec. 2.6.

2.4.1 Crosstalk and detection efficiency

Noise sources such as crosstalk and detection efficiency necessarily result in photon detection being treated as a stochastic process. Both crosstalk and detection efficiency can be included into the propagators in both cases by substituting the zeros and ones, appearing in the ideal radiative and nonradiative detection matrices (Eqs. 16-17), with probabilities between zero and one. In such a way, the resulting propagators obtained from these detection matrices, in turn, incorporate into the likelihood the effects of crosstalk and detection efficiency into the model.

Here, in the presence of crosstalk, for clarity, we add a superscript to the radiative detection matrix of Eq. 17 for the k-th photon, $\mathbf{D}_k^{rad-ct}$. The elements of this detection matrix for the $a_n \to b_n$ transition, when a photon intended for channel j is registered in channel i reads

$$(\mathbf{D}_k^{rad-ct})_{a_n \to b_n} = \begin{cases} 0 & \text{Nonradiative transitions} \\ \phi_{ji} & \text{Radiative transitions} \end{cases}$$

where ϕ_{ji} is the probability for this event (upon transition from superstate a_n to b_n). Further, detector efficiencies can also be accounted for in these probabilities in order to represent the combined effects of crosstalk, arising from spectral overlap, and absence of detection channel registration. When we do so, we recover $\sum_i \phi_{ji} \leq 1$ (for cases where i and j can be both the same or different), as not all emitted photons can be accounted for by the detection channels.

This new detection matrix above results in the following modification to the radiative propagator of Eq. 19 for the k-th photon

$$\mathbf{\Pi}_k^{(r)rad-ct} = (\mathbb{I} + \epsilon \mathbf{G} + \mathcal{O}(\epsilon^2)) \odot \mathbf{D}_k^{rad-ct} = \epsilon \mathbf{G}_k^{rad-ct} + \mathcal{O}(\epsilon^2).$$

The second equality above follows by recognizing that the identity matrix multiplied, element-wise, by $\mathbf{D}_k^{rad-ct}$ is zero. By definition, $\mathbf{G}_k^{rad-ct}$ is the remaining nonzero product.

On the other hand, for time periods when no photons are detected, the nonradiative detection matrices in Eq. 16 become

$$(\mathbf{D}_n)_{a_n \to b_n} = (\mathbf{D}^{non-ct})_{a_n \to b_n} = \begin{cases} 1 & \text{Nonradiative transitions} \\ 1 - \sum_j \phi_{ij} & \text{Radiative transitions} \end{cases},$$

where the sum gives the probability of the photon intended for channel i to be registered in any channel. The nonradiative propagator of Eq. 18 for an infinitesimal period of size ϵ in the presence of crosstalk and inefficient detectors is now

$$\mathbf{\Pi}^{(r)non-ct} = (\mathbb{I} + \epsilon\mathbf{G} + \mathcal{O}(\epsilon^2)) \odot \mathbf{D}^{non-ct} = \mathbf{exp}(\epsilon\mathbf{G}^{non-ct}) + \mathcal{O}(\epsilon^2), \tag{39}$$

where $\mathbf{G}^{non-ct} = \mathbf{G} \odot \mathbf{D}^{non-ct}$. With the propagators incorporating crosstalk and detection efficiency now defined, the evolution during an interphoton period between the $(k-1)$-th photon and the k-th photon of size $(T_k - T_{k-1})$ is now governed by the product

$$\mathbf{\Pi}_k^{non-ct}\mathbf{\Pi}_k^{(r)rad-ct} = \epsilon\mathbf{\Pi}_k^{non-ct}\mathbf{G}_k^{rad-ct} + \mathcal{O}(\epsilon^2), \tag{40}$$

where the nonradiative propagators in Eq. 39 have now been merged into a single propagator $\mathbf{\Pi}_k^{non-ct} = \mathbf{exp}((T_k - T_{k-1})\mathbf{G}^{non-ct})$ following the same procedure as Eq. 20.

Finally, inserting Eq. 40 for each interphoton period into the likelihood of Eq. 14, we arrive at the final likelihood incorporating crosstalk and detection efficiency as

$$\begin{aligned} L = \epsilon^K \boldsymbol{\rho}_{start}\ \mathbf{\Pi}_1^{non-ct}\mathbf{G}_1^{rad-ct}\ \mathbf{\Pi}_2^{non-ct}\mathbf{G}_2^{rad-ct}\ \ldots\ \mathbf{\Pi}_{K-1}^{non-ct}\mathbf{G}_{K-1}^{rad-ct} \\ \times \mathbf{\Pi}_K^{non-ct}\mathbf{G}_K^{rad-ct}\ \mathbf{\Pi}_{end}^{non-ct}\ \boldsymbol{\rho}_{norm}^T + \mathcal{O}(\epsilon^{K+1}). \end{aligned}$$

After incorporating crosstalk and detector efficiencies into our model, we briefly explain the calibration of the crosstalk probabilities/detection efficiencies ϕ_{ij}. To calibrate these parameters, two samples, one containing only donor dyes and another containing only acceptor dyes, are individually excited with a laser under the same power to determine the number of donor photons n_{di}^{raw} and number of acceptor photons n_{ai}^{raw} detected in channel i.

From photon counts recorded for the donor only sample, assuming ideal detectores with 100% efficiency, we can compute the crosstalk probabilities for donor photons going to channel i, ϕ_{di}, using the photon count ratios as $\phi_{di} = n_{di}^{raw}/n_d^{em}$ where n_d^{em} is the absolute number of emitted donor photons. Similarly, crosstalk probabilities for acceptor photons going to channel i, ϕ_{ai}, can be estimated as $\phi_{ai} = n_{ai}^{raw}/n_a^{em}$ where n_a^{em} is the absolute number of emitted acceptor photons. In the matrix form, these crosstalk factors for a two-detector setup can be written as

$$\mathbf{A} = \begin{bmatrix} \phi_{a1} & \phi_{d1} \\ \phi_{a2} & \phi_{d2} \end{bmatrix}. \tag{41}$$

Using this matrix, for the donor only sample, we can now write

$$\begin{bmatrix} n_{d1}^{raw} \\ n_{d2}^{raw} \end{bmatrix} = \mathbf{A} \begin{bmatrix} 0 \\ n_d^{em} \end{bmatrix} = \begin{bmatrix} \phi_{d1} \\ \phi_{d2} \end{bmatrix} n_d^{em}, \tag{42}$$

and similarly for the acceptor only sample

$$\begin{bmatrix} n_{a1}^{raw} \\ n_{a2}^{raw} \end{bmatrix} = \mathbf{A} \begin{bmatrix} n_a^{em} \\ 0 \end{bmatrix} = \begin{bmatrix} \phi_{a1} \\ \phi_{a2} \end{bmatrix} n_a^{em}. \tag{43}$$

However, it is difficult to estimate the absolute number of emitted photons n_d^{em} and n_a^{em} experimentally, and therefore the crosstalk factors in $\mathbf{A}$ can only be determined up to multiplicative factors of n_d^{em} and n_a^{em}.

Since scaling the photon counts in an smFRET trace by an overall constant does not affect the FRET efficiency estimates (determined by photon count ratios) and escape rates (determined by changes in FRET efficiency), we only require crosstalk factors up to a constant as in the last equation.

For this reason, one possible solution toward determining the matrix elements of $\mathbf{A}$ up to one multiplicative constant is to first tune dye concentrations such that the ratio $n_d^{em}/n_a^{em} = 1$, which can be accomplished experimentally. This allows us to write the crosstalk factors in the matrix form up to a constant as follows

$$\mathbf{A} = \begin{bmatrix} \phi_{a1} & \phi_{d1} \\ \phi_{a2} & \phi_{d2} \end{bmatrix} \propto \begin{bmatrix} n_{a1}^{raw} & n_{d1}^{raw} \\ n_{a2}^{raw} & n_{d2}^{raw} \end{bmatrix}. \tag{44}$$

It is common to set the multiplicative factor in Eq. 44 by the total donor photons counts $\sum_j n_{dj}^{raw}$ to give

$$\mathbf{A} = \begin{bmatrix} \phi_{a1} & \phi_{d1} \\ \phi_{a2} & \phi_{d2} \end{bmatrix} \equiv \begin{bmatrix} \frac{n_{a1}^{raw}}{\sum_j n_{dj}^{raw}} & \frac{n_{d1}^{raw}}{\sum_j n_{dj}^{raw}} \\ \\ \frac{n_{a2}^{raw}}{\sum_j n_{dj}^{raw}} & \frac{n_{d2}^{raw}}{\sum_j n_{dj}^{raw}} \end{bmatrix}. \tag{45}$$

We note that from the convention adopted here, we have $\phi_{d1} + \phi_{d2} = 1$.

Furthermore, in situations where realistic detectors affect the raw counts, the matrix elements of $\mathbf{A}$ as computed above automatically incorporate the effects of detector inefficiencies including the fact that $\sum_j \phi_j \leq 1$.

Additionally, the matrix $\mathbf{A}$ can be further generalized to account for more than two detectors by appropriately expanding the size of the matrix dimensions to coincide with the number of detectors. Calibration of the matrix elements then follows the same procedure as above.

Now, in performing single photon FRET analysis, we will use directly the elements of $\mathbf{A}$ in constructing our measurement matrix. However, it is also common, to compute the matrix elements of $\mathbf{A}$ from what is termed the route correction matrix (RCM) [61] typically used in binned photon analysis. The RCM is defined as the inverse of $\mathbf{A}$ to obtain corrected counts n_d^{em} and n_a^{em} up to a proportionality constant as

$$\mathbf{RCM} \propto \begin{bmatrix} \phi_{d2} & -\phi_{d1} \\ -\phi_{a2} & \phi_{a1} \end{bmatrix}. \tag{46}$$

Example V: Detection Matrices with Crosstalk and Detector Efficiencies

For our example with two system states, we had earlier shown detection matrices for ideal detectors. Here, we incorporate crosstalk and detector efficiencies into these matrices. Moreover, we assume a realistic RCM [62] given as

$$\mathbf{RCM} \propto \begin{bmatrix} 1.0 & -0.22 \\ 0.0 & 1.02 \end{bmatrix}.$$

However, following the convention of Eq. 45, we scale the matrix provided by a sum of absolute values of its first row elements, namely 1.22, leading to effective crosstalk factors ϕ_{ij} given as

$$\phi_{a1} = 0.84,\ \phi_{a2} = 0.0,\ \phi_{d1} = 0.18, \text{ and } \phi_{d2} = 0.82.$$

As such, these values imply approximately 18% crosstalk from donor to acceptor channel and 84% detection efficiency for acceptor channel without any crosstalk using the convention adopted in Eq. 45. Now, we modify the ideal radiative detection matrices by replacing their nonzero elements with the calibrated ϕ_{ij}'s above

$$\mathbf{D}_{d/a}^{rad-ct} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_{d2}/\phi_{d1} & 0 & 0 & 0 & 0 & 0 \\ \phi_{a2}/\phi_{a1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_{d2}/\phi_{d1} & 0 & 0 \\ 0 & 0 & 0 & \phi_{a2}/\phi_{a1} & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.82/0.18 & 0 & 0 & 0 & 0 & 0 \\ 0/0.84 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.82/0.18 & 0 & 0 \\ 0 & 0 & 0 & 0/0.84 & 0 & 0 \end{bmatrix}.$$

Similarly, we modify the ideal nonradiative detection matrix by replacing the zero elements by $1 - \phi_{a1} - \phi_{a2} = 0.16$ and $1 - \phi_{d2} - \phi_{d1} = 0$ as follows

$$\mathbf{D}^{non-ct} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0.16 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0.16 & 1 & 1 \end{bmatrix}.$$

2.4.2 Effects of detector dead time

Typically, a detection channel i becomes inactive (dead) after the detection of a photon for a period δ_i as specified by the manufacturer. Consequently, radiative transitions associated with that channel cannot be monitored during that period.

To incorporate this detector dead period into our likelihood model, we break an inter-photon period between the $(k-1)$-th and k-th photon into two intervals: the first interval with an inactive detector and the second one when the detector is active. Assuming that the $(k-1)$-th photon is detected in the ith channel, the first interval is thus δ_{i_k} long. As such,

we can define the detection matrix for this interval as

$$(\mathbf{D}^{i_k-dead})_{a_n\to b_n} = \begin{cases} 1 & \text{All transitions not intended for channel } i_k \\ 0 & \text{All transitions intended for channel } i_k \end{cases}.$$

Next, corresponding to this detection matrix, we have the propagator

$$\mathbf{\Pi}_k^{i_k-dead} = \mathbf{exp}(\delta_{i_k}\,(\mathbf{G}\odot\mathbf{D}^{i_k-dead})) = \mathbf{exp}(\delta_{i_k}\,\mathbf{G}^{i_k-dead}),$$

that evolves the superstate during the detector dead time. This propagator can now be used to incorporate the detector dead time into Eq. 20 to represent the evolution during the period between the $(k-1)$-th and k-th photons as

$$\mathbf{\Pi}_{k-1}^{i_{k-1}-dead}\mathbf{\Pi}_k^{non}\mathbf{\Pi}_k^{(r)rad} = \epsilon\mathbf{\Pi}_{k-1}^{i_{k-1}-dead}\mathbf{\Pi}_k^{non}\mathbf{G}_k^{rad} + \mathcal{O}(\epsilon^2), \tag{47}$$

where $\mathbf{\Pi}_k^{non}\mathbf{\Pi}_k^{(r)rad}$ describes the evolution when the detector is active.

Finally, inserting Eq. 47 for each interphoton period into the likelihood for the HMM with a second order structure in Eq. 14, we arrive at the following likelihood that includes detector dead time

$$L \propto \boldsymbol{\rho}_{start}\;\mathbf{\Pi}_1^{non}\mathbf{G}_1^{rad}\;\mathbf{\Pi}_1^{i_1-dead}\mathbf{\Pi}_2^{non}\mathbf{G}_2^{rad}\;\mathbf{\Pi}_2^{i_2-dead}\;\ldots\;\mathbf{\Pi}_K^{non}\mathbf{G}_K^{rad}\;\mathbf{\Pi}_K^{i_K-dead}\mathbf{\Pi}_{end}^{non}\;\boldsymbol{\rho}_{norm}^T. \tag{48}$$

To provide an explicit example on the effect of the detector dead time on the likelihood, we take a detour for pedagogical reasons. In this context, we consider a very simple case of one detection channel (dead time δ) observing a fluorophore with two photophysical states, ground (ψ_1) and excited (ψ_2), illuminated by a laser. The data in this case contains only photon arrival times

$$\{T_1, T_2, T_3, \ldots, T_K\}. \tag{49}$$

The generator matrix containing the photophysical transition rates for this setup is

$$\mathbf{G} = \begin{bmatrix} * & \lambda_{\psi_1\to\psi_2} \\ \lambda_{\psi_2\to\psi_1} & * \end{bmatrix} = \begin{bmatrix} * & \lambda_{ex} \\ \lambda_d & * \end{bmatrix},$$

where the $*$ along the diagonal represents the negative row-sum of the remaining elements, λ_{ex} is the excitation rate, and λ_d is the donor relaxation rate.

Here, all transitions are possible during detector dead times as there are no observations. As such, the dead time propagators in the likelihood (Eq. 48) are simply expressed as exponentials of the full generator matrix, that is, $\mathbf{\Pi}_k^{i_k-dead} = \mathbf{exp}(\delta\mathbf{G})$, leaving the normalization of the propagated probability vector $\boldsymbol{\rho}$ unchanged, *e.g.*, just as we had seen in Eq. 8.

As we will see, these dead times, similar to detector inefficiencies, simply increase our uncertainty over parameters we wish to learn, such as kinetics, by virtue of providing less information. By contrast, background emissions and cross talk, provide false information. However, the net effect is the same: all noise sources increasing uncertainty.

2.4.3 Adding the detection IRF

Due to various sources of noise impacting the detection timing electronics (also known as jitter), the time elapsed between photon arrival and detection is itself a hidden (latent) random variable [22]. Under continuous illumination, we say that this stochastic delay in time is sampled from a probability density, $f(\tau)$, termed the *detection* instrument response function (IRF). To incorporate the detection IRF into the likelihood of Eq. 48, we convolute the propagators with $f(\tau)$ as follows

$$\begin{aligned} L \propto \boldsymbol{\rho}_{start} & \left(\int_0^{\tau_{IRF}} d\tau_1 \mathbf{\Pi}_1^{non}(\Delta T_1 - \tau_1)\ \mathbf{G}_1^{rad}\ \mathbf{\Pi}_1^{i_1-dead}(\tau_1)\ f(\tau_1) \right) \\ & \times \left(\int_0^{\tau_{IRF}} d\tau_2 \mathbf{\Pi}_2^{non}(\Delta T_2 - \tau_2)\ \mathbf{G}_2^{rad}\ \mathbf{\Pi}_2^{i_2-dead}(\tau_2)\ f(\tau_2) \right) \\ & \vdots \\ & \times \left(\int_0^{\tau_{IRF}} d\tau_K \mathbf{\Pi}_K^{non}(\Delta T_K - \tau_K)\ \mathbf{G}_K^{rad}\ \mathbf{\Pi}_K^{i_K-dead}(\tau_K)\ f(\tau_K) \right)\ \mathbf{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{norm}^T, \end{aligned} \quad (50)$$

where we have used dead time propagators $\mathbf{\Pi}_k^{i_k-dead}$ to incorporate detector inactivity during the period between photon reception and detector reactivation. Moreover, we have $\mathbf{\Pi}_k^{non}(\Delta T_k - \tau_k) = \mathbf{exp}\left((\Delta T_k - \tau_k)\mathbf{G}^{non}\right)$ as described in Eq. 18.

To facilitate the computation of this likelihood, we use the fact that typical acquisition devices record at discrete (but very small) time intervals. For instance, a setup with the smallest acquisition time of 16 ps and a detection IRF distribution that is approximately 100 ps wide will have the detection IRF spread over, roughly, six acquisition periods. This allows each convolution integral to be discretized over the six acquisition intervals and computed in parallel, thereby avoiding extra real computational time associated to this convolution other than the overhead associated with parallelization.

2.5 Illumination Features

After discussing detector effects, we continue here by further considering different illumination features. For simplicity alone, our likelihood computation until now assumed continuous illumination with a uniform intensity. More precisely, the element λ_{ex} of the generator matrix in Eq. 4 was assumed to be time-independent. Here, we generalize our formulation and show how other illumination setups (such as pulsed illumination and alternating laser excitation, ALEX [63]) can be incorporated into the likelihood by simply assigning a time dependence to the excitation rate $\lambda_{ex}(t)$.

2.5.1 Pulsed illumination

Here, we consider an smFRET experiment where the FRET pair is illuminated using a laser for a very short period of time (a pulse), δ_{pulse}, at regular intervals of size τ; see Fig. 3(a). Now, as in the case of continuous illumination with constant intensity, the likelihood for a set of observations acquired using pulsed illumination takes a similar form to Eq. 21 involving

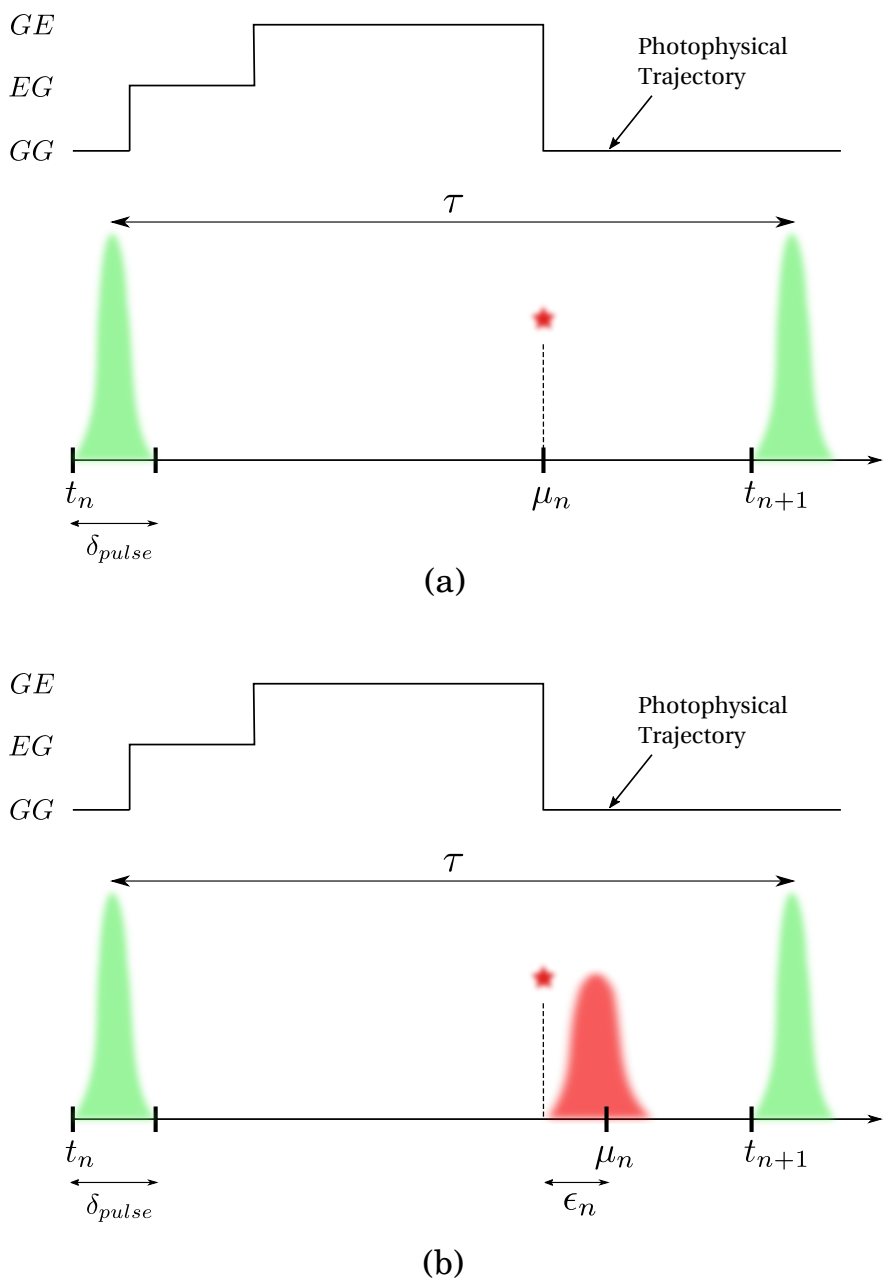

Figure 3: **Events over a pulsed illumination experiment pulse window**. Here, the beginning of the n-th interpulse window of size τ is marked by time t_n. The FRET labels originally in the state GG (donor and acceptor, respectively, in ground states) are excited by a high intensity burst (shown by the green) to the state EG (only donor excited) for a very short time δ_{pulse}. If FRET occurs, the donor transfers its energy to the acceptor and resides in the ground state leaving the FRET labels in the GE state (only acceptor excited). The acceptor then emits a photon to be registered by the detector at microtime μ_n. When using ideal detectors, the microtime is the same as the photon emission time as shown in panel (a). However, when the timing hardware has jitter (shown in red), a small delay ϵ_n is added to the microtime as shown in panel (b).

products of matrices as follows

$$L \propto \boldsymbol{\rho}_{start}\ \mathbf{Q}_1\ \mathbf{Q}_2\ \mathbf{Q}_3\ \ldots\ \ \mathbf{Q}_{N-1}\ \mathbf{Q}_N\ \boldsymbol{\rho}_{norm}^T, \tag{51}$$

where $\mathbf{Q}_n$, with $n = 1, \ldots, N$, denotes the propagator evolving the superstate during the n-th interpulse period between the $(n-1)$-th and the n-th pulse.

To derive the structure of $\mathbf{Q}_n$ during the n-th interpulse period, we break it into two portions: 1) pulse with nonzero laser intensity where the evolution of the FRET pair is described by the propagator $\mathbf{\Pi}_n^{pulse}$ introduced shortly; 2) dark period with zero illumination intensity where the evolution of the FRET pair is described by the propagator $\mathbf{\Pi}_n^{dark}$ introduced shortly. Furthermore, depending on whether a photon is detected or not over the n-th interpulse period the propagators $\mathbf{\Pi}_n^{pulse}$ and $\mathbf{\Pi}_n^{dark}$ assume different forms.

First, when no photons are detected, we have

$$\mathbf{\Pi}_n^{pulse} = \mathbf{exp}\left(\int_0^{\delta_{pulse}} d\delta\, \mathbf{G}^{non}(\delta)\right), \text{ and} \tag{52}$$

$$\mathbf{\Pi}_n^{dark} = \mathbf{exp}\left((\tau - \delta_{pulse})\, \mathbf{G}^{dark}\right), \tag{53}$$

where the integration over the pulse period now involves a time dependent $\mathbf{G}^{non}$ due to temporal variations in $\lambda_{ex}(t)$. The integral in Eq. 52 is sometimes termed the excitation IRF though we will not use this convention here. For this reason, when we say IRF, we imply detection IRF alone. Additionally, $\mathbf{G}^{dark}$ is the same as $\mathbf{G}^{non}$ except for the excitation rate that is now set to zero due to lack of illumination. Finally, the propagator for an interpulse period with no photon detection can now be written as

$$\mathbf{Q}_n = \mathbf{\Pi}_n^{pulse}\mathbf{\Pi}_n^{dark} = \mathbf{exp}\left(\int_0^{\delta_{pulse}} d\delta\, \mathbf{G}^{non}(\delta)\right)\ \mathbf{exp}\left((\tau - \delta_{pulse})\, \mathbf{G}^{dark}\right). \tag{54}$$

On the other hand, if a photon is detected sometime after a pulse (as in Fig. 3(a)), the pulse propagator remains as in Eq. 52. However, the propagator $\mathbf{\Pi}_n^{dark}$ must now be modified to include a radiative generator matrix $\mathbf{G}_n^{rad}$ similar to Eq. 20

$$\mathbf{\Pi}_n^{dark} = \mathbf{exp}\left((\mu_n - \delta_{pulse})\, \mathbf{G}^{dark}\right)\ \mathbf{G}_n^{rad}\, \mathbf{exp}\left((\tau - \mu_n)\, \mathbf{G}^{dark}\right), \tag{55}$$

where μ_n is the photon arrival time measured with respect to the n-th pulse (also termed microtime) as shown in Fig. 3(a). Here, the two exponential terms describe the evolution of the superstate before and after the photon detection during the dark period.

Moreover, we can construct the propagator for situations where a photon is detected during a pulse itself in a similar fashion. Here, the propagator $\mathbf{\Pi}_n^{dark}$ remains the same as in Eq. 53 but $\mathbf{\Pi}_n^{pulse}$ must now be modified to include the radiative generator matrix $\mathbf{G}_n^{rad}$ as

$$\mathbf{\Pi}_n^{pulse} = \mathbf{exp}\left(\int_0^{\mu_n} d\delta\, \mathbf{G}^{non}(\delta)\right)\ \mathbf{G}_n^{rad}\ \mathbf{exp}\left(\int_{\mu_n}^{\delta_{pulse}} d\delta\, \mathbf{G}^{non}(\delta)\right). \tag{56}$$

The propagators derived so far in this section assumed ideal detectors. We now describe a procedure to incorporate the IRF into this formulation. This is specially significant in accurate estimation of fluorophores' lifetimes, which is commonly done in pulse illumination

smFRET experiments. To incorporate the IRF, we follow the same procedure as in Sec. 2.4.3 and introduce convolution between the IRF function $f(\epsilon)$ and propagators above involving photon detections. That is, when there is a photon detected during the dark period, we modify the propagator $\mathbf{\Pi}_n^{dark}$ as

$$\mathbf{\Pi}_n^{dark} = \int_0^{\delta_{IRF}} d\epsilon_n \, \mathbf{exp}\left((\mu_n - \delta_{pulse} - \epsilon_n)\,\mathbf{G}^{dark}\right)\mathbf{G}_n^{rad} \times \mathbf{exp}\left((\tau - \mu_n + \epsilon_n)\,\mathbf{G}^{dark}\right) f(\epsilon_n), \tag{57}$$

while the $\mathbf{\Pi}_n^{pulse}$ stays the same as in Eq. 52. Here, ϵ_n is the stochastic delay in photon detection resulting from the IRF as shown in Fig. 3(b).

Moreover, when there is a photon detected during a pulse, the propagator $\mathbf{\Pi}_n^{pulse}$ of Eq. 56 can be modified in a similar fashion to accommodate the IRF, while the propagator $\mathbf{\Pi}_n^{dark}$ remains the same as in Eq. 53.

The propagators $\mathbf{Q}_n$ presented in this section involve integrals over large generator matrices that are analytically intractable and computationally expensive when considering large pulse numbers. Therefore, we follow a strategy similar to Sec. 2.3.8 for binned likelihood to approximate these propagators.

To reduce the complexity of the calculations, we start by making realistic approximations. Given the timescale separation between the interpulse period (typically tens of nanoseconds) and the system kinetics (typically of seconds timescale) in a pulsed illumination experiment, it is possible to approximate the system state trajectory as being constant during an interpulse period. In essence, rather than treating the system state trajectory as a continuous time process, we discretize the trajectory such that system transitions only occur at the beginning of each interpulse period. This allows us to separate the photophysical part of the generator matrix $\mathbf{G}_\psi$ in Eq. 4 from the portion describing the evolution of the system under study $\mathbf{G}_\sigma$ given in Eq. 3. Here, by contrast to the likelihood shown in Sec. 2.5.1 for pulsed illumination, we can now independently compute photophysical and system likelihood portions, as described below.

To derive the likelihood, we begin by writing the system state propagator during an interpulse period as

$$\mathbf{\Pi}_\sigma = \mathbf{exp}\left(\tau \mathbf{G}_\sigma\right), \tag{58}$$

Furthermore, we must incorporate observations into these propagators by multiplying each system transition probability in $\mathbf{\Pi}_\sigma$, $\pi_{\sigma_i \to \sigma_j}$, with the observation probability if that transition had occurred. We organize these observation probabilities using our newly defined detection matrices $\mathbf{D}_n^\sigma$ similar to Sec. 2.3.6, and write the modified propagators as

$$\mathbf{\Pi}_n^\sigma = \mathbf{\Pi}_\sigma \odot \mathbf{D}_n^\sigma, \tag{59}$$

where $\odot$ again represents the element-by-element product. Here, the elements of $\mathbf{D}_n^\sigma$ depend on the photophysical portion of the generator matrix $\mathbf{G}_\psi$ and their detailed derivations are shown in the third companion manuscript [64]. We note here that propagator matrix dimensions are now $M_\sigma \times M_\sigma$ making them computationally less expensive than in the continuous illumination case. Finally, the likelihood for the pulsed illuminated smFRET data with these new propagators reads

$$L = p(\boldsymbol{w}|\boldsymbol{\rho}_{start}, \mathbf{\Pi}_\sigma, \mathbf{G}_\psi) \propto \boldsymbol{\rho}_{start}\mathbf{\Pi}_1^\sigma\mathbf{\Pi}_2^\sigma \ldots \mathbf{\Pi}_N^\sigma \boldsymbol{\rho}_{norm}^T, \tag{60}$$

which, similar to the case of binned likelihood under continuous illumination (see Sec. 2.3.8), sums over all possible system state trajectories.

We will later use this likelihood to put forward an inverse model to learn transition probabilities (elements of $\mathbf{\Pi}_\sigma$) and photophysical transition rates appearing in $\mathbf{G}_\psi$.

2.6 Background Emissions

Here, we consider background photons registered by detectors from sources other than the labeled system under study [2]. The majority of background photons comprise ambient photons, photons from the illumination laser entering the detectors, and dark current (false photons registered by detectors) [22].

Due to the uniform laser intensity in the continuous illumination case, considered in this section, we may model all background photons using a single distribution from which waiting times are drawn. Often, such distributions are assumed (or verified) to be Exponential with fixed rates for each detection channel [65, 66]. Here we model the waiting time distribution for background photons arising from both origins as a single Exponential as is often the most common case. However, in the pulsed illumination case, laser source and the two other sources of background require different treatments due to nonuniform laser intensity. That is, the ambient photons and dark current are still modeled by an Exponential distribution though it is often further approximated as a Uniform distribution given that interpulse period if much shorter than the average background waiting time. The full formulation describing all background sources under pulsed illumination is provided in the third companion manuscript [64].

We now proceed to incorporate background into the likelihood under continuous illumination. We do so, as mentioned earlier, by assuming an Exponential distribution for the background, which effectively introduces new photophysical transitions into the model. As such, these transitions may be incorporated by expanding the full generator matrix $\mathbf{G}$ (described in Sec. 2.3) appearing in the likelihood, thereby leaving the structure of the likelihood itself intact, *c.f.*, Eq. 21.

To be clear, in constructing the new generator matrix, we treat background in each detection channel as if originating from fictitious independent emitters with constant emission rates (exponential waiting time). Furthermore, we assume that an emitter corresponding to channel i is a two state system with photophysical states denoted by

$$\{\psi_{i,1}^{bg},\ \psi_{i,2}^{bg}\}.$$

Here, each transition to the other state coincides with a photon emission with rate λ_i^{bg}. As such, the corresponding background generator matrix for channel i can now be written as

$$\mathbf{G}_i^{bg} = \begin{bmatrix} * & \lambda_{\psi_{i,1}^{bg}\to\psi_{i,2}^{bg}} \\ \lambda_{\psi_{i,2}^{bg}\to\psi_{i,1}^{bg}} & * \end{bmatrix} = \begin{bmatrix} * & \lambda_i^{bg} \\ \lambda_i^{bg} & * \end{bmatrix}.$$

Since the background emitters for each channel are independent of each other, the expanded generator matrix $\mathbf{G}$ for the combined setup (system-FRET composite plus background) can now be computed. This can be achieved by combining the system-FRET composite

state space and the background state spaces for all of the total C detection channels using Kronecker sums [67] as

$$\mathbf{G} = \mathbf{G}_{no-bg} \oplus \mathbf{G}_1^{bg} \oplus \mathbf{G}_2^{bg} \oplus \ldots \oplus \mathbf{G}_C^{bg},$$

where the symbol $\oplus$ denotes the matrix Kronecker sum, and $\mathbf{G}_{no-b}$ represents previously shown generator matrices without any background transition rates.

The propagators needed to compute the likelihood can now be obtained by exponentiating the expanded generator matrix above as

$$\begin{aligned}\mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}) = \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_{no-bg}) \otimes \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_1^{bg}) \\ \otimes\, \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_2^{bg}) \otimes \ldots \otimes \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_C^{bg}),\end{aligned}$$

where the symbol $\otimes$ denotes the matrix Kronecker product (tensor product) [67].

Furthermore, the same detection matrices defined earlier to include only nonradiative transitions or only radiative transitions, and their generalization with crosstalk and detection efficiency, can be used to obtain nonradiative and radiative propagators, as shown in Sec. 2.3.6.

Consequently, as mentioned earlier, by contrast to incorporating the effects of dead time or IRF, addition of background sources do not entail any changes in the basic structure (arrangement of propagators) of the likelihood appearing in Eq. 21.

Example VI: Background

To provide a concrete example for background, we again return to our FRET pair with two system states. The background free full generator matrix for this system-FRET composite was provided in the example box in Sec. 2.3 as (in units of ms^{-1})

$$\mathbf{G}_{no-bg} = \begin{bmatrix} -12 & 10.0 & 0 & 2.0 & 0.0 & 0.0 \\ 277000 & -347002 & 70000 & 0.0 & 2.0 & 0.0 \\ 285000 & 0.0 & -285002 & 0.0 & 0.0 & 2.0 \\ 1.0 & 0.0 & 0.0 & -11 & 10.0 & 0 \\ 0.0 & 1.0 & 0.0 & 277000 & -2777001 & 2500000 \\ 0.0 & 0.0 & 1.0 & 285000 & 0.0 & -285001 \end{bmatrix}.$$

Here, we expand the above generator matrix to incorporate background photons entering two channels ($i = 1, 2$) at rates of $\lambda_1^{bg} = 1$ ms^{-1} and $\lambda_2^{bg} = 0.5$ ms^{-1}. We do so by performing a Kronecker sum of $\mathbf{G}_{no-bg}$ with the following generator matrix for the background

$$\mathbf{G}_{bg} = \mathbf{G}_1^{bg} \oplus \mathbf{G}_2^{bg} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \oplus \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} -1.5 & 0.5 & 1 & 0 \\ 0.5 & -1.5 & 0 & 1 \\ 1 & 0 & -1.5 & 0.5 \\ 0 & 1 & 0.5 & -1.5 \end{bmatrix},$$

resulting in

$$\mathbf{G} = \mathbf{G}_{no-bg} \oplus \mathbf{G}_{bg}.$$

Here, $\mathbf{G}$ is a 24×24 matrix and we do not include its explicit from.

2.7 Fluorophore Characteristics: Quantum Yield, Blinking, Photobleaching, and Direct Acceptor Excitation

As demonstrated for background in the previous section, to incorporate new photophysical transitions such as fluorophore blinking and photobleaching into the likelihood, we must modify the full generator matrix $\mathbf{G}$. This can again be accomplished by adding extra photophysical states, relaxing nonradiatively, to the fluorophore model. These photophysical states can have long or short lifetimes depending on the specific photophysical phenomenon at hand. For example, donor photobleaching can be included by introducing a third donor photophysical state into the matrix of Eq. 4 without any escape transitions as follows

$$\mathbf{G}_{\sigma_i}^{\psi} = \begin{bmatrix} * & \lambda_{\psi_1 \to \psi_2} & 0 & 0 \\ \lambda_{\psi_2 \to \psi_1} & * & \lambda_{\sigma_i, \psi_2 \to \psi_3} & \lambda_{\psi_2 \to \psi_4} \\ \lambda_{\psi_3 \to \psi_1} & 0 & * & 0 \\ 0 & 0 & 0 & * \end{bmatrix} = \begin{bmatrix} * & \lambda_{ex} & 0 & 0 \\ \lambda_d & * & \lambda_{\sigma_i}^{FRET} & \lambda_{bleach} \\ \lambda_a & 0 & * & 0 \\ 0 & 0 & 0 & * \end{bmatrix},$$

where ψ_1 is the lowest energy combined photophysical state for the FRET labels, ψ_2 represents the excited donor, ψ_3 represents the excited acceptor, and ψ_4 represents a photobleached

donor, respectively. Additionally, λ_d and λ_a denote donor and acceptor relaxation rates, respectively, λ_{bleach} represents permanent loss of emission from the donor (photobleaching), and $\lambda_{\sigma_i}^{FRET}$ represents FRET transitions when the system is in its i-th system state.

Fluorophore blinking can be implemented similarly, except with a nonzero escape rate out of the new photophysical state, allowing the fluorophore to resume emission after some time [52, 68]. Here, assuming that the fluorophore cannot transition into the blinking photophysical state from the donor ground state results in the following generator matrix

$$\mathbf{G}_{\sigma_i}^{\psi} = \begin{bmatrix} * & \lambda_{\psi_1\to\psi_2} & 0 & 0 \\ \lambda_{\psi_2\to\psi_1} & * & \lambda_{\sigma_i,\psi_2\to\psi_3} & \lambda_{\psi_2\to\psi_4} \\ \lambda_{\psi_3\to\psi_1} & 0 & * & 0 \\ \lambda_{\psi_4\to\psi_1} & 0 & 0 & * \end{bmatrix} = \begin{bmatrix} * & \lambda_{ex} & 0 & 0 \\ \lambda_d & * & \lambda_{\sigma_i}^{FRET} & \lambda_{blink} \\ \lambda_a & 0 & * & 0 \\ \lambda_{unblink} & 0 & 0 & * \end{bmatrix}.$$

So far, we have ignored direct excitation of acceptor dyes in the likelihood model. This effect can also be incorporated into the likelihood by assigning a nonzero value to the transition rate $\lambda_{\psi_1\to\psi_3}$, that is,

$$\mathbf{G}_{\sigma_i}^{\psi} = \begin{bmatrix} * & \lambda_{\psi_1\to\psi_2} & \lambda_{\psi_1\to\psi_3} \\ \lambda_{\psi_2\to\psi_1} & * & \lambda_{\sigma_i,\psi_2\to\psi_3} \\ \lambda_{\psi_3\to\psi_1} & 0 & * \end{bmatrix} = \begin{bmatrix} * & \lambda_{ex} & \lambda_{direct} \\ \lambda_d & * & \lambda_{\sigma_i}^{FRET} \\ \lambda_a & 0 & * \end{bmatrix}.$$

Other photophysical phenomena can also be incorporated into our likelihood by following the same procedure as above. Finally, just as when adding background, the structure of the likelihood (arrangement of the propagators) when treating photophysics (including adding the effect of direct acceptor excitation) stays the same as in Eq. 21.

2.8 Synthetic Data Generation

In the previous subsections, we described how to compute the likelihood, which is the sum of probabilities over all possible superstate trajectories that could give rise to the observations made by a detector, as demonstrated in Sec. 2. Here, we demonstrate how one such superstate trajectory can be simulated to produce synthetic photon arrival data using the Gillespie algorithm [69], as described in the next section followed by the addition of detector artefacts. We then use the generated data to test our BNP-FRET sampler.

2.8.1 Gillespie and detector artefacts

The Gillespie algorithm generates two sets of random variables. The times at which superstates change (indexed 1 through N). These times occur anywhere along a continuous time grid. The next set of random variables are the states associated to the superstate preceding the time at which the superstate changes.

We designate the sequence of superstates

$$\{b_1, b_2, \ldots, b_N\},$$

where $b_n \in \{\phi_1, \phi_2, \ldots, \phi_{M_\phi}\}$. Here, unlike earlier in Sec. 2.3, the time index n on superstates b_n is not on a regular temporal grid.

Now, to generate the superstate sequence above, we first randomly draw the first superstate, b_1, from the set of possible superstates given their corresponding probabilities. Next, we draw the second superstate b_2 of the sequence using the set of transition rates out of the first state with self-transitions excluded by construction. Now, after choosing b_2, we generate the holding time h_1 (the time spent in b_1) from the Exponential distribution with rate constant associated with transitions $b_1 \to b_2$. Finally, we repeat the two previous steps to sequentially generate the full sequence of superstates along with the corresponding holding times.

More formally, we generate a trajectory, by first sampling the initial superstate as

$$b_1 \sim \mathbf{Categorical}_{\phi_{1:M_\phi}}(\boldsymbol{\rho}_{start}),$$

where $\boldsymbol{\rho}_{start}$ is the initial probability vector and the Categorical distribution is the generalization of the Bernoulli distribution for more than two possible outcomes. The remaining superstates can now be sampled as

$$b_{n+1}|b_n, \mathbf{G} \sim \mathbf{Categorical}_{\phi_{1:M_\phi}}\left(\frac{\lambda_{b_n\to\phi_1}}{\lambda_{b_n}}, \frac{\lambda_{b_n\to\phi_2}}{\lambda_{b_n}}, \ldots, \frac{\lambda_{b_n\to\phi_{M_\phi}}}{\lambda_{b_n}}\right),$$

where $\lambda_{b_n} = \sum_i \lambda_{b_n\to\phi_i}$ is the escape rate for the superstate b_n and rates for self-transitions are zero. The above equation reads as follows: "the superstate b_{n+1} is drawn (sampled) from a Categorical distribution given the superstate b_n and the generator matrix $\mathbf{G}$".

Once the n-th superstate b_n is chosen, the holding time h_n (the time spent in b_n) is sampled as follows

$$h_n|b_n, \mathbf{G} \sim \mathbf{Exponential}(\lambda_{b_n}).$$

Finally, with ideal detectors, the detection channel c_k is assigned deterministically to the k-th photon emitted at time T_k^{em}, which can be computed by summing all the holding times preceding the corresponding radiative transition.

Furthermore, in the presence of detection effects, such as crosstalk, detection efficiency, and IRF, we must add to the stochastic output of the Gillespie simulation another layer of stochasticity originating from the measurement model. That is, we stochastically assign detection channel and detection times to an emitted photon, as described below.

In the presence of crosstalk and inefficient detectors, we choose the detection channel for the k-th photon emitted upon a radiative transition as

$$c_k \sim \mathbf{Categorical}_{\{\varnothing,1,2\}}(p_\varnothing^k, p_1^k, p_2^k), \tag{61}$$

where $p_\varnothing^k$, p_1^k and p_2^k, respectively, denote the probability of the photon going undetected, being detected in channel 1 and channel 2.

Moreover, in the presence of the IRF, we assign a stochastic delay ϵ_k, sampled from a probability distribution $f(\epsilon)$, to the absolute photon emission time T_k^{em}. This results in the detection time, $T_k = T_k^{em} + \epsilon_k$, as registered by the timing hardware.

Additionally, when photophysical effects (such as blinking and photobleaching) and background are present, we can generate a superstate trajectory following the same procedure as above using the generator matrices $\mathbf{G}$ incorporating these effects as described in the previous sections.

Finally, we obtain our desired smFRET trace (see Fig. 4) consisting of photon arrival times $T_{1:K}$ and detection channels $c_{1:K}$ as

$$\{(T_1, c_1), (T_2, c_2), (T_3, c_3), \ldots, (T_K, c_K)\}.$$

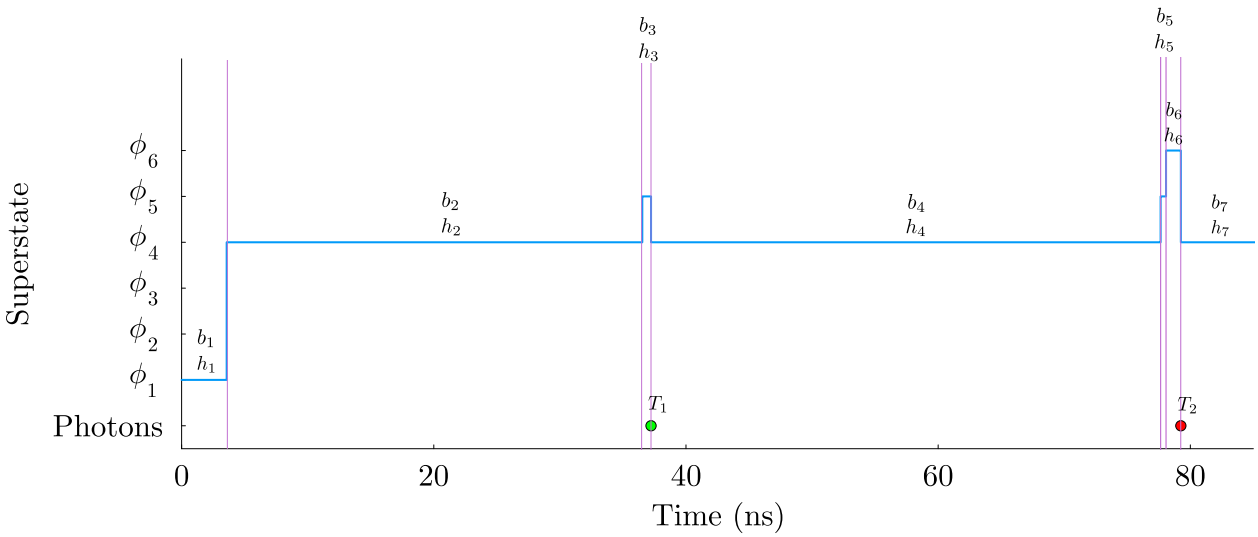

Figure 4: **Simulated data**. Here, we show a superstate trajectory (in blue) generated using Gillespie algorithm for a system-FRET composite with two system states $\sigma_{1:2}$ and three photophysical states. Collectively, superstates $\phi_{1:3}$ correspond to photophysical states when the system resides in σ_1 and superstates $\phi_{4:6}$ correspond to photophysical states when system resides in σ_2. The pink vertical lines mark the time points where transitions between the superstates occur. The variables $b_{1:7}$ and $h_{1:7}$ between each set of vertical lines represent the superstates and associated holding times, respectively. The green and red dots show the photon detections at times T_1 and T_2 in the channels 1 and 2 , respectively. The first photon is detected upon transition $b_3 \to b_4$ (or $\phi_5 \to \phi_4$), while the second photon is detected upon transition $b_6 \to b_7$ (or $\phi_6 \to \phi_4$). For this plot, we have used very fast system transition rates of $\lambda_{\sigma_1 \to \sigma_2} = 0.001\text{ns}^{-1}$ and $\lambda_{\sigma_1 \to \sigma_2} = 0.002\text{ns}^{-1}$ for demonstrative purposes only.

3 Inverse Strategy

Now, armed with the likelihood for different experimental setups and a means by which to generate synthetic data (or having experimental data at hand), we proceed to learn the parameters of interest. Assuming precalibrated detector parameters, these include transition rates that enter the generator matrix $\mathbf{G}$, and elements of $\boldsymbol{\rho}_{start}$. However, accurate estimation of the unknowns requires an inverse strategy capable of dealing with all existing sources of uncertainty in the problem such as photon's stochasticity and detector noise. This naturally leads us to adopt a Bayesian inference framework where we employ Monte Carlo methods to learn distributions over the parameters.

We begin by defining the distribution of interest over the unknown parameters we wish to learn termed the posterior. The posterior is proportional to the product of the likelihood

and prior distributions using Bayes' rule as follows

$$p(\mathbf{G}, \boldsymbol{\rho}_{start}|\boldsymbol{w}) \propto L(\boldsymbol{w}|\mathbf{G}, \boldsymbol{\rho}_{start})\, p(\mathbf{G}, \boldsymbol{\rho}_{start}), \tag{62}$$

where the last term $p(\mathbf{G}, \boldsymbol{\rho}_{start})$ is the joint prior distribution over $\mathbf{G}$ and $\boldsymbol{\rho}_{start}$ defined over the same domains as the parameters. The prior is often selected on the basis of computational convenience. The influence of the prior distribution on the posterior diminishes as more data is incorporated through the likelihood. Furthermore, the constant of proportionality is the inverse of the absolute probability of the collected data, $1/p(\boldsymbol{w})$, and can be safely ignored as generation of Monte Carlo samples only involves ratios of posterior distributions or likelihoods.

Additionally, the ϵ^K factor in the likelihood first derived in Eq. 21 can be absorbed into the proportionality constant as it does not depend on any of the parameters of interest, resulting in the following expression for the posterior (in the absence of detector dead time and IRF for simplicity)

$$p(\mathbf{G}, \boldsymbol{\rho}_{start}|\boldsymbol{w}) \propto \boldsymbol{\rho}_{start}\ \boldsymbol{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \boldsymbol{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \boldsymbol{\Pi}_{K-1}^{non}\mathbf{G}_{K-1}^{rad}\ \boldsymbol{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \boldsymbol{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{end}^T \times p(\mathbf{G}, \boldsymbol{\rho}_{start}). \tag{63}$$

Next, assuming *a priori* that different transition rates are independent of each other and initial probabilities, we can simplify the prior as follows

$$p(\mathbf{G}, \boldsymbol{\rho}_{start}) = p(\boldsymbol{\rho}_{start}) \prod_{i,j} p(\lambda_{\phi_i \to \phi_j}), \tag{64}$$

where we select the Dirichlet prior distribution over initial probabilities as this prior is conveniently defined over a domain where the probability vectors, drawn from it, sum to unity. That is,

$$p(\boldsymbol{\rho}_{start}) = \mathbf{Dirichlet}(\boldsymbol{\zeta}), \tag{65}$$

where the Dirichlet distribution is a multivariate generalization of the Beta distribution and $\boldsymbol{\zeta}$ is a vector of the same size as the superstate space. Typically parameters of the prior are termed hyperparameters and as such $\boldsymbol{\zeta}$ collects as many hyperparameters as its size.

Additionally, we select Gamma prior distributions for individual rates. That is,

$$p(\lambda_{\phi_i \to \phi_j}) = \mathbf{Gamma}\left(\lambda_{\phi_i \to \phi_j}; \alpha, \frac{\lambda_{ref}}{\alpha}\right), \tag{66}$$

guaranteeing positive values. Here, α and λ_{ref} (a reference rate parameter) are hyperparameters of the Gamma prior. For simplicity, these hyperparameters are usually chosen (with appropriate units) such that the prior distributions are very broad, minimizing their influence on the posterior.

Furthermore, to reduce computational cost, the number of parameters we need to learn can be reduced by reasonably assuming the system was at steady state immediately preceding the time at which the experiment began. That is, instead of sampling $\boldsymbol{\rho}_{start}$ from the posterior, we compute $\boldsymbol{\rho}_{start}$ by solving the time-independent master equation,

$$\boldsymbol{\rho}_{start}\mathbf{G} = 0.$$

Therefore, the posterior in Eq. 63 now reduces to

$$p(\mathbf{G}|\boldsymbol{w}) \propto \boldsymbol{\rho}_{start}\ \boldsymbol{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \boldsymbol{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \boldsymbol{\Pi}_{K-1}^{non}\mathbf{G}_{K-1}^{rad}\ \boldsymbol{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \boldsymbol{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{end}^T \times p(\mathbf{G}). \tag{67}$$

In the following subsections, we first describe a parametric inverse strategy, *i.e.*, assuming a known number of system states, for sampling parameters from the posterior distribution in Eq. 67 using Monte Carlo methods. Next, we generalize this inverse strategy to a non-parametric case where we also deduce the number of system states.

3.1 Parametric Sampler: BNP-FRET with Fixed Number of System States

Now with the posterior, Eq. 67, at hand and assuming steady state $\boldsymbol{\rho}_{start}$, here we illustrate a sampling scheme to deduce the transition rates of the generator matrix $\mathbf{G}$.

As our posterior of Eq. 67 does not assume a standard form amenable to analytical calculations, we must iteratively draw numerical samples of the transition rates within $\mathbf{G}$ using Markov Chain Monte Carlo (MCMC) techniques. Specifically, we adopt a Gibbs algorithm to, sequentially and separately, generate samples for individual transition rates at each MCMC iteration. To do so, we first write the posterior of Eq. 67 using the chain rule as follows

$$p(\mathbf{G}|\boldsymbol{w}) = p(\lambda_{\phi_i\to\phi_j}|\mathbf{G}\backslash\lambda_{\phi_i\to\phi_j}, \boldsymbol{w})p(\mathbf{G}\backslash\lambda_{\phi_i\to\phi_j}|\boldsymbol{w}), \tag{68}$$

where the backslash after $\mathbf{G}$ indicates exclusion of the following rate parameters and $\boldsymbol{w}$ denotes the set of observations as introduced in Sec. 2.3.2. Here, the first term on the right hand side is the conditional posterior for the individual rate $\lambda_{\phi_i\to\phi_j}$. The second term is considered a constant in the corresponding Gibbs step as it does not depend on $\lambda_{\phi_i\to\phi_j}$. Moreover, following the same logic, the priors $p(\mathbf{G}\backslash\lambda_{\phi_i\to\phi_j})$ (see Eq. 67) for the remaining rate parameters in the posterior on the left are also considered constant. Therefore, from Eqs. 67 & 68, we can write the conditional posterior for $\lambda_{\phi_i\to\phi_j}$ above as

$$p(\lambda_{\phi_i\to\phi_j}|\mathbf{G}\backslash\lambda_{\phi_i\to\phi_j}, \boldsymbol{w}) \propto \boldsymbol{\rho}_{start}\ \boldsymbol{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \boldsymbol{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \boldsymbol{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \boldsymbol{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{end}^T \times \mathbf{Gamma}\left(\lambda_{\phi_i\to\phi_j}; \alpha, \frac{\lambda_{ref}}{\alpha}\right). \tag{69}$$

Just as with the posterior over all parameters, this conditional posterior shown above does not take a closed form allowing for direct sampling.

As such, we turn to the Metropolis-Hastings (MH) algorithm [70] to draw samples from this conditional posterior, where new samples are drawn from a proposal distribution q and accepted with probability

$$\alpha(\lambda^*_{\phi_i\to\phi_j}, \lambda_{\phi_i\to\phi_j}) = \mathbf{min}\left\{1, \frac{p(\lambda^*_{\phi_i\to\phi_j}|\boldsymbol{w}, \mathbf{G}\backslash\lambda_{\phi_i\to\phi_j})\ q(\lambda_{\phi_i\to\phi_j}|\lambda^*_{\phi_i\to\phi_j})}{p(\lambda_{\phi_i\to\phi_j}|\boldsymbol{w}, \mathbf{G}\backslash\lambda_{\phi_i\to\phi_j})\ q(\lambda^*_{\phi_i\to\phi_j}|\lambda_{\phi_i\to\phi_j})}\right\}, \tag{70}$$

where the asterisk represents the proposed rate values from the proposal distribution q.

To construct an MCMC chain of samples, we begin by initializing the chain for each transition rate $\lambda_{\phi_i \to \phi_j}$, by random values drawn from the corresponding prior distributions. We then iteratively sweep the whole set of transition rates in each MCMC iteration by drawing new values from the proposal distribution q.

A computationally convenient choice for the proposal is a Normal distribution leading to a simpler acceptance probability in Eq. 70. This is due to its symmetry resulting in $q(\lambda_{\phi_i \to \phi_j} | \lambda^*_{\phi_i \to \phi_j}) = q(\lambda^*_{\phi_i \to \phi_j} | \lambda_{\phi_i \to \phi_j})$. However, a Normal proposal distribution would allow negative transition rates naturally forbidden leading to rejection in the MH step and thus inefficient sampling. Therefore, it is convenient to propose new samples either drawn from a Gamma distribution or, as shown below, from a Normal distribution in logarithmic space to allow for exploration along the full real line as follows

$$\log(\lambda^*_{\phi_i \to \phi_j}/\kappa) \,\Big|\, \log(\lambda_{\phi_i \to \phi_j}/\kappa), \sigma^2 \sim \textbf{Normal}\left(\log(\lambda_{\phi_i \to \phi_j}/\kappa), \sigma^2\right),$$

where $\kappa = 1$ is an auxiliary parameter in the same units as $\lambda_{\phi_i \to \phi_j}$ introduced to obtain a dimensionless quantity within the logarithm.

The variable transformation above now requires introduction of Jacobian factors in the acceptance probability as follows

$$\alpha(\lambda^*_{\phi_i \to \phi_j}, \lambda_{\phi_i \to \phi_j}) = \textbf{min}\left\{1, \frac{p(\lambda^*_{\phi_i \to \phi_j} | \boldsymbol{w}, \mathbf{G} \backslash \lambda_{\phi_i \to \phi_j})}{p(\lambda_{\phi_i \to \phi_j} | \boldsymbol{w}, \mathbf{G} \backslash \lambda_{\phi_i \to \phi_j})} \frac{(\partial \log(\lambda_{\phi_i \to \phi_j}/\kappa)/\partial \lambda_{\phi_i \to \phi_j})}{(\partial \log(\lambda_{\phi_i \to \phi_j}/\kappa)/\partial \lambda_{\phi_i \to \phi_j})^*}\right\},$$

where the derivative terms represent the Jacobian and the proposal distributions are canceled by virtue of using a symmetric Normal distribution.

The acceptance probability above depends on the difference of the current and proposed values for a given transitions rate. In other words, smaller differences between the current and proposed values often lead to larger acceptance probabilities. This difference is determined by the covariance of the Normal proposal distribution σ^2 which needs to be tuned for each rate individually to achieve optimal performance of the BNP-FRET sampler, or very approximately, one-fourth acceptance rate for the proposals [71].

This whole algorithm can now be summarized in the following pseudocode

```
# Initialize Chain of Samples
j = 1
for i = 1 : M_σ × M_σ
```

$$\lambda_i^{(j)} \sim \textbf{Gamma}\left(\alpha, \frac{\lambda_{ref}}{\alpha}\right)$$

```
end
```

```
# Iteratively sample from the posterior using Gibbs algorithm
for j = 2 : Draws
 for i = 1 : M_σ × M_σ

        # Propose new sample
        log(λ_i^*) ~ Normal(log(λ_i^(j−1)), σ^2),

        # Compute acceptance probability
        α(λ_i^*, λ_i^(j−1)) = min{1, p(λ_i^*|w, G\λ_i)/p(λ_i^(j−1)|w, G\λ_i) (∂ log(λ_i)/∂λ_i)^(j−1)/(∂ log(λ_i)/∂λ_i)^*}

        if α(λ_i^*, λ_i^(j−1)) > rand()
         # Accept proposal
          λ_i^(j) = exp(λ_i^*)
        else
         # Reject proposal
          λ_i^(j) = λ_i^(j−1)
        end

 end
end
```

3.2 Nonparametrics: Predicting the Number of System States

After describing our inverse strategy for a known number of system states (*i.e.*, parametric inference), we turn to more realistic scenarios where we may not know the number of system states which, in turn, leads to an unknown number of superstates (*i.e.*, nonparametric inference). In the following subsections, we first describe the BNP framework for continuous illumination and then proceed to illustrate our BNP strategy under pulsed illumination. Such BNP frameworks introduced herein, eventually, provide us with distributions over the number of system states simultaneously, and self-consistently, with other model parameters.

3.2.1 Bernoulli process for continuous illumination

The number of system states is often unknown and cannot *a priori* be set by hand. Therefore, in order to learn the states warranted by the data, we turn to the Bayesian nonparametric paradigm. That is, we first define an infinite-dimensional version of the generator matrix in Eq. 2 and multiply each of its elements by a Bernoulli random variable b_i (also termed loads). These loads, indexed by i, allow us to turn on/off portions of the generator matrix

associated with transitions between specific system states (including self-transitions). We can write the nonparametric generator matrix as follows

$$\mathbf{G} = \begin{bmatrix} b_1^2 \mathbf{G}_{\sigma_1}^{\psi} - \sum_{j \neq 1} b_1 b_j \lambda_{\sigma_1 \to \sigma_j} \mathbb{I} & b_1 b_2 \lambda_{\sigma_1 \to \sigma_2} \mathbb{I} & \dots \\ b_2 b_1 \lambda_{\sigma_2 \to \sigma_1} \mathbb{I} & b_2^2 \mathbf{G}_{\sigma_2}^{\psi} - \sum_{j \neq 2} b_2 b_j \lambda_{\sigma_2 \to \sigma_j} \mathbb{I} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} * & b_1^2 \lambda_{\psi_1 \to \psi_2} & b_1^2 \lambda_{\psi_1 \to \psi_3} & b_1 b_2 \lambda_{\sigma_1 \to \sigma_2} & 0 & 0 & \dots \\ b_1^2 \lambda_{\psi_2 \to \psi_1} & * & b_1^2 \lambda_{\psi_2 \to \psi_3}^{(1)} & 0 & b_1 b_2 \lambda_{\sigma_1 \to \sigma_2} & 0 & \dots \\ b_1^2 \lambda_{\psi_3 \to \psi_1} & b_1^2 \lambda_{\psi_3 \to \psi_2} & * & 0 & 0 & b_1 b_2 \lambda_{\sigma_1 \to \sigma_2} & \dots \\ b_1 b_2 \lambda_{\sigma_2 \to \sigma_1} & 0 & 0 & * & b_2^2 \lambda_{\psi_1 \to \psi_2} & b_2^2 \lambda_{\psi_1 \to \psi_3} & \dots \\ 0 & b_1 b_2 \lambda_{\sigma_2 \to \sigma_1} & 0 & b_2^2 \lambda_{\psi_2 \to \psi_1} & * & b_2^2 \lambda_{\psi_2 \to \psi_3}^{(2)} & \dots \\ 0 & 0 & b_1 b_2 \lambda_{\sigma_2 \to \sigma_1} & b_2^2 \lambda_{\psi_3 \to \psi_1} & b_2^2 \lambda_{\psi_3 \to \psi_2} & * & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where a load value of 1 represents an "active" system state, while "inactive" system states (not warranted by the data) get a load value of 0. Here, there are two loads associated to every transition because there is a pair of states corresponding to each transition. Within this formalism, the number of active loads is the number of system states estimated by the BNP-FRET sampler. As before, $*$ represents negative row-sums.

The full set of loads, $\mathbf{b} = \{b_1, b_2, \dots, b_\infty\}$, now become quantities we wish to learn. In order to leverage Bayesian inference methods to learn the loads, the previously defined posterior distribution (Eq. 67) now reads as follows

$$p(\mathbf{b}, \mathbf{G} | \boldsymbol{w}) \propto L(\boldsymbol{w} | \mathbf{b}, \mathbf{G}, \boldsymbol{\rho}_{start})\, p(\mathbf{G}) p(\mathbf{b}), \tag{71}$$

where the prior $p(\mathbf{b})$ is Bernoulli while the remaining prior, $p(\mathbf{G})$, can be assumed to be the same as in Eq. 66.

As in the case of the parametric BNP-FRET sampler presented in Sec. 3.1, we generate samples from this nonparametric posterior employing a similar Gibbs algorithm. To do so, we first initialize the MCMC chains of loads and rates by taking random values from their priors. Next, to construct the MCMC chains, we iteratively draw samples from the posterior in two steps: 1) sequentially sample all rates using the MH algorithm; then 2) loads by direct sampling, one-by-one from their corresponding conditional posteriors. Here, step (1) is identical to the parametric case of Sec. 3.1 and we only focus on the second step in what follows.

To sample the i-th load, the corresponding conditional posterior reads [72]

$$p(b_i | \mathbf{b} \backslash b_i, \mathbf{G}, \boldsymbol{w}) \propto L(\boldsymbol{w} | \mathbf{b}, \mathbf{G}, \boldsymbol{\rho}_{start})\ \textbf{Bernoulli}\left(b_i; \frac{1}{1 + \frac{M_\sigma^{max} - 1}{\gamma}}\right), \tag{72}$$

where the backslash after $\mathbf{b}$ indicates exclusion of the following load and M_σ^{max} and γ are hyperparameters. Here γ sets the *a priori* expected number of system states.

A note on the interpretation of M_σ^{max} is in order. When dealing with nonparametrics, we nominally must consider an infinite set of loads and priors for these loads called Bernoulli process priors [73]. Samplers for such process priors are available though inefficient [74, 75]. However, for computational convenience, it is possible to introduce a large albeit finite number of loads set to M_σ^{max}. It can be shown that parameter inference are unaffected by this choice of cutoff [73, 76, 77] when setting the success probability to $1/(1+\frac{M_\sigma^{max}-1}{\gamma})$ as in the Bernoulli distribution of Eq. 72. This is because such a choice forces the mean (expected number of system states) of the full prior on loads $\prod_i p(b_i)$ to be finite ($=\gamma$).

Since the conditional posterior in the equation above must be discrete and describes probabilities of the load being either active or inactive, it must itself follow a Bernoulli distribution with updated parameters

$$p(b_i|\mathbf{b}\backslash b_i, \mathbf{G}, \boldsymbol{w}) = \mathbf{Bernoulli}(b_i; q_i),$$

where

$$q_i = \frac{L(\boldsymbol{w}|b_i=1, \mathbf{b}\backslash b_i, \mathbf{G}, \boldsymbol{\rho}_{start})}{L(\boldsymbol{w}|b_i=1, \mathbf{b}\backslash b_i, \mathbf{G}, \boldsymbol{\rho}_{start}) + L(\boldsymbol{w}|b_i=0, \mathbf{b}\backslash b_i, \mathbf{G}, \boldsymbol{\rho}_{start})}.$$

The Bernoulli form of this posterior allows direct sampling of the loads.

We will apply this method for synthetic and experimental data in the second companion manuscript of this series [78].

3.2.2 iHMM methods for pulsed illumination

Under pulsed illumination, the Bernoulli process prior described earlier for continuous illumination can in principle be used as is to estimate the number of system states and the transition rates. However, in this section, we will describe a computationally cheaper inference strategy applicable to the simplified likelihood of Eq. 60 assuming system state transitions occurring only at the beginning of each pulse. The reduction in computational expense is achieved by directly learning the elements of the propagator $\mathbf{\Pi}_\sigma$ of Eq. 58, identical for all interpulse periods. In doing so, we learn transition probabilities for the system states instead of learning rates though we will continue learning rates for photophysical states. This avoids expensive matrix exponentials for potentially large system state numbers required for computing the propagators under continuous illumination.

Now, to infer the transition probabilities in $\mathbf{\Pi}_\sigma$, the dimensions of which are unknown owing to an unknown number of system states, as well as transition rates among the photophysical states (elements of $\mathbf{G}_\psi$ in Eq. 4), and initial probabilities, we must place suitable priors on these parameters yielding the following posterior

$$p(\boldsymbol{\rho}_{start}, \mathbf{\Pi}_\sigma, \mathbf{G}_\psi|\boldsymbol{w}) \propto p(\boldsymbol{w}|\boldsymbol{\rho}_{start}, \mathbf{\Pi}_\sigma, \mathbf{G}_\psi)p(\boldsymbol{\rho}_{start})p(\mathbf{G}_\psi)p(\mathbf{\Pi}_\sigma), \tag{73}$$

where we have immediately written the joint prior as a product prior over $\boldsymbol{\rho}_{start}$, $\mathbf{G}_\psi$, and $\mathbf{\Pi}_\sigma$. Next, for $\boldsymbol{\rho}_{start}$ and $\mathbf{G}_\psi$ we use the same priors as in Eqs. 65-66. However, as the number of system states is unknown, $\mathbf{\Pi}_\sigma$ requires special treatment. To learn $\mathbf{\Pi}_\sigma$, it is convenient to adopt the infinite HMM (iHMM) [41, 48] due to the discrete nature of system state transitions over time.

As the name suggests, the iHMM leverages infinite system state spaces ($M_\sigma \to \infty$ in Eq. 3) similar to the Bernoulli process prior of Sec. 3.2.1. However, unlike the Bernoulli process, all system states remain permanently active. The primary goal of an iHMM is then to infer transition probabilities between system states, some of which, not warranted by the data, remain very small and set by the (nonparametric) prior that we turn to shortly. Thus the effective number of system states can be enumerated from the most frequently visited system states over the course of a learned trajectory.

Within this iHMM framework, we place an infinite dimensional version of the Dirichlet prior, termed the Dirichlet process prior [41, 48, 79], as priors over each row of the propagator $\mathbf{\Pi}_\sigma$. That is,

$$\boldsymbol{\pi}_m \sim \mathbf{DirichletProcess}(\alpha\boldsymbol{\beta}), \qquad m = 1, 2, \ldots, \tag{74}$$

where $\boldsymbol{\pi}_m$ is the mth row of $\mathbf{\Pi}_\sigma$. Here, the hyperparameters of the Dirichlet process prior include the concentration parameter α which determines the sparsity of the $\boldsymbol{\pi}_m$ and the hyper parameter $\boldsymbol{\beta}$ which is a probability vector, also known as base distribution. Together $\alpha\boldsymbol{\beta}$ are related to the $\boldsymbol{\zeta}$ introduced earlier for the (finite) Dirichlet distribution of Eq. 65.

Next, as the base distribution itself is unknown and all transitions out of each state should be likely to revisit the same set of states, we must place the same base distribution on all Dirichlet process priors placed on the rows of the propagator. To sample this unique base, we again choose a Dirichlet process prior [41, 80–82], that is,

$$\boldsymbol{\beta} \sim \mathbf{DirichletProcess}(\xi\boldsymbol{\gamma}),$$

where we may set $\xi = 1$ and $\boldsymbol{\gamma}$ is a vector of hyperparameters of size M_σ.

Now, to deduce the unknown parameters, we need to draw samples from the posterior in Eq. 73. However, due to the nonanalytical form of the posterior we cannot jointly sample our posterior. Thus, as before, we adopt a Gibbs sampling strategy to sequentially and separately draw samples for each parameter. Here, we only illustrate our Gibbs sampling step for the transition probabilities $\boldsymbol{\pi}_m$. Our Gibbs steps for the remaining parameters are similar to Sec. 3.1. The complete procedure is described in the third companion manuscript [64].

Similar to the Bernoulli process prior, there are two common approaches to draw samples within the iHMM framework: slice sampling using the exact Dirichlet process prior and finite truncation [41, 48, 83, 84]. Just as before for the case of continuous illumination, we truncate the Dirichlet process prior to a finite Dirichlet distribution and fix its dimensionality to a finite (albeit large) number which we set to M_σ^{max} to improve the sampling. It can then be shown that for large enough M_σ^{max} the number of system states visited becomes independent of M_σ^{max} [72].

As before, to numerically sample the transition probabilities $\boldsymbol{\pi}_m$ from our full posterior in Eq. 73 through MCMC, we choose our initial samples from the priors

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathbf{Dirichlet}(\xi\boldsymbol{\gamma}), \\ \boldsymbol{\pi}_m &\sim \mathbf{Dirichlet}(\alpha\boldsymbol{\beta}), \qquad m = 1, 2, \ldots, M_\sigma^{max} \end{aligned}$$

where we chose elements of $\boldsymbol{\gamma}$ to be $1/M_\sigma^{max}$ to ascribe similar weights across the state space *a priori*.

3.3 Likelihood Computation in Practice

As shown in Sec. 2.5.1, the likelihood typically takes the following generic form

$$L \propto \boldsymbol{\rho}_{start}\ \mathbf{Q}_1\ \mathbf{Q}_2\ \mathbf{Q}_3\ \ldots\ \ \mathbf{Q}_{K-1}\ \mathbf{Q}_K\ \boldsymbol{\rho}_{norm}^T, \tag{75}$$

where $\mathbf{Q}_i$ are matrices whose exact form depends on which effects we incorporate into our likelihood. Computing this last expression would typically lead to underflow as likelihood values quickly drop below floating-point precision.

For this reason, it is convenient to introduce the logarithm of this likelihood. In order to derive the logarithm of the likelihood of Eq. 75, we rewrite the likelihood as a product of multiple terms as follows

$$L \propto (\boldsymbol{\rho}_{start}\ \boldsymbol{\rho}_{norm}^T)\ (\boldsymbol{\rho}_1\ \mathbf{Q}_1\ \boldsymbol{\rho}_{norm}^T)\ (\boldsymbol{\rho}_2\ \mathbf{Q}_2\ \boldsymbol{\rho}_{norm}^T)\ \ldots\ \ (\boldsymbol{\rho}_{K-1}\ \mathbf{Q}_{K-1}\ \boldsymbol{\rho}_{norm}^T)\ (\boldsymbol{\rho}_K\ \mathbf{Q}_K\ \boldsymbol{\rho}_{norm}^T),$$

where $\boldsymbol{\rho}_i$ are the normalized probability vectors given by the following recursive formula

$$\begin{aligned}\boldsymbol{\rho}_1 &= \boldsymbol{\rho}_{start}, \text{ and}\\ \boldsymbol{\rho}_i &= \frac{\boldsymbol{\rho}_{i-1}\ \mathbf{Q}_{i-1}}{(\boldsymbol{\rho}_{i-1}\ \mathbf{Q}_{i-1}\ \boldsymbol{\rho}_{norm}^T)}.\end{aligned}$$

Now, using the recursion relation above, the log-likelihood can be written as

$$\begin{aligned}\log(L) =& \log(\boldsymbol{\rho}_{start}\ \boldsymbol{\rho}_{end}^T) + \log(\boldsymbol{\rho}_1\ \mathbf{Q}_1\ \boldsymbol{\rho}_{norm}^T) + \log(\boldsymbol{\rho}_2\ \mathbf{Q}_2\ \boldsymbol{\rho}_{norm}^T) + \log(\boldsymbol{\rho}_3\ \mathbf{Q}_3\ \boldsymbol{\rho}_{norm}^T) + \ldots\\ & \log(\boldsymbol{\rho}_{K-1}\ \mathbf{Q}_{K-1}\ \boldsymbol{\rho}_{norm}^T) + \log(\boldsymbol{\rho}_K\ \mathbf{Q}_K\ \boldsymbol{\rho}_{norm}^T) + const,\end{aligned}$$

where *const* is a constant.
Note that $\boldsymbol{\rho}_{start}\ \boldsymbol{\rho}_{norm}^T = 1$. The pseudocode to compute the log-likelihood is as follows

$\boldsymbol{\rho} = \boldsymbol{\rho}_{start}$
$p = \text{sum}(\boldsymbol{\rho}) = 1$
$\log(L) = \log(p) = 0$

for $i = 1 : K$

$$\begin{aligned}\mathbf{Q} &= \ldots\\ \boldsymbol{\rho} &= \boldsymbol{\rho}\mathbf{Q}\\ p &= \text{sum}(\boldsymbol{\rho})\\ \log(L) &= \log(L) + \log(p)\\ \boldsymbol{\rho} &= \boldsymbol{\rho}/p\end{aligned}$$

end
return $\log(L)$

4 Results

In this section, we present results using our BNP-FRET sampler described above. Specifically, here we benchmark the parametric (*i.e.*, fixed number of system states) version of our sampler using synthetic data, while the two subsequent manuscripts [64, 78] focus on the nonparametric (*i.e.*, unknown number of system states) analysis of experimental data.

For simplicity alone, we begin by analyzing data from an idealized system with two system states using different photon budgets and excitation rates. Next, we consider more realistic examples incorporating the following one at a time: 1) crosstalk and detection efficiency; 2) background emission; 3) IRF; and then 4) end with a brief discussion on the unknown number of system states. We demonstrate when these features become relevant, as well as the overall robustness and generality of the BNP-FRET sampler.

For now, we assume continuous illumination for all parametric examples and use the following priors for the analyses. The prior used for the FRET rates are

$$\lambda_{\sigma_i}^{FRET} \sim \mathbf{Gamma}(1, 1\,\mathrm{ns}^{-1}),$$

and use the following prior over the system transition rates

$$\lambda_{\sigma_i \to \sigma_j} \sim \mathbf{Gamma}(1, 10^{-6}\,\mathrm{ns}^{-1}).$$

As discussed earlier in Sec. 3.1, it is more convenient to work within logarithmic space where we use the following proposal distributions to update the parameter values

$$\begin{aligned} \log(\lambda_{ex}^{*}) | \log(\lambda_{ex}), \sigma_{ex} &\sim \mathbf{Normal}(\log(\lambda_{ex}), \sigma_{ex}^2), \\ \log(\lambda_{\sigma_i}^{FRET*}) | \log(\lambda_{\sigma_i}^{FRET}), \sigma_{FRET} &\sim \mathbf{Normal}(\log(\lambda_{\sigma_i}^{FRET}), \sigma_{FRET}^2), \text{ and} \\ \log(\lambda_{\sigma_i \to \sigma_j}^{*}) | \log(\lambda_{\sigma_i \to \sigma_j}), \sigma_{sys} &\sim \mathbf{Normal}(\log(\lambda_{\sigma_i \to \sigma_j}), \sigma_{sys}^2), \end{aligned}$$

where $*$ denotes proposed rates and where it is understood that all rates appearing in the logarithm have been divided through by a unit constant in order for the argument of the logarithm to remain dimensionless.

For efficient exploration of the parameter space by the BNP-FRET sampler and upon extensive experimentation with acceptance ratios, we found it prudent to alternate between two sets of variances, $\{\sigma_{ex}^2 = 10^{-5}, \sigma_{FRET}^2 = 0.01, \sigma_{sys}^2 = 0.1\}$ and $\{\sigma_{ex}^2 = 10^{-5}, \sigma_{FRET}^2 = 0.5, \sigma_{sys}^2 = 5.0\}$ to generate an MCMC chain. This ensures that we propose samples of different orders of magnitude. As an intuitive guide, the more data we have, the sharper we expect our posterior over our rates to be and thus, the smaller both variances should be in our proposal distributions.

In the examples presented in the next few subsections, for computational simplicity, we fix the escape rates for the donor and acceptor excited photophysical states as well as the background rates for each detection channel in our simulations, as they can be precalibrated from experiments.

4.1 Parametric Examples

4.1.1 Photon budget and excitation rate

Here, we perform Bayesian analysis on synthetically generated data (as described in Sec. 2.8) for the simplest case where the number of system states is an input to the BNP-FRET

sampler. To generate data, we use the following generator matrix

$$\mathbf{G} = \begin{bmatrix} * & 10.0 & 0 & 2.0 & 0.0 & 0.0 \\ 2.77 \times 10^5 & * & 1.11 \times 10^5 & 0.0 & 2.0 & 0.0 \\ 2.85 \times 10^5 & 0.0 & * & 0.0 & 0.0 & 2.0 \\ 1.0 & 0.0 & 0.0 & * & 10.0 & 0 \\ 0.0 & 1.0 & 0.0 & 2.77 \times 10^5 & * & 0.91 \times 10^5 \\ 0.0 & 0.0 & 1.0 & 2.85 \times 10^5 & 0.0 & * \end{bmatrix} \text{ms}^{-1},$$

where the elements are motivated from real experiments [85]. Using this generator matrix, we generated a superstate trajectory as described in Sec. 2.8. We analyzed 430000 photons from the generated data using our BNP-FRET sampler. The resulting posterior distribution for transitions between system states and FRET efficiencies (computed as $\varepsilon_{\sigma_i}^{FRET} = \lambda_{\sigma_i}^{FRET}/(\lambda_d + \lambda_{\sigma_i}^{FRET})$ for the i-th system state) is shown in Fig. 5. As we will see for all examples, the finiteness of data always leads to some error as evident from the slight offset of the peaks of the distribution from the ground truth.

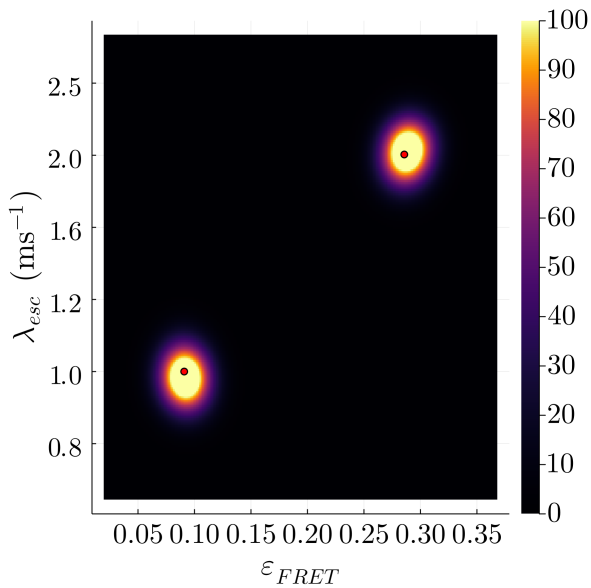

Figure 5: **Learned bivariate posterior for the system state escape rates λ_{esc} and FRET efficiencies ε_{FRET} given synthetic data.** To produce this plot, we analyzed synthetic data generated using an excitation rate of $\lambda_{ex} = 10$ ms^{-1}, and escape rates $\lambda_{esc} = 1$ & 2 ms^{-1} and FRET efficiencies of 0.09 & 0.29 for the two system states, respectively. The ground truth is shown with red dots. The FRET efficiencies estimated by our sampler are $0.288^{+0.007}_{-0.006}$ and $0.092^{+0.003}_{-0.003}$. Furthermore, predicted escape rates are $2.03^{+0.16}_{-0.17}$ ms^{-1} and $0.98^{+0.10}_{-0.07}$ ms^{-1}. The small bias away from the ground truth is due to the finiteness of data. We have smoothed the distributions, for illustrative purposes only, using kernel density estimation (KDE) available through the Julia Plots package.

The effects of limited photon budget become significant especially when system kinetics occur across multiple timescales with the most photon starved state characterized by the

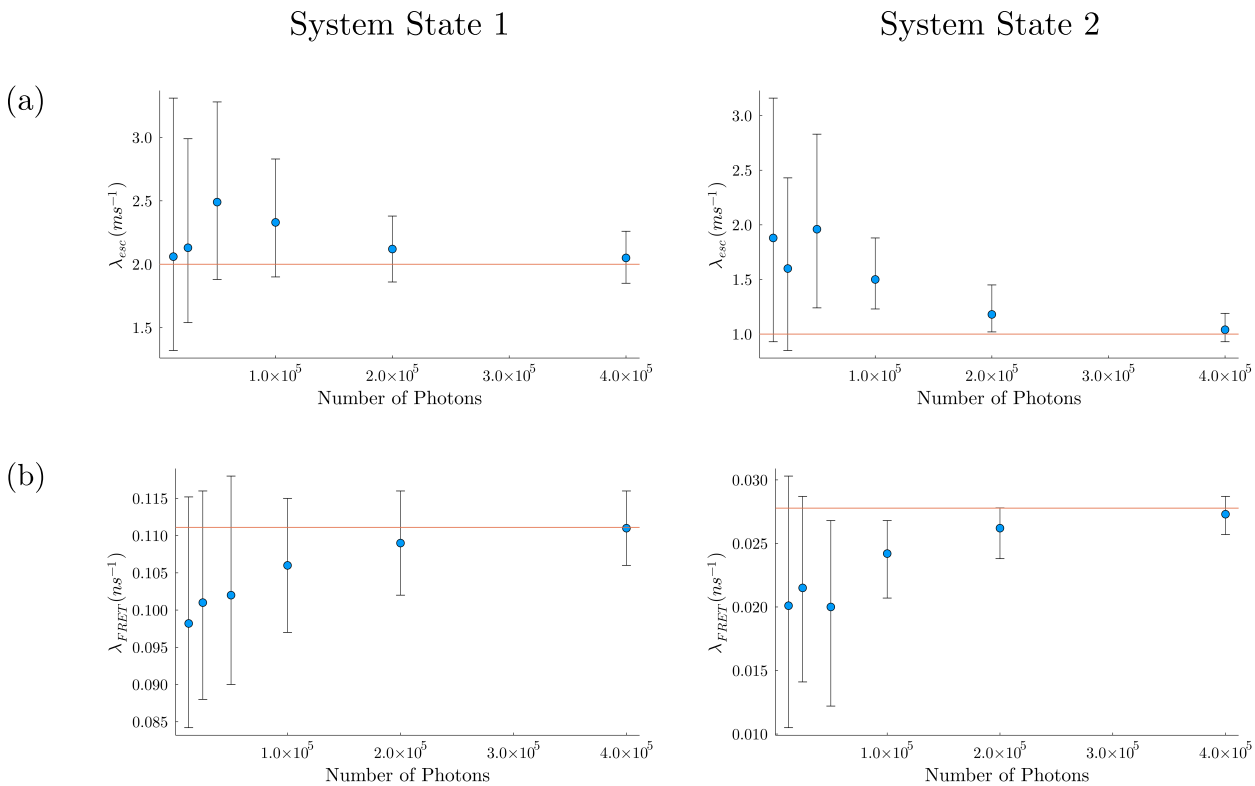

Figure 6: **System and FRET transition rates as functions of the number of photons used for analysis.** To produce these plots using the same kinetic parameters as in Fig. 5. Next, we analyzed the data considering only the first 12500 photons and then increased the photon budget by a factor of two for each subsequent analysis. Furthermore, we generated 9000 MCMC samples for each analysis to compute statistical quantities. In panel (a), we show two plots corresponding to the two system transition rates (λ_{esc}). The blue dots represent the median values (50% quantile), and the ends of the attached confidence intervals represent 5% and 95% quantiles. The ground truths are shown with red horizontal lines. We show similar plots for FRET transition rates (λ_{FRET}) in panel (b). In all of the plots, we see that as the photon budget is increased, the confidence intervals shrink (the posterior gets narrower/sharper). With a budget of 400000 photons, the confidence intervals represent less than 10% error in the estimates and contain the ground truths in all of the plots.

largest escape rate. In this case, it is useful to quantify how many photons are typically required to assess any escape rate (with the fastest setting the lower photon count bound needed) to obtain below 15% error in parameter estimates.

To quantify the number of photons, ignoring background and detector effects, we define a dimensionless quantity which we call the "photon budget index" predicting the photon budget needed to accurately estimate the transition rates in the model as

$$s = \frac{K\lambda_{ex}}{\lambda_{probe} M_\sigma}, \tag{76}$$

where K is the total number of photons in a single photon smFRET trace (photon budget),

λ_{ex} is the excitation rate, λ_{probe} represents the escape rate (timescale) that we want to probe, and M_σ is the number of system states. The parameters in the numerator control the amount of data available and the temporal resolution. On the other hand, the parameters in the denominator are the properties of the system under investigation and represent the required resolution.

From experimentation, we have found a photon budget index of approximately 10^6 to be a safe lower threshold for keeping errors below 15% (this error cutoff is a user choice) in parameter estimates. In the simple parametric example above, we have $K = 4.3 \times 10^5$, $\lambda_{ex} = 10$ ms^{-1}, and the fastest transition that we want to probe is $\lambda_{probe} = 2$ ms^{-1}, and $M_\sigma = 2$, which corresponds to a photon budget index of 1.08×10^7. In Fig. 6, we also demonstrate the reduction in errors (confidence interval size) for parameters of the same system as the photon budget is increased from 12500 photons to 400000 photons. For each of those cases, we used 9000 MCMC samples to compute statistical metrics such as quantiles.

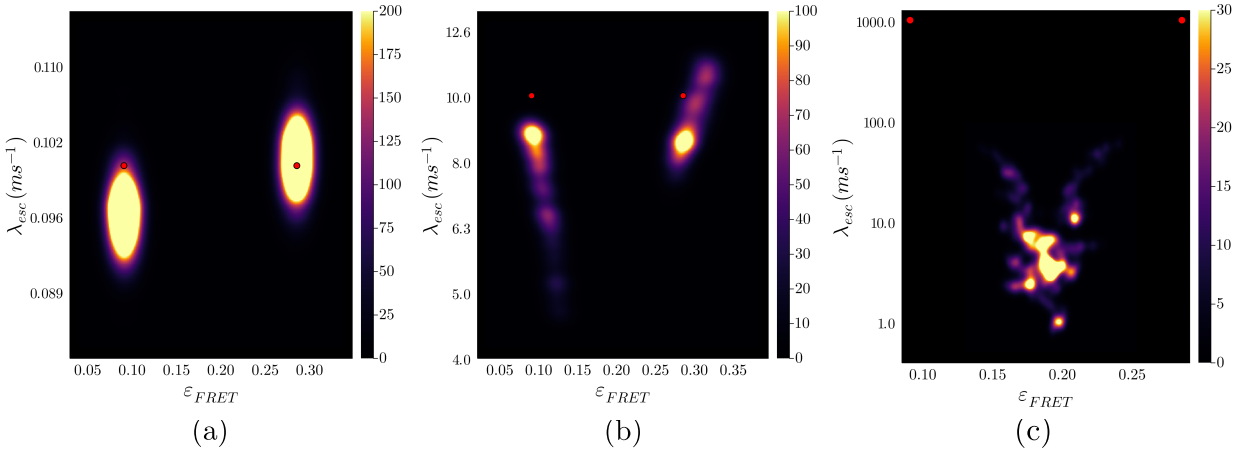

Figure 7: **Learned bivariate posterior for the system state escape rates λ_{esc} (log-scale) and FRET efficiencies ε_{FRET} from synthetic data**. For all synthetic smFRET traces, we use an excitation rate of 10 ms^{-1} and FRET efficiences of 0.29 and 0.09 for the two system states, respectively. The three panels correspond to different timescales being probed with transitions rates: (a) 0.1 ms^{-1}; (b) 10 ms^{-1}; and (c) 1000 ms^{-1}. The ground truth values are shown with red dots. The bias in the parameter estimates increases as faster kinetics are probed, demonstrating deterioration of the information content of the collected data resulting in expectedly poor estimation assuming a fixed photon budget of 670000. This can also be seen quantitatively by calculating the confidence intervals reported below for each case. The FRET efficiencies estimated by our sampler for the slowest case in panel (a) are $0.286^{+0.002}_{-0.002}$ and $0.091^{+0.001}_{-0.001}$ ms^{-1}, and the corresponding escape rates are $0.101^{+0.004}_{-0.005}$ and $0.096^{+0.004}_{-0.004}$ ms^{-1}. For the intermediate case in panel (b), FRET efficiencies estimated by our sampler are $0.200^{+0.117}_{-0.110}$ and $0.102^{+0.022}_{-0.014}$, and predicted escape rates are $8.47^{+2.42}_{-3.17}$ ms^{-1} and $7.67^{+1.32}_{-2.66}$ ms^{-1}. For the fastest case in panel (c), FRET efficiencies estimated by our sampler are $0.189^{+0.025}_{-0.027}$ and $0.189^{+0.016}_{-0.029}$, and predicted escape rates are $5.00^{+26.9}_{-3.63}$ ms^{-1} and $3.49^{+27.21}_{-2.49}$ ms^{-1}. Poorer confidence intervals for larger escape rates reflect larger uncertainty due to lack of information.

We further investigate the effect of another quantity which appears in the photon budget

index, that is, excitation rate on the parameter estimates. To do so, we generate three new synthetic datasets, each containing ≈ 670000 photons, using the same excitation rate of 10 ms^{-1}, and FRET efficiencies of 0.28 and 0.09 for the two system states, respectively, as before. However, the kinetics differ across these datasets so that they have system state transition rates well below, equal to, and well above the excitation rate. As such, for the first dataset, we probe slower kinetics compared to the excitation rate with system state transition rates set at 0.1 ms^{-1}. In the next two datasets, the molecule changes system states at a much faster rates of 10 ms^{-1} and 1000 ms^{-1}, respectively.

The results obtained for these FRET traces using our Bayesian methods are shown in Fig. 7. The bias in the posterior away from the ground truth increases as faster kinetics are probed in Fig. 7 from left to right. The results for the case with the fastest transition rates of 1000 ms^{-1} in Fig. 7(c) show a marked deterioration of the predictions, as the information content is not sufficient to separate the two FRET efficiencies resulting in estimated values close to the average of the ground truth values (≈ 0.185). This lack of information is also reflected in the uncertainties corresponding to each escape rate as shown in Fig. 7. Moreover, the predicted transition rates are of the same order as the excitation rate itself due to lack of temporal resolution available to probe such fast kinetics. To conclude, the excitation rate used to collect smFRET data and the total number of photons available determine the amount of information needed to resolve transitions among system states. As such, the ability of Bayesian methods to naturally propagate error from finiteness of information into parameter estimates make them indispensable tools for quantitative smFRET data analysis. This is by contrast to maximum likelihood-based methods which provide only inaccurate point estimates on account of limited data.

4.1.2 An example with crosstalk

Here, we demonstrate how our method handles cases when significant crosstalk is present. To show this, we use the same dynamical parameters and photon budget as in the previous subsection for generating synthetic data but allow 5% of the donor photons to be stochastically detected by the acceptor channel. We then analyze the data with two versions of our method, one that incorporates crosstalk and one that ignores it altogether. Our results show that neglecting crosstalk necessarily leads to artefactually higher FRET efficiency estimates. This is clearly seen in Fig. 8(a). As expected, incorporation of crosstalk into the likelihood as shown in Sec. 2.4.1 results in a smaller bias. In this case, both ground truths falls within the range of posteriors for the corrected model; see Fig. 8(b). Furthermore, as shown in Fig. 9 top panels, we note that donor crosstalk again results in overestimation in FRET efficiencies. However, when we correct for crosstalk, our BNP-FRET sampler starts learning FRET efficiencies with ground truths falling within the range of 95% confidence intervals; Fig. 9 bottom panels. As expected, our simulations in Fig. 9 also show that uncertainty increases with increasing crosstalk and parameter estimation fails for crosstalk values beyond 60%.

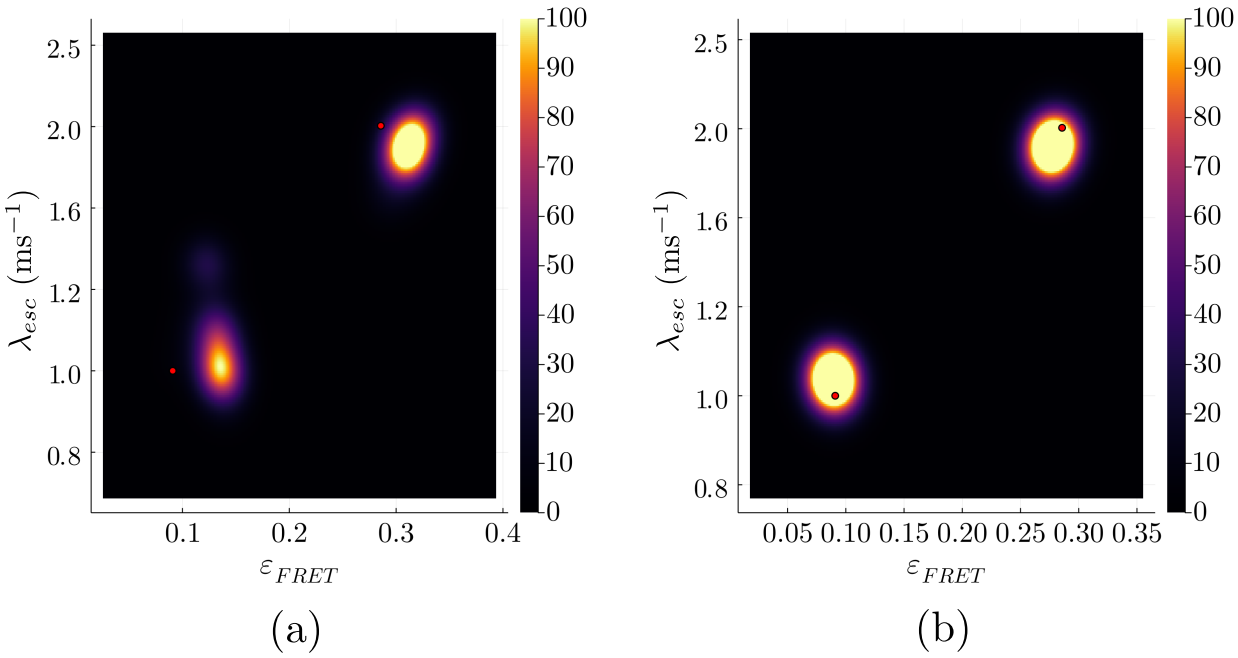

Figure 8: **Learned bivariate posterior for the system state escape rates λ_{esc} (log-scale) and FRET efficiencies ε_{FRET} for synthetic data with crosstalk**. The ground truth is shown with red dots. In panel (a), we show the learned posterior using the model that doesn't correct for cross talk consistently shows deviation from the ground truth with higher FRET efficiency estimates on account of more donor photons being detected as acceptor photons. The FRET efficiencies estimated by our sampler for this case are $0.313^{+0.008}_{-0.009}$ and $0.135^{+0.005}_{-0.006}$, and the predicted escape rates are $1.88^{+0.15}_{-0.14}$ ms^{-1} and $1.04^{+0.14}_{-0.08}$ ms^{-1}. For the corrected case shown in panel (b), FRET efficiencies estimated by our sampler are $0.276^{+0.006}_{-0.009}$ and $0.088^{+0.004}_{-0.005}$. Furthermore, predicted escape rates are $1.87^{+0.13}_{-0.14}$ ms^{-1} and $1.06^{+0.12}_{-0.10}$ ms^{-1}. The corrected model mitigates this bias as demonstrated by the posterior in panel (b).

4.1.3 An example with background emissions

In section 2.6, we had shown a way to include background emissions in the forward model. For the current example, we again choose the same kinetic parameters for the system and the FRET pair as in Fig. 5 but now some of the photons come from background sources with rates $\lambda_i^{bg} = \lambda_{ex}/10 = 1$ ms^{-1} for the i-th channel. Addition of a uniform background would again lead to higher FRET efficiency estimates due to excess photons detected in each channel, if left uncorrected in the model, as can be seen in Fig 10(a) and Fig. 11 top panels. By comparison to the uncorrected method, our results migrate toward the ground truth when analyzed with the full method; see Fig 10(b) and Fig. 11 bottom panels. Furthermore, as shown in Fig. 11, when background photons account for more than approximately 40% of detected photons, relative uncertainties in estimated transition rates become larger than 25% indicating unreliable results.

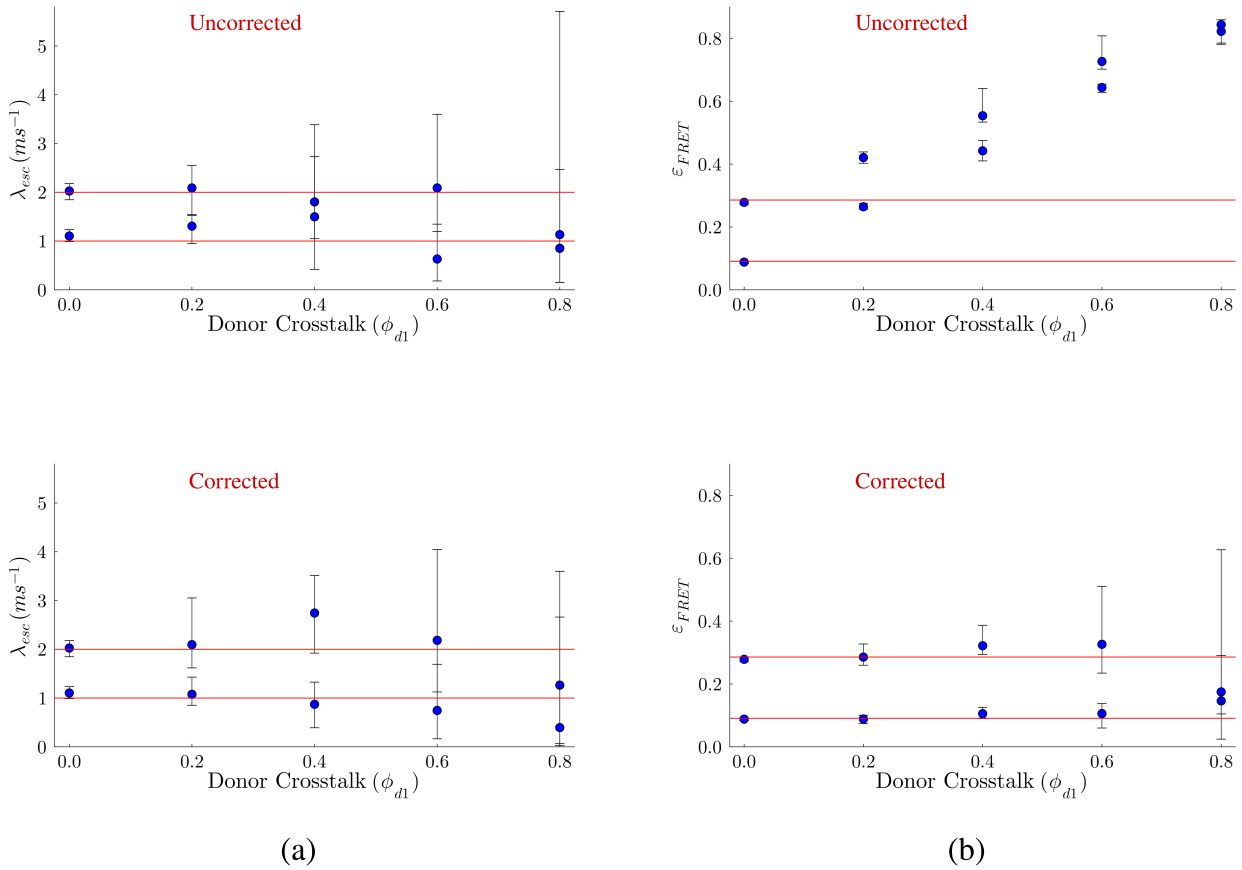

Figure 9: **System transition rates λ_{esc} and FRET efficiencies ε_{FRET} as functions of increasing donor crosstalk probability ϕ_{d1}.** To produce these plots, we generated synthetic data with excitation and escape rates as Fig. 5. In each plot, the blue dots represent the median values (50% quantile), and the ends of the attached confidence intervals represent the 5% and 95% quantiles. Furthermore, the ground truths are shown with red horizontal lines. In panel (a), our two plots show system transition rates estimated by the BNP-FRET sampler when corrected and uncorrected for crosstalk. We show similar plots for FRET efficiencies in panel (b). In all plots, we see that as donor crosstalk is increased, the confidence intervals grow (the posterior gets wider) and the estimates become unreliable after $\phi_{d1} > 0.6$. Additionally, as expected, if uncorrected for, the FRET efficiencies start to merge with increasing crosstalk due to most photons being detected in acceptor channel (labeled 1).

4.1.4 An example with IRF

To demonstrate the effect of the IRF as described in Sec. 2.4.3, we generated new synthetic data for a single fluorophore (with no FRET for simplicity alone) with an escape rate of $\lambda_d = 2.0$ ns^{-1} (similar to that of Cy3 dye [86]) being excited by a continuous-wave laser at a high excitation rate $\lambda_{ex} = 0.01$ ns^{-1}. For simplicity, we approximate the IRF with a truncated Gaussian distribution about 96 ps wide with mean at 48 ps. We again analyze the data with two versions of our method, both incorporating and neglecting the IRF. The results

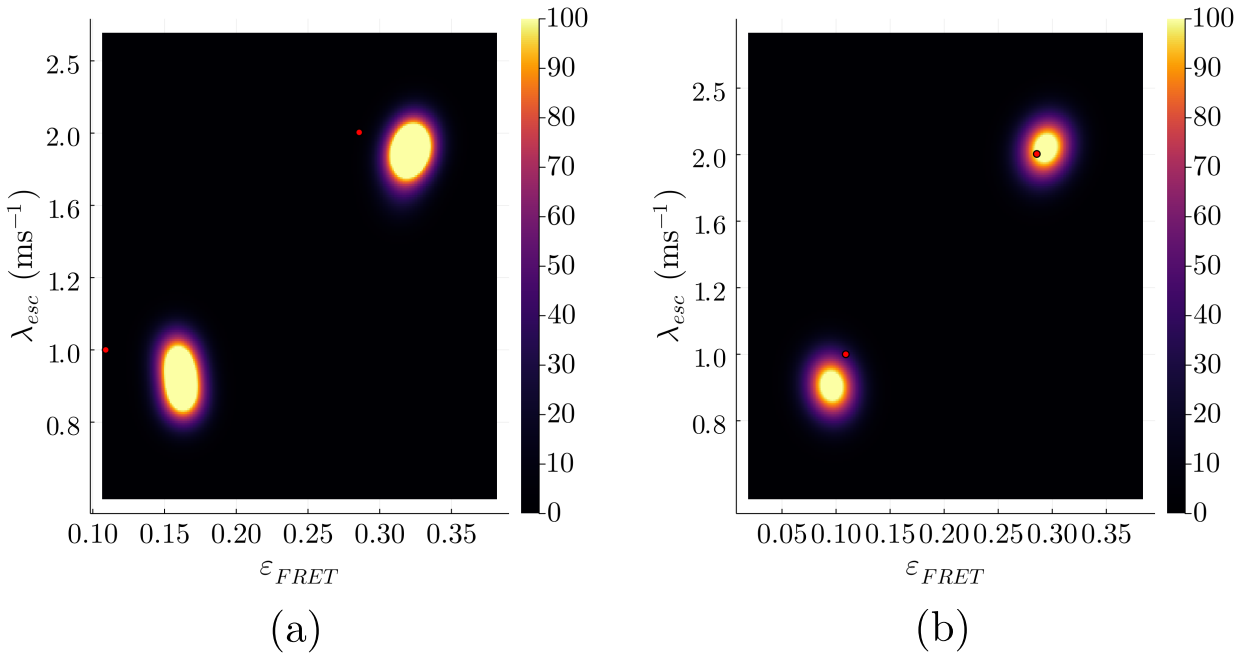

Figure 10: **Learned bivariate posterior for the system state escape rates λ_{esc} (log-scale) and FRET efficiencies ε_{FRET} given synthetic data with background emissions**. The ground truth is shown with red dots. The learned posterior distribution using the model that doesn't correct for background emissions (panel (a)) consistently shows bias away from the ground truth with higher estimates for FRET efficiencies on account of extra background photons. The FRET efficiencies estimated by our sampler for this uncorrected case are $0.319^{+0.009}_{-0.007}$ and $0.160^{+0.004}_{-0.004}$, and the predicted escape rates are $1.86^{+0.19}_{-0.15}$ ms^{-1} and $0.95^{+0.09}_{-0.10}$ ms^{-1}. The corrected model mitigates this bias as demonstrated by the posterior in panel (b) as demonstrated by learned FRET efficiencies of $0.276^{+0.006}_{-0.009}$ and $0.088^{+0.004}_{-0.005}$. Furthermore, predicted escape rates are $1.87^{+0.13}_{-0.14}$ ms^{-1} and $1.06^{+0.12}_{-0.10}$ ms^{-1}.

are depicted in Fig. 12, where the posterior is narrower when incorporating the IRF. This is especially helpful when accurate lifetime determination is important in discriminating between different system states. By contrast, accurate determination of lifetimes (which span ns timescales) do not impact the determination of much slower system kinetics from one system state to the next.

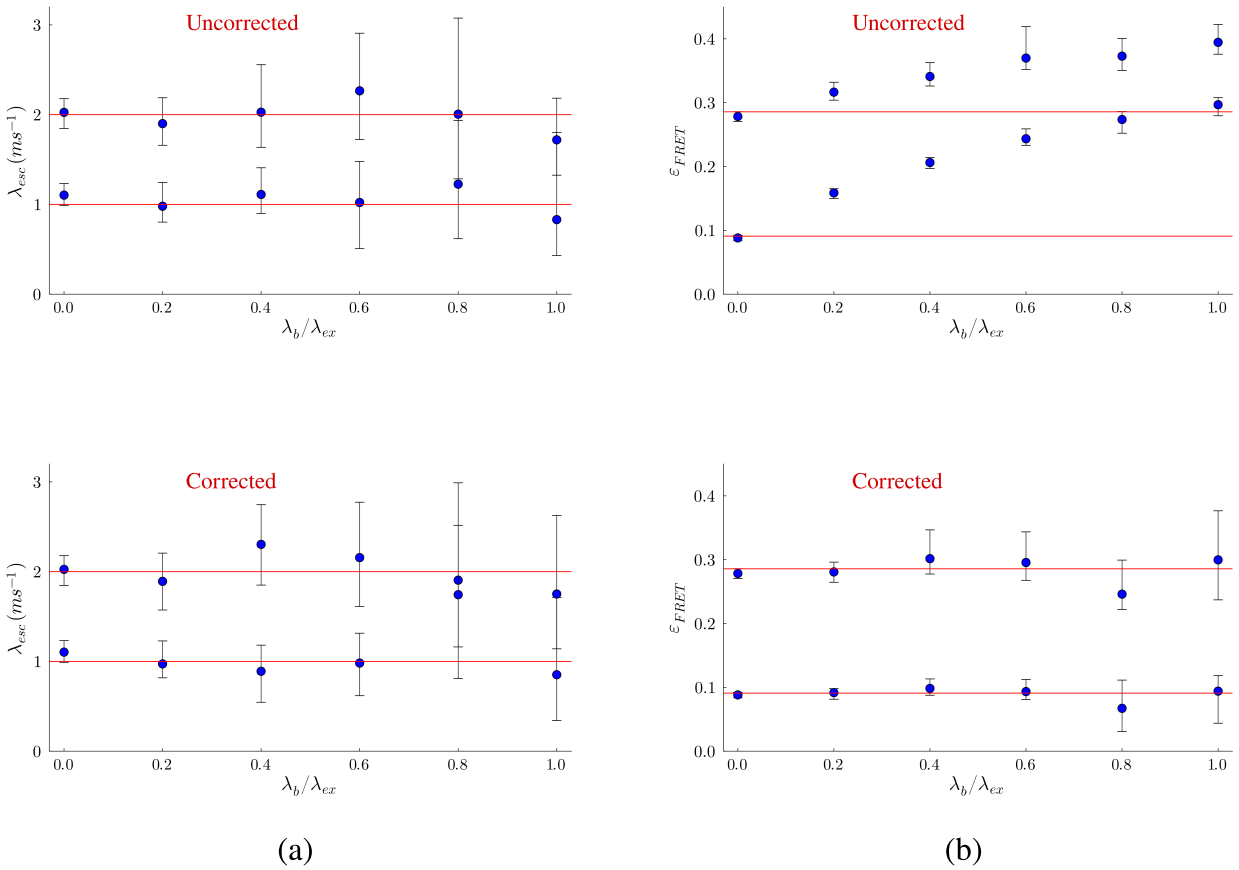

Figure 11: **System transition rates λ_{esc} and FRET efficiencies ε_{FRET} as functions of increasing donor and acceptor background fraction $\lambda_{bg}/\lambda_{ex}$.** To produce these plots, we generated synthetic data with an excitation rate of $\lambda_{ex} = 10$ ms^{-1}, and escape rates $\lambda_{esc} = 1$ and 2 ms^{-1} for the two system states, respectively, same as Fig. 5, while increasing the fraction of background photons (donor and acceptor) from 0 to 50% ($\lambda_{bg}/\lambda_{ex} = 1$). In each plot, the blue dots represent the median values (50% quantile), and the ends of the attached confidence intervals represent the 5% and 95% quantiles. Furthermore, ground truths are shown with red horizontal lines. In panel (a), we show two plots showing system transition rates estimated by the BNP-FRET sampler when corrected and uncorrected for crosstalk. We show similar plots for FRET efficiencies in panel (b). In all plots, we see that as background is increased, the confidence intervals get bigger (the posterior gets wider) and the estimates become unreliable after $\lambda_{bg}/\lambda_{ex} > 0.6$. Additionally, as expected, if unaccounted for, FRET efficiencies start to merge with increasing background as photons originating from FRET events significantly reduce.

4.2 A Nonparametric Example

Here, we demonstrate our method in learning the number of system states by analyzing approximately 600 ms (≈120000 photons) of synthetic smFRET time trace data with three system states under pulsed illumination with 25 ns interpulse window; see Fig. 13. This example utilizes the iHMM method described in Sec. 3.2.2 and discussed in greater depth

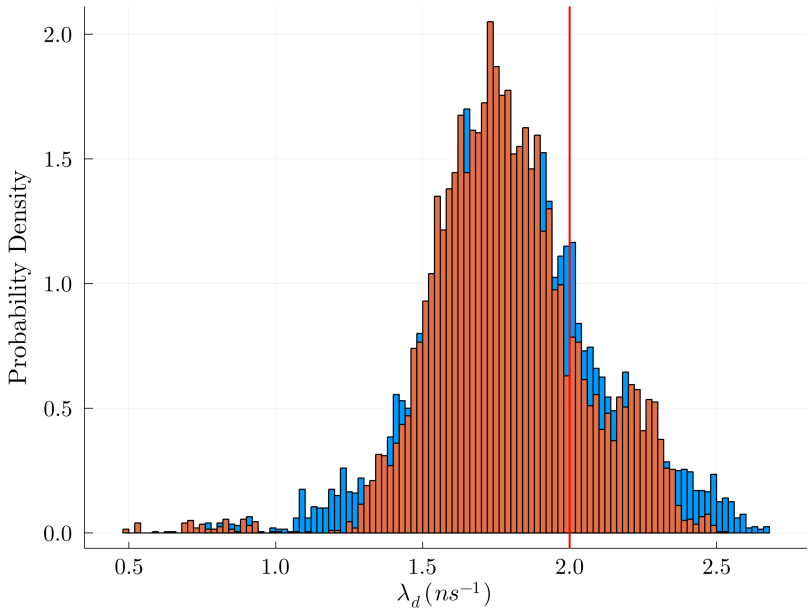

Figure 12: **Effects of IRF**. Both histograms show the fluorophore's inverse lifetime with ground truth shown by a red line. The bias in the peaks away from the ground truth arises from the limited amounts of data being used to learn the posterior shown. Corrected model (orange) reduces the histogram's breadth as compared with the uncorrected model (blue). We conclude from the small effects of correcting for the IRF that, predictably, the IRF may be less important under continuous illumination. By contrast, under pulsed illumination to be explored in the third companion manuscript [64], the IRF will play a more significant role.

in the third companion manuscript [64]. Using realistic values from the third companion manuscript [64], we set the excitation probability per pulse to be 0.005. Furthermore, kinetics are set at 1.2 ms escape rates for the highest and lowest FRET system states, and an escape rate of 2.4 ms for an intermediate system state. Our BNP method simultaneously recovered the correct system state transition probabilities and thereby the number of system states along with other parameters, including donor and acceptor relaxation rates and the per-pulse excitation probability. By comparison, a parametric version of the same method must assume a fixed number of system states *a priori*. Assuming, say, two system states results in both higher-FRET system states being combined together into one system state with a FRET efficiency of 0.63 and a lifetime of about 0.6 ms; see Fig. 13(c).

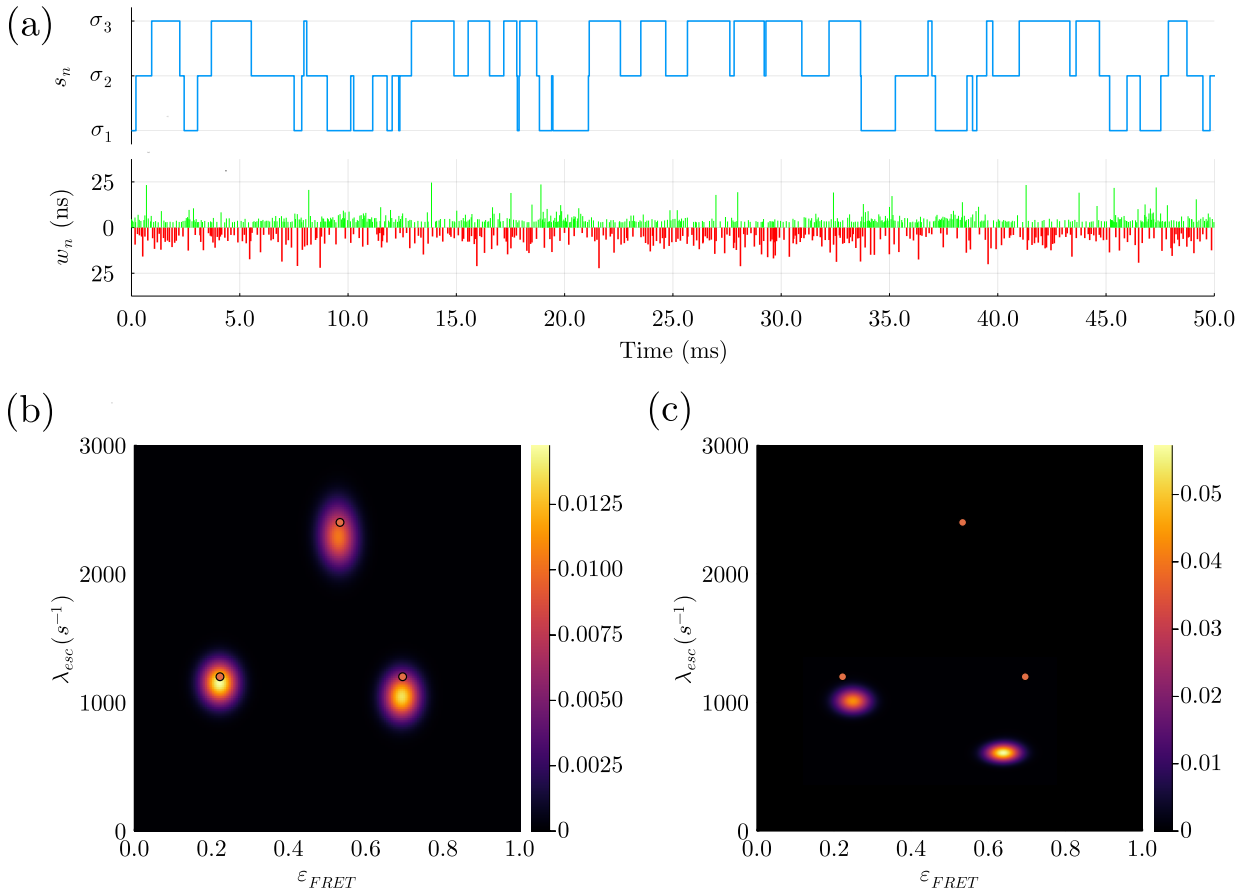

Figure 13: **Demonstration of nonparametric analysis on synthetic pulsed data**. In (a) we show simulated data for a pulsed illumination experiment. It illustrates a trajectory with three system states labeled s_n in blue and the corresponding photon arrivals with the vertical length of the red or green line denoting the lifetime observed in nanoseconds. Here (b) and (c) show the bivariate posterior for the escape rates λ_{esc} and FRET efficiencies ε_{FRET}. The ground truth is again shown with red dots. The learned posterior using the two system state parametric model (panel (c)) combines high FRET states into one averaged system state with a long lifetime. The nonparametric model correctly infers three system states, as shown in (b).

5 Discussion

In this paper, we have presented a complete framework to analyze single photon smFRET data which includes a photon-by-photon likelihood, detector effects, fluorophore characteristics, and different illumination methods. We demonstrated how modern Bayesian methods can be used to obtain full distributions over the parameters, and discussed limitations posed by the photon budget and excitation rate. Additionally, we have shown how to implement a nonparametric inverse strategy to learn an unknown number of system states.

Our method readily accommodates details relevant to specialized smFRET applications. For instance, we can analyze spatial and temporal dependence of excitation by simple modification of generator matrices included in Eq. 21. This is useful in experiments employing pulsed illumination, (Sec. 2.5), as well as alternating-laser excitation (ALEX). In particular, ALEX is used to directly excite the acceptor label, either as a way to gain qualitative information about the sample [63, 87], to reduce photobleaching [63, 87], or to study intermolecular interactions [16]. Similarly, the generator matrix in Eq. 21 can easily be expanded to include any number of labels, extending our method beyond two colors. Three color smFRET experiments have revealed simultaneous interactions between three proteins [88], monitored conformational subpopulations of molecules [89], and improved our understanding of protein folding and interactions [2, 16, 90–92].

As the likelihood (Eq. 21) involves as many matrix exponentials as detected photons, the computational cost of our method scales approximately linearly with the number of photons and quadratically with the number of system states. For instance, it took 5 hours to process the data used to generate Fig. 5 on a regular desktop computer. Additions to our model which increase computational cost include: 1) IRF; 2) pulsed illumination; and 3) BNPs. The computational cost associated to the IRF is attributed to the integral required (see Sec. 2.4.3). The cost of the likelihood computation in the pulsed illumination case depends linearly with pulse number, rather than with photon number (see Eq. 51). This greatly increases the computational cost in cases where photon detections are infrequent. Finally, BNPs necessarily expand the dimensions of the generator matrix whose exponentiation is required (Eq. 21) resulting in longer burn-in time for our MCMC chains.

As a result, we have optimized the computational cost with respect to the physical conditions of the system being studied. First, inclusion of the IRF can be parallelized, potentially reducing the time-cost to a calculation over a single data acquisition period. In our third companion manuscript [64], dealing with pulsed illumination, we improve computational cost by making the assumption that fluorophore relaxation occurs within the window between consecutive pulses, thereby reducing our second order structure herein to a first order HMM, and allowing for faster computation of the likelihood in pulses where no photon is detected. In [64], we mitigate the computational cost of (3) by assuming physically motivated timescale separation.

As it stands, our framework applies to smFRET experiments on immobilized molecules. However, it is often the case that molecules labeled with FRET pairs are allowed to diffuse freely through a confocal volume, such as in the study of binding and unbinding events [17], protein-protein interactions [17, 63], and unhindered conformational dynamics of freely diffusing proteins [63, 93]. Photon-by-photon analysis of such data is often based on correlation methods which suffer from bulk averaging [40, 42, 43]. We believe our framework has the potential to extend Refs. [50, 73, 94] to learn both the kinetics and diffusion coefficients of single molecules.

Additionally, our current framework is restricted to models with discrete system states. However, smFRET can also be used to study systems that are better modeled as continuous, such as intrinsically disordered proteins which include continuous changes not always well approximated by discrete system states [17, 95]. Adopting an adaptation of Ref. [54] should allow us to generalize this framework and instead infer energy landscapes, perhaps relevant to protein folding [96, 97], continuum ratchets as applied to motor protein kinetics [98], and

the stress modified potential landscapes of mechano-sensitive molecules [99].

To conclude, we have presented a general framework and demonstrated the importance of incorporating various features into the likelihood while learning full distributions over all unknowns including system states. In the following two companion manuscripts [64, 78], we specialize our method, and computational scheme, to continuous and pulsed illumination. We then apply our method to interactions of the intrinsically disordered proteins NCBD and ACTR [78] under continuous illumination, and the kinetics of the Holliday junction under pulsed illumination [64].

6 Code availability

The BNP-FRET software package is available on Github at
`https://github.com/LabPresse/BNP-FRET`

7 Acknowledgments

We thank Weiqing Xu and Dr. Zeliha Kilic for regular feedback and help, especially during the development of the nonparametrics samplers, and Dr. Irina Gopich for discussions. S. P. acknowledges support from the NIH NIGMS (R01GM130745) for supporting early efforts in nonparametrics and NIH NIGMS (R01GM134426) for supporting single photon efforts. Bulk of the computations were performed on Agave and Sol supercomputers at ASU.

Bibliography

[1] Luling Wu, Chusen Huang, Ben P. Emery, Adam C. Sedgwick, Steven D. Bull, Xiao-Peng He, He Tian, Juyoung Yoon, Jonathan L. Sessler, and Tony D. James. Förster resonance energy transfer (FRET)-based small-molecule sensors and imaging agents. *Chemical Society Reviews*, 49:5110, 2020.

[2] Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule FRET. *Nature Methods*, 5(6):507, 2008.

[3] Alexander P Demchenko. *Introduction to fluorescence sensing*. Springer Science & Business Media, 2008.

[4] Ammasi Periasamy and Richard Day. *Molecular imaging: FRET microscopy and spectroscopy*. Elsevier, 2011.

[5] Elizabeth Rhoades, Eugene Gussakovsky, and Gilad Haran. Watching proteins fold one molecule at a time. *Proceedings of the National Academy of Sciences*, 100(6):3197, 2003.

[6] Boris Martinac. Single-molecule FRET studies of ion channels. *Progress in Biophysics and Molecular Biology*, 130:192, 2017.

[7] Hoi Sung Chung, Kevin McHale, John M. Louis, and William A. Eaton. Single-molecule fluorescence experiments determine protein folding transition path times. *Science*, 335(6071):981, 2012.

[8] S Pressé, K Ghosh, R Phillips, and KA Dill. Dynamical fluctuations in biochemical reactions and cycles. *Physical Review E*, 82(3):031905, 2010.

[9] Benjamin Schuler. Single-molecule FRET of protein structure and dynamics - a primer. *Journal of Nanobiotechnology*, 11(1):S2, 2013.

[10] Oana Coban, Laura C Zanetti-Dominguez, Daniel R Matthews, Daniel J Rolfe, Gregory Weitsman, Paul R Barber, Jody Barbeau, Viviane Devauges, Florian Kampmeier, Martyn Winn, et al. Effect of phosphorylation on EGFR dimer stability probed by single-molecule dynamics and FRET/FLIM. *Biophysical journal*, 108(5):1013, 2015.

[11] Kangkan Halder, Nicole Dölker, Qui Van, Ingo Gregor, Achim Dickmanns, Imke Baade, Ralph H Kehlenbach, Ralf Ficner, Jörg Enderlein, Helmut Grubmüller, et al. MD simulations and FRET reveal an environment-sensitive conformational plasticity of importin-β. *Biophysical journal*, 109(2):277, 2015.

[12] Tara Sabir, Gunnar F. Schröder, Anita Toulmin, Peter McGlynn, and Steven W. Magennis. Global structure of forked DNA in solution revealed by high-resolution single-molecule FRET. *Journal of the American Chemical Society*, 133(5):11881191, 2011.

[13] Carey Phelps, Brett Israels, Davis Jose, Morgan C. Marsh, Peter H. von Hippel, and Andrew H. Marcus. Using microsecond single-molecule FRET to determine the assembly pathways of T4 ssDNA binding protein onto model DNA replication forks. *Proceedings of the National Academy of Sciences*, 114(18):E3612, 2017.

[14] Laura E. Baltierra-Jasso, Michael J. Morten, Linda Laflör, Steven D. Quinn, and Steven W. Magennis. Crowding-induced hybridization of single DNA hairpins. *Journal of the American Chemical Society*, 137(51):16020, 2015.

[15] Yong Wang, Yanxin Liu, Hannah A DeBerg, Takeshi Nomura, Melinda Tonks Hoffman, Paul R Rohde, Klaus Schulten, Boris Martinac, and Paul R Selvin. Single molecule FRET reveals pore size and opening mechanism of a mechano-sensitive ion channel. *eLife*, 3:e01834, 2014.

[16] Janghyun Yoo, Jae-Yeol Kim, John M. Louis, Irina V. Gopich, and Hoi Sung Chung. Fast three-color single-molecule FRET using statistical inference. *Nature Communications*, 11(1), 2020.

[17] Benjamin Schuler and William A Eaton. Protein folding studied by single-molecule FRET. *Current Opinion in Structural Biology*, 18(1):16, 2008.

[18] Sinan Kilic, Suren Felekyan, Olga Doroshenko, Iuliia Boichenko, Mykola Dimura, Hayk Vardanyan, Louise C. Bryan, Gaurav Arya, Claus A. M. Seidel, and Beat Fierz. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. *Nature Communications*, 9(1):235, 2018.

[19] Eitan Lerner, Thorben Cordes, Antonino Ingargiola, Yazan Alhadid, SangYoon Chung, Xavier Michalet, and Shimon Weiss. Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science*, 359(6373):eaan1133, 2018.

[20] Th. Förster. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, 437(1-2):55, 1948.

[21] Garth A. Jones and David S. Bradshaw. Resonance energy transfer: From fundamental theory to recent applications. *Frontiers in Physics*, 7:100, 2019.

[22] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov. Invited review article: Single-photon sources and detectors. *Review of Scientific Instruments*, 82(7):071101, 2011.

[23] Sean A. McKinney, Chirlmin Joo, and Taekjip Ha. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal*, 91(5):1941, 2006.

[24] Jonathan E. Bronson, Jingyi Fei, Jake M. Hofman, Ruben L. Gonzalez, Jr., and Chris H. Wiggins. Learning rates and states from biophysical time series: A Bayesian approach to model selection and single-molecule FRET data. *Biophysical Journal*, 97(12):3196, 2009.

[25] Ioannis Sgouralis, Shreya Madaan, Franky Djutanta, Rachael Kha, Rizal F. Hariadi, and Steve Pressé. A Bayesian nonparametric approach to single molecule Förster resonance energy transfer. *The Journal of Physical Chemistry B*, 123(3):675, 2019.

[26] Wolfgang Becker. *Advanced time-correlated single photon counting applications*. Springer, 2015.

[27] Sebastian Isbaner, Narain Karedla, Daja Ruhlandt, Simon Christoph Stein, Anna Chizhik, Ingo Gregor, and Jörg Enderlein. Dead-time correction of fluorescence lifetime measurements and fluorescence lifetime imaging. *Optics Express*, 24(9):9429, 2016.

[28] Ivan Rasnik, Sean A McKinney, and Taekjip Ha. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Methods*, 3(11):891, 2006.

[29] Christian G. Hübner, Alois Renn, Indrek Renge, and Urs P. Wild. Direct observation of the triplet lifetime quenching of single dye molecules by molecular oxygen. *The Journal of Chemical Physics*, 115(21):9619, 2001.

[30] R. E. Dale, J. Eisinger, and W. E. Blumberg. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophysical Journal*, 26(2):161, 1979.

[31] Benjamin Schuler. Single-molecule fluorescence spectroscopy of protein folding. *ChemPhysChem*, 6(7):1206, 2005.

[32] Zeliha Kilic, Ioannis Sgouralis, Wooseok Heo, Kunihiko Ishii, Tahei Tahara, and Steve Pressé. Extraction of rapid kinetics from smFRET measurements using integrative detectors. *Cell Reports Physical Science*, 2(5):100409, 2021.

[33] Irina V. Gopich and Attila Szabo. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *The Journal of Physical Chemistry B*, 107(21):5058, 2003.

[34] Michael Andrec, Ronald M. Levy, and David S. Talaga. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden Markov models. *The Journal of Physical Chemistry A*, 107(38):7454, 2003.

[35] Bettina G. Keller, Andrei Kobitski, Andres Jäschke, G. Ulrich Nienhaus, and Frank Noé. Complex RNA folding kinetics revealed by single-molecule FRET and hidden Markov models. *Journal of the American Chemical Society*, 136(12):4534, 2014.

[36] Johannes Thomsen, Magnus Berg Sletfjerding, Simon Bo Jensen, Stefano Stella, Bijoya Paul, Mette Galsgaard Malle, Guillermo Montoya, Troels Christian Petersen, and Nikos S Hatzakis. DeepFRET, a software for rapid and automated single-molecule FRET data classification using deep learning. *eLife*, 9:e60404, 2020.

[37] Irina V. Gopich and Attila Szabo. Decoding the pattern of photon colors in single-molecule FRET. *The Journal of Physical Chemistry B*, 113(31):10965, 2009.

[38] Zeliha Kilic, Ioannis Sgouralis, and Steve Pressé. Generalizing HMMs to continuous time for fast kinetics: Hidden Markov jump processes. *Biophysical Journal*, 120(3):409, 2021.

[39] Paul David Harris, Alessandra Narducci, Christian Gebhardt, Thorben Cordes, Shimon Weiss, and Eitan Lerner. Multi-parameter photon-by-photon hidden Markov modeling. *Nature communications*, 13(1):1, 2022.

[40] Irina V. Gopich and Attila Szabo. Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis. *The Journal of Chemical Physics*, 124(15):154712, 2006.

[41] Ioannis Sgouralis and Steve Pressé. An introduction to infinite HMMs for single-molecule data analysis. *Biophysical Journal*, 112(10):2021, 2017.

[42] Irina V. Gopich and Attila Szabo. Single-molecule FRET with diffusion and conformational dynamics. *The Journal of Physical Chemistry B*, 111(44):12925, 2007.

[43] Menahem Pirchi, Roman Tsukanov, Rashid Khamis, Toma E. Tomov, Yaron Berger, Dinesh C. Khara, Hadas Volkov, Gilad Haran, and Eyal Nir. Photon-by-photon hidden Markov model analysis for microsecond single-molecule FRET kinetics. *The Journal of Physical Chemistry B*, 120(51):13065, 2016.

[44] Eitan Lerner, Antonino Ingargiola, and Shimon Weiss. Characterizing highly dynamic conformational states: The transcription bubble in RNAP-promoter open complex as an example. *The Journal of Chemical Physics*, 148(12):123315, 2018.

[45] Jake Lever, Martin Krzywinski, and Naomi Altman. Model selection and overfitting. *Nature Methods*, 13(9):703, 2016.

[46] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209, 1973.

[47] Samuel J. Gershman and David M. Blei. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1, 2012.

[48] Ioannis Sgouralis, Miles Whitmore, Lisa Lapidus, Matthew J. Comstock, and Steve Pressé. Single molecule force spectroscopy at high data acquisition: A Bayesian nonparametric analysis. *The Journal of Chemical Physics*, 148(12):123320, 2018.

[49] Meysam Tavakoli, J Nicholas Taylor, Chun-Biu Li, Tamiki Komatsuzaki, and Steve Pressé. Single molecule data analysis: An introduction. *Advances in Chemical Physics*, 162:205–305, 2017.

[50] Meysam Tavakoli, Sina Jazani, Ioannis Sgouralis, Omer M. Shafraz, Sanjeevi Sivasankar, Bryan Donaphon, Marcia Levitus, and Steve Pressé. Pitching single-focus confocal data analysis one photon at a time with Bayesian nonparametrics. *Physical Review X*, 10(1):011021, 2020.

[51] Meysam Tavakoli, Sina Jazani, Ioannis Sgouralis, Wooseok Heo, Kunihiko Ishii, Tahei Tahara, and Steve Pressé. Direct photon-by-photon analysis of time-resolved pulsed excitation data using Bayesian nonparametrics. *Cell Reports Physical Science*, 1(11):100234, 2020.

[52] J. Shepard Bryan IV, Ioannis Sgouralis, and Steve Pressé. Diffraction-limited molecular cluster quantification with Bayesian nonparametrics. *Nature Computational Science*, 2(2):102, 2022.

[53] Mohamadreza Fazel, Sina Jazani, Lorenzo Scipioni, Alexander Vallmitjana, Enrico Gratton, Michelle A. Digman, and Steve Pressé. High resolution fluorescence lifetime maps from minimal photon counts. *ACS Photonics*, 9(3):1015, 2022.

[54] J Shepard Bryan IV, Prithviraj Basak, John Bechhoefer, and Steve Pressé. Inferring potential landscapes from noisy trajectories of particles within an optical feedback trap. *Iscience*, 25(9):104731, 2022.

[55] J. Shepard Bryan, Ioannis Sgouralis, and Steve Pressé. Inferring effective forces for Langevin dynamics using Gaussian processes. *The Journal of Chemical Physics*, 152(12):124106, 2020.

[56] Lekha Patel, Nils Gustafsson, Yu Lin, Raimund Ober, Ricardo Henriques, and Edward Cohen. A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores. *The Annals of Applied Atatistics*, 13(3):1397, 2019.

[57] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257, 1989.

[58] Alexa L Mattheyses, Adam D Hoppe, and Daniel Axelrod. Polarized fluorescence resonance energy transfer microscopy. *Biophysical Journal*, 87(4):2787, 2004.

[59] Fiona Gordon, Saioa Elcoroaristizabal, and Alan G. Ryder. Modelling Förster resonance energy transfer (FRET) using anisotropy resolved multi-dimensional emission spectroscopy (ARMES). *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1865(2):129770, 2021.

[60] Gerald W. Gordon, Gail Berry, Xiao Huan Liang, Beth Levine, and Brian Herman. Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. *Biophysical Journal*, 74(5):2702, 1998.

[61] Stephan Benke, Andrea Holla, Bengt Wunderlich, Andrea Soranno, Daniel Nettels, and Benjamin Schuler. Combining rapid microfluidic mixing and three-color single-molecule FRET for probing the kinetics of protein conformational changes. *The Journal of Physical Chemistry B*, 125(24):6617, 2021.

[62] Franziska Zosel, Andrea Soranno, Karin J. Buholzer, Daniel Nettels, and Benjamin Schuler. Depletion interactions modulate the binding between disordered proteins in crowded environments. *Proceedings of the National Academy of Sciences*, 117(24):13480, 2020.

[63] Achillefs N. Kapanidis, Ted A. Laurence, Nam Ki Lee, Emmanuel Margeat, Xiangxu Kong, and Shimon Weiss. Alternating-laser excitation of single molecules. *Accounts of Chemical Research*, 38(7):523, 2005.

[64] Matthew Safar, Ayush Saurabh, Bidyut Sarkar, Mohamadreza Fazel, Kunihiko Ishii, Tahei Tahara, Ioannis Sgouralis, and Steve Pressé. Single photon smFRET. III. application to pulsed illumination. *bioRxiv*, 2022.

[65] Alex Dytso and H. Vincent Poor. Estimation in Poisson noise: Properties of the conditional mean estimator. *IEEE Transactions on Information Theory*, 66(7):4304, 2020.

[66] R Alléaume, F Treussart, J-M Courty, and J-F Roch. Photon statistics characterization of a single-photon source. *New Journal of Physics*, 6:85, 2004.

[67] J. Brewer. Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, 25(9):772, 1978.

[68] Geoffrey C. Rollins, Jae Yen Shin, Carlos Bustamante, and Steve Pressé. Stochastic approach to the molecular counting problem in superresolution microscopy. *Proceedings of the National Academy of Sciences*, 112(2):E110, 2015.

[69] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403, 1976.

[70] 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, 1953.

[71] Andrew Gelman, Walter R Gilks, and Gareth O Roberts. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110, 1997.

[72] Ioannis Sgouralis and Steve Pressé. An introduction to infinite HMMs for single-molecule data analysis. *Biophysical Journal*, 112(10):2021, 2017.

[73] 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):3662, 2019.

[74] Zoubin Ghahramani and Thomas Griffiths. Infinite latent feature models and the indian buffet process. *Advances in neural information processing systems*, 18, 2005.

[75] Romain Thibaux and Michael I Jordan. Hierarchical beta processes and the indian buffet process. In *Artificial intelligence and statistics*, page 564. PMLR, 2007.

[76] John Paisley and Lawrence Carin. Nonparametric factor analysis with beta process priors. In *Proceedings of the 26th annual international conference on machine learning*, page 777, 2009.

[77] Luai Al Labadi and Mahmoud Zarepour. On approximations of the beta process in latent feature models: Point processes approach. *Sankhya A*, 80(1):59, 2018.

[78] Ayush Saurabh, Matthew Safar, Mohamadreza Fazel, Ioannis Sgouralis, and Steve Pressé. Single photon smFRET. II. application to continuous illumination. *bioRxiv*, 2022.

[79] Mohamadreza Fazel, Alexander Vallmitjana, Lorenzo Scipioni, Enrico Gratton, Michelle A Digman, and Steve Presse. Fluorescence lifetime: Beating the IRF and interpulse window. *bioRxiv*, 2022.

[80] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566, 2006.

[81] Jim Pitman. Poisson–Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. *Combinatorics, Probability and Computing*, 11(5):501, 2002.

[82] Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639, 1994.

[83] Radford M Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249, 2000.

[84] Alan E Gelfand, Athanasios Kottas, and Steven N MacEachern. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, 100(471):1021, 2005.

[85] Franziska Zosel, Davide Mercadante, Daniel Nettels, and Benjamin Schuler. A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nature Communications*, 9(1):3332, 2018.

[86] Matthew E. Sanborn, Brian K. Connolly, Kaushik Gurunathan, and Marcia Levitus. Fluorescence properties and photophysics of the sulfoindocyanine cy3 linked covalently to DNA. *The Journal of Physical Chemistry B*, 111(37):11064, 2007.

[87] Johannes Hohlbein, Timothy D. Craggs, and Thorben Cordes. Alternating-laser excitation: Single-molecule FRET and beyond. *Chemical Society Reviews*, 43:1156, 2014.

[88] Yuansheng Sun, Horst Wallrabe, Cynthia F. Booker, Richard N. Day, and Ammasi Periasamy. Three-color spectral FRET microscopy localizes three interacting proteins in living cells. *Biophysical Journal*, 99(4):1274, 2010.

[89] Jean-Pierre Clamme and Ashok A. Deniz. Three-color single-molecule fluorescence resonance energy transfer. *ChemPhysChem*, 6(1):74, 2005.

[90] Sungchul Hohng, Chirlmin Joo, and Taekjip Ha. Single-molecule three-color FRET. *Biophysical Journal*, 87(2):1328, 2004.

[91] Steve Pressé, Jack Peterson, Julian Lee, Phillip Elms, Justin L MacCallum, Susan Marqusee, Carlos Bustamante, and Ken Dill. Single molecule conformational memory extraction: P5ab RNA hairpin. *The Journal of Physical Chemistry B*, 118(24):6597, 2014.

[92] Steve Pressé, Julian Lee, and Ken A Dill. Extracting conformational memory from single-molecule kinetic data. *The Journal of Physical Chemistry B*, 117(2):495, 2013.

[93] Ashok A. Deniz, Maxime Dahan, Jocelyn R. Grunwell, Taekjip Ha, Ann E. Faulhaber, Daniel S. Chemla, Shimon Weiss, and Peter G. Schultz. Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. *Proceedings of the National Academy of Science*, 96(7):3670, 1999.

[94] Sina Jazani, Ioannis Sgouralis, and Steve Pressé. A method for single molecule tracking using a conventional single-focus confocal setup. *The Journal of Chemical Physics*, 150(11):114108, 2019.

[95] Benjamin Schuler and Hagen Hofmann. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Current Opinion in Structural Biology*, 23(1):36, 2013.

[96] Serdal Kirmizialtin, Lei Huang, and Dmitrii E. Makarov. Topography of the free-energy landscape probed via mechanical unfolding of proteins. *The Journal of Chemical Physics*, 122(23):234915, 2005.

[97] Benjamin A. Shoemaker, Jin Wang, and Peter G. Wolynes. Structural correlations in protein folding funnels. *Proceedings of the National Academy of Sciences*, 94(3):777, 1997.

[98] Anatoly B. Kolomeisky and Michael E. Fisher. Molecular motors: A theorist's perspective. *Annual Review of Physical Chemistry*, 58(1):675, 2007.

[99] Sai Sriharsha M. Konda, Stanislav M. Avdoshenko, and Dmitrii E. Makarov. Exploring the topography of the stress-modified energy landscapes of mechanosensitive molecules. *The Journal of Chemical Physics*, 140(10):104114, 2014.