Single Photon smFRET. I. Theory and Conceptual Basis

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

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

July 21, 2022

Contents

1 Introduction **2**

2 Forward Model **6**
2.1 smFRET Data . . . 6
2.2 Likelihood . . . 7
2.2.1 Absence of Observations . . . 9
2.2.2 Introducing Observations . . . 9
2.2.3 Recursion Formulas . . . 10
2.2.4 Reduced Propagators . . . 12
2.2.5 Likelihood for the Second Order HMM . . . 13
2.2.6 Continuum Limit . . . 13
2.2.7 Final Likelihood . . . 14
2.3 Detector Effects . . . 14
2.3.1 Crosstalk and detection efficiency . . . 15
2.3.2 Effects of detector dead time . . . 16
2.3.3 Adding the IRF . . . 17
2.4 Illumination Features . . . 18
2.5 Background Emissions . . . 20

2.6 Fluorophore Characteristics: Quantum Yield, Blinking, Photobleaching, and Acceptor Direct Excitation . . . 21
2.7 Synthetic Data Generation . . . 22
2.7.1 Dynamics . . . 23

3 Inverse Strategy **24**
3.1 Parametric Sampler: BNP-FRET with Fixed Number of System States . . . 26
3.2 Nonparametrics: Predicting the Number of System States . . . 28
3.2.1 Beta-Bernoulli Process for Continuous Illumination . . . 29
3.2.2 iHMM Methods for Pulsed Illumination . . . 30
3.3 Likelihood Computation in Practice . . . 32

4 Results **33**
4.1 Parametric Examples . . . 34
4.1.1 Photon Budget and Excitation Rate . . . 34
4.1.2 An Example with Crosstalk . . . 36
4.1.3 An Example with Background Emissions . . . 37
4.1.4 An Example with IRF . . . 37
4.2 A Nonparametric Example . . . 38

5 Discussion **39**

6 Acknowledgments **42**

Bibliography **43**

Abstract

We present a unified conceptual framework and 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 an smFRET experiment, including: 1) fluorophore photophysics; 2) continuous time dynamics 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, we propose a second order hidden Markov model (HMM) and Bayesian nonparametrics (BNP) on account of items 1, 2 and 5 on the list, respectively. In companion manuscripts II and III, we discuss the direct effects of these key complexities on the inference of parameters for continuous and pulsed illumination, respectively.

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

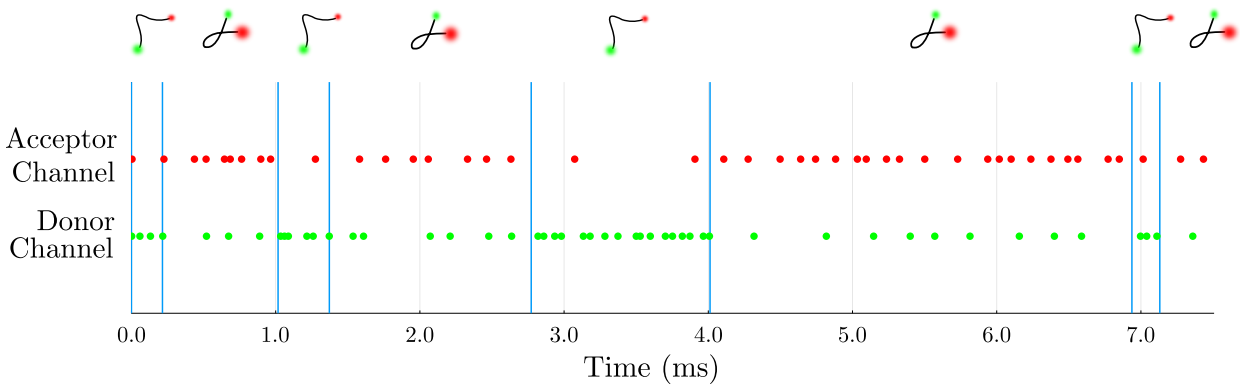

Figure 1: **A cartoon figure illustrating smFRET data**. For the experiments considered here, the dynamics along the reaction coordinate defined along the donor-acceptor distance are monitored using single photon arrival data. In the figure above, photon arrivals are represented by green dots for photons 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.

dynamics of proteins [5–7], nucleic acids [8], and their interactions [9, 10]. 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 [11], the intermediate stages of protein folding [12, 13], and the chromatin interactions modulated by the helper protein HP1α involved in allowing genetic transcription when the chromatin is tightly packed [14].

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 [15]. The donor, in turn, is illuminated by a laser in continuous or pulsating fashion to record a trace of photon arrival times, brightness values, and detection channels.

Upon excitation by the laser, a 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 [16, 17]. The distance dependence of this energy transfer is the key to using smFRET as a molecular ruler. Furthermore, this property 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

$$\varepsilon = \frac{n_A}{n_A + n_D},$$

where n_D and n_A are the number of donor and acceptor photons detected in a given time period, respectively.

The aim of smFRET is to capture on-the-fly changes in 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 detec-

tor's probabilistic response to an incoming photon [18]; 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 the course of an experiment [19–21].

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 [18, 22, 23].

Item 4 refers to a collection of effects including limited quantum yield and variable brightness due to blinking of dyes caused by nonradiative pathways [24, 25], photobleaching or permanent deactivation of the dyes [2, 24, 25], 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, 22], or a donor-acceptor pair's relative misalignment or positioning resulting in false signals and inaccurate characterization of the separation between labeled molecules [2, 26].

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. Binning is either achieved by directly binning photon arrivals when using single photon detectors [19, 27] or by integrating intensity over a few pixels when using widefield detectors [28].

While binned data analyses can be used to determine the number and connectivity of system states [29]–by computing average FRET efficiencies over bin time windows and using them in turn to construct FRET efficiency histograms [19, 21, 27, 30–32]–they come at the cost of averaging dynamics existing below a time bin not otherwise easily accessible [28, 33, 34]. 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, dynamics over binned time traces have also been extracted by supplementing these techniques with a hidden Markov model (HMM) treatment [19, 21, 30–32].

Using HMMs, histogeram analyses 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 kinetics between the hypothesized system states.

What is more, the histogram analysis' accuracy is determined by bin sizes where large bins may result in averaging of dynamics. 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 [34]. To address these challenges, we must infer continuous time trajectories below the bin size through, for example, the use of Markov jump processes [28] while retaining a binned, *i.e.*, discrete measurement model.

When single photon data is available, and to avoid the binning issues inherent to HMM analysis [28, 35, 36]–and leverage, for example, direct noise properties of detectors known for single photon arrivals (*e.g.*, IRF), and knowledge of excited photophysical state lifetimes available from pulsed illumination–single photon arrivals can be directly employed in the analysis. This, naturally, comes with added computational cost [33].

Often to help mitigate computational costs, further approximations on the system dynamics are often invoked such as assuming system dynamics to be much slower than FRET

label excitation and relaxation rates. This approximation helps decouple photophysical and system (molecular) dynamics [12, 33, 37, 38].

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 [12, 33, 37, 38]. 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 are 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 [21]; 2) system states may differ primarily in kinetics but not FRET efficiency [35]; 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) [19] or variational Bayesian [20] 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 dynamics in state space) and a unique maximum in the likelihood, both of are violated in smFRET data [20]. Furthermore, BIC and other such methods provide point estimates rather than full probabilities over system states ignoring uncertainty from items 1-4 propagated into model selection [39].

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 kinetics as random variables [21, 35, 36]. 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 specialize to the Bayesian nonparametric (BNP) paradigm [21, 36, 40–47]. 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 papers, 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, warranting 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 [48] 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 two subsequent companion manuscripts.

2 Forward Model

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

1. a biomolecular complex labeled with a FRET dye pair is always referred to as a *system*,
2. the configurations through which a system transitions are termed *system states*,
3. FRET dyes undergo quantum mechanical transitions between *photophysical states*,
4. a system-FRET combination is always referred to as a *composite*, and
5. a composite undergoes transitions among its *superstates*.

2.1 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. Upon relaxation, both donor and acceptor can emit photons via direct excitation by the laser or nonradiative transfer of energy from donor to acceptor. These photons are collected by single photon detectors, and their arrival times and detection channels are recorded. 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 varies 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.*, the 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. [49] and Ref. [50] 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.

2.2 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 $M_\phi = M_\psi \times M_\sigma$ superstates, $\{\phi_1, \ldots, \phi_{M_\phi}\}$, corresponding to all possible ordered pairs (ψ_j, σ_k) of the system and photophysical states. To be precise, we define $\phi_i \equiv (\psi_j, \sigma_k)$, where $i = (k-1)M_\psi + j$.

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 [35]

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

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. Here, $\mathbf{G}$ is the generator matrix of size $M_\phi \times M_\phi$ containing all transition rates $\lambda_{\phi_i \to \phi_j}$ between superstates (excluding self-transitions), with each diagonal element equal to the negative sum of the remaining transition rates within the corresponding row, that is, $\lambda_{\phi_i \to \phi_i} = -\sum_{j \neq i} \lambda_{\phi_i \to \phi_j}$, forcing each row-sum to be zero. Furthermore, assuming no simultaneous transitions among system states and photophysical states, that is, $\lambda_{(\psi_i, \sigma_j) \to (\psi_l, \sigma_m)} = 0$, we can simplify notation further by redefinitions, $\lambda_{(\psi_i, \sigma_j) \to (\psi_i, \sigma_k)} \equiv \lambda_{\sigma_j \to \sigma_k}$ and $\lambda_{(\psi_i, \sigma_j) \to (\psi_k, \sigma_j)} \equiv \lambda^{(j)}_{\psi_i \to \psi_k}$ for photophysics occurring in the j-th system state. This leads to the following form for the generator matrix containing

blocks of exclusively photophysical and exclusively system transition rates, respectively:

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

where the matrices on the diagonal $\mathbf{G}_\psi^{(i)}$ are the photophysical parts of the generator matrix when the system is in its i-th system state. Additionally, $\mathbb{I}$ is the identity matrix of size M_ψ.

The explicit form of $\mathbf{G}_\psi^{(i)}$ depends on the photophysical transitions that are allowed in the model. For instance, if the FRET pair is allowed to go from its ground photophysical state (ψ_1) to excited donor (ψ_2) or excited acceptor (ψ_3) only, the matrix is given as

$$\mathbf{G}_\psi^{(i)} = \begin{bmatrix} * & \lambda^{(i)}_{\psi_1\to\psi_2} & \lambda^{(i)}_{\psi_1\to\psi_3} \\ \lambda^{(i)}_{\psi_2\to\psi_1} & * & \lambda^{(i)}_{\psi_2\to\psi_3} \\ \lambda^{(i)}_{\psi_3\to\psi_1} & 0 & * \end{bmatrix} = \begin{bmatrix} * & \lambda_{ex} & \lambda_{direct} \\ \lambda_d & * & \lambda^{(i)}_{FRET} \\ \lambda_a & 0 & * \end{bmatrix}, \tag{2}$$

where the $*$ along the diagonal represents 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 $\lambda^{(i)}_{FRET}$ is the donor to acceptor FRET transition rate when the system is in its i-th system state. Note that the transition rates other than for FRET here do not depend on the system state.

With the generator matrix in hand, we now look for the solutions to the master equation above. Due to its linearity, the master equation simply accommodates the following analytical solution:

$$\boldsymbol{\rho}(t) = \boldsymbol{\rho}(t_0)\ \mathbf{exp}\left((t-t_0)\mathbf{G}\right) \equiv \boldsymbol{\rho}(t_0)\ \boldsymbol{\Pi}(t-t_0), \tag{3}$$

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 $\boldsymbol{\Pi}(t-t_0)$). The exponential maps the transition rates $\lambda_{\phi_i\to\phi_j}$ in the generator matrix to their corresponding transition probabilities $\pi_{\phi_i\to\phi_j}$ that populate 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, $\sum_j \pi_{\phi_i\to\phi_j} = 1$).

When there are no observations, any transition among the set of superstates is allowed. However, when monitoring the system using suitable detectors, observations rule out specific transitions at the observation time. For example, 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 detection 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 a continuous quantity defined at every time point.

Additionally, since smFRET traces can only specify the radiative transitions of the FRET labels at photon arrival times, uncertainty remains about the occurrences of other unmonitored transitions (*e.g.*, between system states). To put it another way, smFRET traces (observations) can only partially specify superstates at any given time.

As mentioned before, our goal is to compute the likelihood for such smFRET traces. This can be accomplished by summing the probabilities for all superstate trajectories that could give rise to a given set of observations from the beginning of an experiment T_{start} to its end T_{end}. Assuming the system ends up in the superstate ϕ_i at T_{end}, this sum over all the possible trajectories is given by the element of the propagated vector $\boldsymbol{\rho}(T_{end})$ corresponding to superstate ϕ_i. Therefore, the likelihood can be computed as

$$L = p(\phi_i) = [\boldsymbol{\rho}(T_{end})]_i. \tag{4}$$

However, as the final superstate at time T_{end} is usually unknown, we often marginalize (sum) over the final superstate to get the likelihood. That is,

$$L = \sum_{i=1}^{M_\phi} p(\phi_i) = \boldsymbol{\rho}(T_{end})\, \boldsymbol{\rho}_{norm}^T, \tag{5}$$

where all elements of the vector $\boldsymbol{\rho}_{norm}$ are 1 to help sum the probabilities in vector $\boldsymbol{\rho}(T_{end})$.

2.2.1 Absence of Observations

For pedagogical reasons, it is helpful to first look at the trivial case where a system-FRET composite evolves as usual 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 and end times of the experiment, since the superstate cannot be specified at any given time by observations. Consequently, the probabilities for all such trajectories must sum to one by virtue of the probability vector $\boldsymbol{\rho}(t)$ remaining normalized the whole time, that is,

$$L = p((T_1, e_1), \ldots, (T_K, e_K)|\mathbf{G}) = \boldsymbol{\rho}(T_{start})\boldsymbol{\Pi}(T_{end} - T_{start})\, \boldsymbol{\rho}_{norm}^T = \boldsymbol{\rho}(T_{end})\, \boldsymbol{\rho}_{norm}^T = 1, \tag{6}$$

where $\{e_1, \ldots, e_K\}$ are the emission times of all emitted photons, not recorded due to lack of detection channels.

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 the presence of detectors partially collapses knowledge of the occupied superstate during the experiment, resulting in a likelihood smaller than one. We do so for the conceptually simpler case of continuous illumination for now.

2.2.2 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 results from ongoing transitions and various detection features (*e.g.*, crosstalk, detector efficiency).

Unlike the case of binned smFRET traces, where observations only depend on the current system state (thereby justifying a regular, "first order", HMM treatment), an observation

in a single photon setup depends on the initial and final superstates of a set of transitions. This naturally enforces a second order HMM structure where each observed random variable depends on two superstates. We keep this point in mind as it will require us to develop filtering formulas for second order HMMs.

Now, in order to apply second order HMMs, we first divide the experiment's time duration into N windows of size $\epsilon = (T_{end} - T_{start})/N$, which we eventually make infinitesimally small. We also consider transitions between superstates in each window (see Fig. 2 (a)). These windows are marked by the times

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

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).

However, to simplify computation of the likelihood, 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 convert our second order HMM into a computationally convenient first order HMM, as 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)). We will eventually take the $\tau \to 0$ limit to retrieve our original second order HMM. The sequence of superstates now looks like

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

With all random variables involved in the forward model now defined, we next describe how to compute the probability of an observation given a pair of consecutive superstates and detection noise, $p(w_n|a_n, b_n, \mathbf{G})$, and the probability of superstate evolution, $p(b_n|a_n, \mathbf{G})$, from the propagator. These probabilities are key to selecting possible superstate trajectories for the computation of the likelihood.

2.2.3 Recursion Formulas

The terminal probability vector $\boldsymbol{\rho}_{end} = \boldsymbol{\rho}(T_{end})$ when observations are present can now be computed by evolving the initial vector $\boldsymbol{\rho}_{start} = \boldsymbol{\rho}(T_{start})$. Following the standard filtering

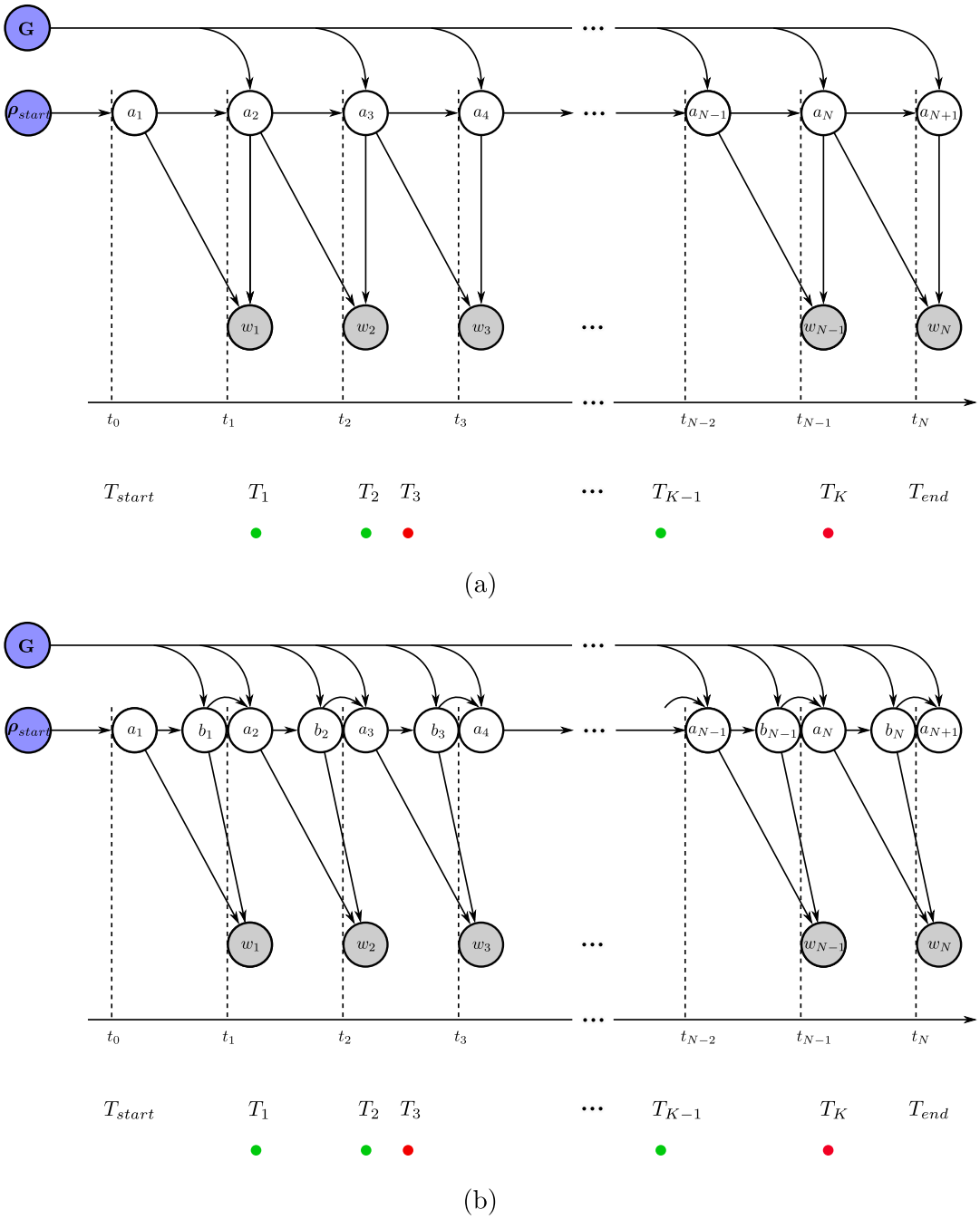

Figure 2: **Graphical models depicting the random variables and parameters involved in the generation of photon arrival data for smFRET experiments**. Circles shaded in blue represent parameters of interest we wish to deduce, namely transition rates and probabilities. The circles shaded in gray correspond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

theory [51], this is most conveniently achieved by recursively marginalizing over all superstates backwards in time starting from the superstate a_{N+1} as follows:

$$L = p(w_{1:N}|\mathbf{G}, \boldsymbol{\rho}_{start}) = \sum_{a_{N+1}} p(w_{1:N}, a_{N+1}|\mathbf{G}, \boldsymbol{\rho}_{start}) \tag{7}$$

$$= \sum_{a_{N+1}} \mathcal{A}_{N+1}(a_{N+1}), \tag{8}$$

where the filter at the beginning of the nth time window is related to the filter at the end of the previous window as follows

$$\begin{aligned}\mathcal{A}_n(a_n) &= p(w_{1:n-1}, a_n|\mathbf{G}, \boldsymbol{\rho}_{start}) \\ &= \sum_{b_{n-1}} p(a_n|b_{n-1}, \mathbf{G})\, \mathcal{B}_{n-1}(b_{n-1}),\end{aligned} \tag{9}$$

and the filter at the end of this window is related to the filter at the beginning of the time window by

$$\begin{aligned}\mathcal{B}_n(b_n) &= p(w_{1:n-1}, b_n|\mathbf{G}, \boldsymbol{\rho}_{start}) \\ &= \sum_{a_n} p(w_n|a_n, b_n, \mathbf{G})\, p(b_n|a_n, \mathbf{G})\, \mathcal{A}_n(a_n).\end{aligned} \tag{10}$$

Here, the last terms in the recursion formula, $\mathcal{A}_1(a_1)$, are the elements of the vector $\boldsymbol{\rho}_{start}$ itself. This two step filtering process is analogous to the forward filter of a first order HMM.

2.2.4 Reduced Propagators

In this section, we derive different terms in the recursive formula above and write them in the matrix form to facilitate both notation and computation. To do so, we first note that the transition probabilities $p(b_n|a_n, \mathbf{G})$ and $p(a_n|b_{n-1}, \mathbf{G})$ in the equations above do not involve observations. As such, they can be computed as elements of the full propagator matrices (exactly as in Eq. 3)

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

and

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

respectively. On the other hand, the term $p(w_n|a_n, b_n, \mathbf{G})$ includes observations and can be structured into a matrix $\mathbf{M}_n$ of the same size as the propagator with elements $(\mathbf{M}_n)_{a_n \to b_n} = p(w_n|a_n, b_n, \mathbf{G})$. 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. 10 can now be written as

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

which related the reduced propagator $(\mathbf{\Pi}_n^{(r)})_{a_n \to b_n}$ is in the presence of observations to the full propagator (no observations). Furthermore, here, the detection matrices cannot yet be computed analytically as the observations w_n allow for an unknown number of transitions within the finite time window (t_{n-1}, t_n). However, they become manageable as the time windows become vanishingly small, as we demonstrate later.

2.2.5 Likelihood for the Second Order HMM

Finally, rewriting the recursion formulas (Eqs. 9 & 10) using the introduced reduced propagator matrices, we arrive at the likelihood

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

Under the limit $\tau \to 0$,

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

where $\mathbb{I}$ is the identity matrix and

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

Therefore, we recover the likelihood of the second order HMM as

$$L = \boldsymbol{\rho}_{start}\, \mathbf{\Pi}_1^{(r)}\, \mathbf{\Pi}_2^{(r)}\, \mathbf{\Pi}_3^{(r)} \ldots \mathbf{\Pi}_{N-1}^{(r)}\, \mathbf{\Pi}_N^{(r)} \boldsymbol{\rho}_{norm}^T\,. \tag{11}$$

The structure of the likelihood above amounts to propagation of the initial probability vector $\boldsymbol{\rho}_{start}$ to the final probability vector $\rho(T_{end})$. However, the sum of probabilities in $\boldsymbol{\rho}(T_{end})$ is less than unity upon propagation due to the reduced number of superstate trajectories in agreement with observations. That is,

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

2.2.6 Continuum Limit

Up until now, observations allowed for an unknown number of transitions in each time window (t_{n-1}, t_n) which made it impossible to compute the exact form of the detection matrices. Taking the continuum limit, where the time windows become vanishingly small ($\epsilon \to 0$ as $N \to \infty$), resolves this problem by reasonably assuming that no more than one transition is allowed per window. This allows us to fully specify the detection matrices $\mathbf{M}_n$.

For example, in an idealized experiment with detectors operating at 100% efficiency, if no photons are detected while detectors are active, only nonradiative transitions can have nonzero probabilities in the detection matrices, that is, we have

$$(\mathbf{M}_n)_{a_n \to b_n} = (\mathbf{M}^{non})_{a_n \to b_n} = \begin{cases} 1 & \text{Nonradiative transitions} \\ 0 & \text{Radiative transitions} \end{cases}. \tag{12}$$

On the other hand, when the k-th photon is recorded in a time window, only radiative transitions have nonzero probabilities, that is,

$$(\mathbf{M}_k^{rad})_{a_n \to b_n} = \begin{cases} 0 & \text{All transitions except for the } k\text{-th photon emission} \\ 1 & k\text{-th photon emission} \end{cases}, \tag{13}$$

where the radiative detection matrices have zeros along their diagonals, since self transitions are nonradiative.

Additionally, only a simple modification of the probabilities in the detection matrices above is required to compute the likelihood when incorporating various features of realistic detectors, as we later demonstrate in Sec. 2.3.

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

$$\mathbf{\Pi}^{(r)non} = (\mathbb{I} + \epsilon\mathbf{G} + \mathcal{O}(\epsilon^2)) \odot \mathbf{M}^{non} = \mathbf{exp}(\epsilon\mathbf{G}^{non}) + \mathcal{O}(\epsilon^2)\,, \text{ and} \tag{14}$$

$$\mathbf{\Pi}_k^{(r)rad} = (\mathbb{I} + \epsilon\mathbf{G} + \mathcal{O}(\epsilon^2)) \odot \mathbf{M}_k^{rad} = \epsilon\mathbf{G}_k^{rad} + \mathcal{O}(\epsilon^2). \tag{15}$$

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

2.2.7 Final Likelihood

With the asymptotic forms of the reduced propagators in Eq. 11 now defined in the last subsection, we have all the ingredients needed to arrive at the final form of the likelihood.

The nonradiative propagators in Eq. 11 within the period between the $(k-1)$-th photon and the k-th photon can now be easily merged into a single propagator $\mathbf{\Pi}_k^{non}$, as the commutative arguments of the exponentials can be trivially 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

$$\mathbf{\Pi}_k^{non}\mathbf{\Pi}_k^{(r)rad} = \epsilon\mathbf{\Pi}_k^{non}\mathbf{G}_k^{rad} + \mathcal{O}(\epsilon^2), \tag{16}$$

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

Inserting Eq.16 for each interphoton period into the likelihood for the second order HMM in Eq. 11, we finally arrive at our desired likelihood

$$L = \epsilon^K \boldsymbol{\rho}_{start}\ \mathbf{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \mathbf{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \mathbf{\Pi}_{K-1}^{non}\mathbf{G}_{K-1}^{rad}\ \mathbf{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \mathbf{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{norm}^T + \mathcal{O}(\epsilon^{K+1}). \tag{17}$$

2.3 Detector Effects

Realistic FRET experiments typically do not have idealized detectors as we assumed in the previous section to illustrate detection matrices. Many emitted photons may simply go undetected when 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 [18].

Additionally, many donor photons may be detected in the channel reserved for the acceptor labels or vice-versa due to overlap in the emission spectrum [52]. This phenomenon,

commonly known as crosstalk or 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 detector receives a photon) [18]. Since dark current is generally calibrated together with the background emissions of the experimental setup, we include it in Sec. 2.5. We describe how the rest of these effects can be incorporated into our model in the following subsections.

2.3.1 Crosstalk and detection efficiency

The existance of crosstalk and detection efficiency add noises that make the detection of an emitted photon itself a stochastic process. These effects can be incorporated into the propagators by substituting zeros and ones in the radiative and nonradiative detection matrices (eqs. 12-13) with associated probabilities. The resulting propagators, in turn, modify the likelihood incorporating the effects of crosstalk and detection efficiency.

For example, the radiative detection matrix in Eq. 13 for the arrival of the k-th photon into channel i becomes

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

where ϕ_{ji} is the probability for photons intended for channel j (upon transition from superstate a_n to b_n) to be registered by channel i. These probabilities represent the combined effect of having crosstalk due to spectral overlap and inefficiencies of detection channels. Therefore, $\sum_i \phi_{ji} \neq 1$, 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. 15 for the k-th photon:

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

On the other hand, for time periods when no photons are detected, the nonradiative detection matrices in Eq. 12 become

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

where we have summed over the probabilities for the photon intended for channel i to not be registered upon entering channel j. The nonradiative propagator of Eq. 14 now modifies in this case to

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

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 is now

governed by the product

$$\mathbf{\Pi}_k^{non-ct}\mathbf{\Pi}_k^{(r)rad-ct} = \epsilon\mathbf{\Pi}_k^{non-ct}\mathbf{G}_k^{rad-ct} + \mathcal{O}(\epsilon^2), \tag{18}$$

where $\mathbf{\Pi}_k^{non-ct}$ combines all nonradiative reduced propagators in the time interval between the $(k-1)$-th and k-th photons.

Finally, inserting Eq.18 for each interphoton period into the likelihood for the second order HMM in Eq. 11, we arrive at the likelihood that incorporates crosstalk and detection efficiency as

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

Now, to calibrate crosstalk probabilities/detection efficiencies ϕ_{ij}, two samples containing only donor dyes and another containing only acceptor dyes can be individually excited with a donor laser under same power to determine the number of photons detected in various channels n_i^{raw}. The ratios of these photon counts can then be computed with respect to total counts (see [53] for more details) to obtain the crosstalk probabilities and detection efficiencies.

In many cases, the calibrated values are encoded in a route correction matrix (RCM) [53] that is used by experimentalists to directly compute corrected counts and is given as (for a two channel setup)

$$\begin{bmatrix} n_1^{corr} \\ n_2^{corr} \end{bmatrix} = \mathbf{RCM} \begin{bmatrix} n_1^{raw} \\ n_2^{raw} \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} n_1^{raw} \\ n_2^{raw} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{21} \\ \phi_{12} & \phi_{22} \end{bmatrix}^{-1} \begin{bmatrix} n_1^{raw} \\ n_2^{raw} \end{bmatrix},$$

where n_i^{corr} are the corrected photon counts for channel i.

Often, the RCM is presented normalized with respect to its $(1,1)$ element, that is,

$$\mathbf{RCM}' = \frac{\mathbf{RCM}}{\mathbf{RCM}_{1,1}} = \begin{bmatrix} 1 & -\frac{\phi_{21}}{\phi_{22}} \\ -\frac{\phi_{12}}{\phi_{22}} & \frac{\phi_{11}}{\phi_{22}} \end{bmatrix}.$$

Once this matrix above is experimentally determined, the crosstalk factors/detection efficiencies ϕ_{ij} can be computed via simple algebraic manipulations. A similar procedure can be performed in cases where more than two channels are present.

2.3.2 Effects of detector dead time

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

To incorporate this detector dead period into the likelihood of an smFRET trace, we follow the same procedure as in the last subsection where we first define appropriate detection matrices for various combinations of active and inactive detectors. For instance, for an inactive channel i_k after the detection of the k-th photon, we can define the detection matrix as

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

Next, corresponding to this detection matrix, we have the propagator $\mathbf{\Pi}_k^{i_k-dead} = \mathbf{exp}(\delta_{i_k}(\mathbf{G}\odot \mathbf{M}^{i_k-dead})) = \mathbf{exp}(\delta_{i_k}\mathbf{G}^{i_k-dead})$, that evolves the superstate during the detection dead period. This propagator can now be combined with the nonradiative and radiative propagators to represent the evolution during the period between $k-1$-th and k-th photons as

$$\mathbf{\Pi}_{k-1}^{i_{k-1}-dead}\mathbf{\Pi}_k^{non}\mathbf{\Pi}_k^{(r)rad} = \epsilon\mathbf{\Pi}_{k-1}^{i_{k-1}-dead}\mathbf{\Pi}_k^{non}\mathbf{G}_k^{rad} + \mathcal{O}(\epsilon^2). \tag{19}$$

Finally, inserting Eq.19 for each interphoton period into the likelihood for the second order HMM in Eq. 11, we arrive at the following likelihood that includes detector inactivity:

$$L \propto \boldsymbol{\rho}_{start}\ \mathbf{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \mathbf{\Pi}_1^{i_1-dead}\mathbf{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \mathbf{\Pi}_2^{i_2-dead}\ \ldots\ \mathbf{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \mathbf{\Pi}_K^{i_K-dead}\mathbf{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{norm}^T. \tag{20}$$

To understand the effect of detector dead time on the likelihood, we take a detour for pedagogical reasons and 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), being illuminated by a laser. The data in this case contains only photon arrival times

$$\{T_1, T_2, T_3, \ldots, T_K\}. \tag{21}$$

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 negative row-sum of the remaining elements, λ_{ex} is the excitation rate, and λ_d is the donor relaxation rate.

Here, since all transitions are allowed during detector dead times, propagators $\mathbf{\Pi}_k^{i_k-dead}$ in the likelihood (Eq. 20) are simply exponentials of the full generator matrix, and therefore do not modify the normalization of the propagated probability vector $\boldsymbol{\rho}$.

In effect, we model the fact that there are no observations (not even non-detections) or data collected during the dead time, and all possible superstate trajectories are allowed. Consequently, the distributions of the parameters of interest are only affected by virtue of having less information (observations) collected during a time period. This is different from other sources of uncertainty, such as background emissions and cross talk, that insert false information into our data by misidentifying photons.

Similarly, in realistic situations, with multiple fluorophore species and associated detection channels, inactive detectors only reduce the amount of information available to learn distributions over parameters of interest.

2.3.3 Adding the IRF

Due to various sources of noise affecting the timing electronics, the time elapsed between photon arrival and detection is itself a hidden (latent) random variable [18] sampled from a probability density, $f(\tau)$, termed the instrument response function (IRF). Here, unlike in the previous subsections, we do not need to define a new set of detection matrices as the stochasticity is now in the detection time and not the detection itself.

To incorporate the IRF in the likelihood, we introduce a convolution of the propagators with $f(\tau)$ for every photon arrival time as follows:

$$
\begin{aligned}
L = \boldsymbol{\rho}_{start} & \left(\int_0^{\tau_{IRF}} d\tau_1 \mathbf{\Pi}_1^{non}(\Delta T_1 - \tau_1)\ \mathbf{G}_1^{rad}\ \mathbf{\Pi}_1^{i_1-dead}(\tau_1)\ f(\tau_1) \right) \\
& \times \left(\int_0^{\tau_{IRF}} d\tau_2 \mathbf{\Pi}_2^{non}(\Delta T_2 - \tau_2)\ \mathbf{G}_2^{rad}\ \mathbf{\Pi}_2^{i_2-dead}(\tau_2)\ f(\tau_2) \right) \\
& \vdots \\
& \times \left(\int_0^{\tau_{IRF}} d\tau_K \mathbf{\Pi}_K^{non}(\Delta T_K - \tau_K)\ \mathbf{G}_K^{rad}\ \mathbf{\Pi}_K^{i_K-dead}(\tau_K)\ f(\tau_K) \right)\ \mathbf{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{norm}^T, \quad (22)
\end{aligned}
$$

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.

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 system with the smallest acquisition time of 16 ps and a detection IRF distribution that is approximately 100 ps wide will have the 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 not adding any extra real computational time other than the overhead associated with parallelization.

2.4 Illumination Features

After discussing the detector effects, we continue here by further considering different illumination procedures. For simplicity alone, our likelihood derivation earlier assumed continuous illumination with a constant intensity. Here, we show how other illumination setups (such as pulsed illumination and alternating laser excitation, ALEX [54])can be incorporated into the likelihood by simply assigning a time dependence to the excitation rates in the generator matrix.

For instance, the likelihood for the case of continuous illumination (Eq. 17) can be modified to accommodate fluorescent labels illuminated by a train of regularly spaced laser pulses separated by a period τ. To do so, we start by introducing uncertainties in the instants at which excitation (δ_n) and detection (μ_n) occur (see Fig. 3). This uncertainty arises due to finite excitation period modeled with a profile of width Δ_{IRF1} (shown in yellow in Fig. 3), and stochastic delay in reporting the photon detection time (also known as jitter) modeled by an IRF of width Δ_{IRF2} (shown in red). Furthermore, Fig. 3 also depicts an example of a photophysical trajectory within an interpulse period. Note that while the figure depicts excitation and relaxation occurring before the next pulse, the framework can treat any combination of events including remaining excited over multiple pulses.

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 involving products of matrices as follows:

$$
L \propto \boldsymbol{\rho}_{start}\ \mathbf{Q}_1\ \mathbf{Q}_2\ \mathbf{Q}_3\ \ldots\ \ \mathbf{Q}_{N-1}\ \mathbf{Q}_N\ \boldsymbol{\rho}_{norm}^T, \quad (23)
$$

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.

Depending on the observations over an interpulse period, these $\mathbf{Q}_n$ can take different forms. For example, when there are no photons detected, the result is

$$\mathbf{Q}_n = \mathbf{exp}\left(\int_0^{\Delta_{IRF1}} d\delta_n \mathbf{G}^{non}(\delta_n)\right) \mathbf{exp}\left((\tau - \Delta_{IRF1})\,\mathbf{G}^{dark}\right), \tag{24}$$

where the first exponential contains integral over excitation delay δ_n to account for time varying excitation rate due to the laser pulse, and the second term represents evolution without any illumination. Here, $\mathbf{G}^{non}$ is the nonradiative generator matrix as defined earlier (see Sec. 2.2.6), and $\mathbf{G}^{dark}$ is the same matrix with the excitation rate set to zero.

On the other hand, if there is a photon detection sometime after a pulse (as in Fig. 3), convolution with the detection IRF must be included (as demonstrated in Sec. 2.3.3) to give

$$\mathbf{Q}_n = \mathbf{exp}\left(\int_0^{\Delta_{IRF1}} d\delta_n \mathbf{G}^{non}(\delta_n)\right) \int_0^{\Delta_{IRF2}} d\epsilon_n\, \mathbf{exp}\left((\mu_n - \Delta_{IRF1} - \epsilon_n)\,\mathbf{G}^{dark}\right) \mathbf{G}^{rad} \times \mathbf{exp}\left((\tau - \mu_n + \epsilon_n)\,\mathbf{G}^{dark}\right) f(\epsilon_n), \tag{25}$$

where the new integration variable ϵ_n represents the stochastic delay in detection timing, and f is the IRF.

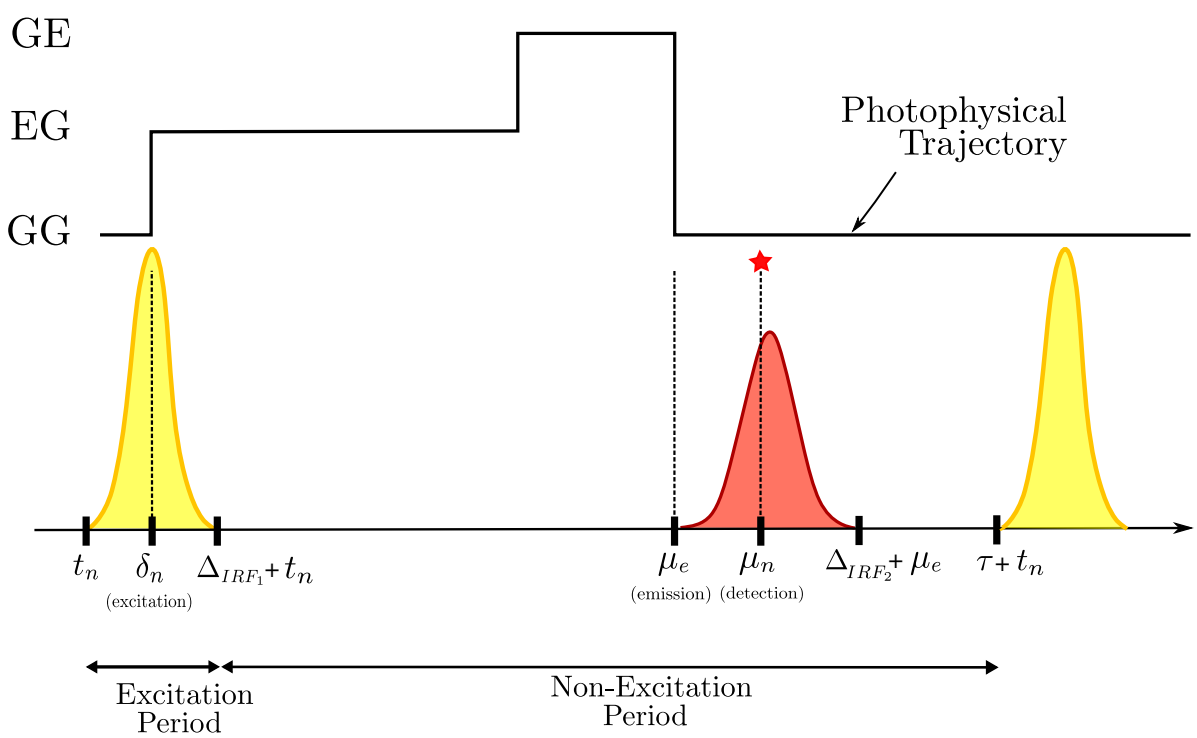

Figure 3: **Events over a pulsed illumination experiment pulse window**. Here, the beginning of the n-th interpulse window of size τ is marked by time t_n. The sample is then excited by a high intensity burst (shown by the yellow Gaussian) for a very short time Δ_{IRF_1}. If excited, the fluorophore then emits a photon at μ_e. However, detection, highlighted with a red star, occurs at time μ_n dictated by the IRF (shown by the red Gaussian).

In cases where the pulses are short enough that emission never occurs during the pulse duration itself, we can avoid calculating the integral in Eq. 24

$$\mathbf{\Pi}^{non} = \mathbf{exp}\left(\int_0^{\Delta_{IRF1}} d\delta_n \mathbf{G}^{non}(\delta_n)\right), \tag{26}$$

and estimate it directly along with other parameters, as it is identical for all of the pulses. Later, in the third companion manuscript [55], we will use this approximation to provide a more computationally advantageous model.

In the most general case, however, it is possible for photon detections to occur during an illumination pulse. The resulting $\mathbf{Q}_n$ would then be

$$\mathbf{Q}_n = \int_0^{\Delta_{IRF2}} d\epsilon_n \mathbf{exp}\left(\int_0^{\mu_n-\epsilon_n} d\delta_n \mathbf{G}^{non}(\delta_n)\right) \mathbf{G}^{rad}\, \mathbf{exp}\left(\int_{\mu_n-\epsilon_n}^{\Delta_{IRF1}} d\delta_n \mathbf{G}^{non}(\delta_n)\right) f(\epsilon_n) \times \mathbf{exp}\left((\tau - \Delta_{IRF1})\,\mathbf{G}^{dark}\right).$$

Since here the integral over the pulse is split into periods before and after detection, we cannot avoid it by learning a propagator like $\mathbf{\Pi}^{non}$. Instead, we must now be able to compute the integral $\int_0^t d\delta_n \mathbf{G}^{non}(\delta_n)$ for any time t between zero and Δ_{IRF1}. To do this, we take advantage of the fact that: 1) we know the laser pulse intensity profile $I(t)$; and 2) the emission intensity of fluorescent molecules is proportional to the laser intensity [56] to approximate the time dependent excitation rates λ_{ex} in $\mathbf{G}^{non}$ as

$$\lambda_{ex}(t) = kI(t),$$

where k is a precalibrated proportionality constant that relates the intensity of incident light to the fluorophore excitation rate λ_{ex}.

2.5 Background Emissions

Here, we consider background photons entering detectors from sources other than the system under study [2]. Additionally, false photons (dark current) may be recorded by the detectors [18]. While these sources of photon detections are typically complicated, we may still assume exponentially distributed waiting times for the background photons with fixed rates for each detection channel [57, 58].

Since background introduces new photophysical transitions into the model, we only need to expand the full generator matrix $\mathbf{G}$ to incorporate it into the likelihood, leaving the form of the likelihood itself exactly the same as in Eq. 17.

In order to include background rates in the generator matrix, we basically treat background in each detection channel as emanating from independent emitters, respectively. These sources have two photophysical states, each with identical rates for transitions among themselves, to ensure that all inter-background photon times are distributed according to the same exponential distribution. We denote these photophysical states with

$$\{\psi_1^{b_i},\ \psi_2^{b_i}\}$$

for the detection channel i. As such, the corresponding background generator matrix for that channel is

$$\mathbf{G}_b^i = \begin{bmatrix} * & \lambda_{\psi_1^{b_i} \to \psi_2^{b_i}} \\ \lambda_{\psi_2^{b_i} \to \psi_1^{b_i}} & * \end{bmatrix} = \begin{bmatrix} * & \lambda_b^i \\ \lambda_b^i & * \end{bmatrix}.$$

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 and background) can now be computed by combining the system-FRET composite state space and the background state spaces for all of the total C detection channels. This is accomplished using Kronecker sums as

$$\mathbf{G} = \mathbf{G}_{no-b} \oplus \mathbf{G}_b^1 \oplus \mathbf{G}_b^2 \oplus \ldots \oplus \mathbf{G}_b^C,$$

where the symbol $\oplus$ denotes the matrix Kronecker sum [59], 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 computed by exponentiating the expanded generator matrix above as

$$\begin{aligned} \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}) = \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_{no-b}) \otimes \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_b^1) \\ \otimes\, \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_b^2) \otimes \ldots \otimes \mathbf{exp}((T_k - T_{k-1})\,\mathbf{G}_b^C), \end{aligned}$$

where the symbol $\otimes$ denotes the matrix Kronecker product [59].

Furthermore, the same detection matrices that were 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. 2.2.6.

Consequently, as mentioned earlier, by contrast to incorporating 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 in Eq. 17.

2.6 Fluorophore Characteristics: Quantum Yield, Blinking, Photobleaching, and Acceptor Direct 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 without radiative transitions to the fluorophore model. These photophysical states can have long or short lifetimes depending on the specific photophysical phenomenon. For example, donor photobleaching can be included by introducing a third donor photophysical state into the matrix in Eq. 2 without any escape transitions as follows:

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

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 $\lambda_{FRET}^{(i)}$ 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 [46, 60], that is,

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

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

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

As we showed before when adding background, the structure of the likelihood (arrangement of the propagators) stays the same as in Eq. 17.

2.7 Synthetic Data Generation

In the previous subsections, we described how to compute the likelihood, which is the sum of probabilities for 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. We then use the generated data to test our BNP-FRET sampler.

For simplicity of demonstration only, we consider a molecule with two system states labeled with an acceptor-donor pair under continuous illumination. With two fluorophores, we have three possible photophysical states:

$$\begin{aligned} \psi_1 &= (G, G), \\ \psi_2 &= (E, G), \text{ and} \\ \psi_3 &= (G, E), \end{aligned}$$

where G denotes ground and E denotes the excited photophysical state, and the ordered pair represents the combined photophysical states of the FRET pair. We have ignored the highly improbable doubly excited photophysical state, (E, E), as only donor wavelengths are typically used for excitation which eliminates the possibility of simultaneous donor and acceptor excitation. Moreover, we assume that the relaxation times of the singly excited photophysical states above are much shorter in relation to typical interphoton times to allow for quick FRET and donor re-excitation.

A molecule labeled with such a FRET pair and two system states $\{\sigma_1, \sigma_2\}$ has in total six superstates: $\{\phi_1, \ldots, \phi_6\}$. As show in Sec. 2.2, we can write the generator matrix that governs the evolution of such a composite as

$$
\begin{aligned}
\mathbf{G} &= \begin{bmatrix} \mathbf{G}_{\psi}^{(1)} - \lambda_{\sigma_1 \to \sigma_2}\mathbb{I} & \lambda_{\sigma_1 \to \sigma_2}\mathbb{I} \\ \lambda_{\sigma_2 \to \sigma_1}\mathbb{I} & \mathbf{G}_{\psi}^{(2)} - \lambda_{\sigma_2 \to \sigma_1}\mathbb{I} \end{bmatrix} \\
&= \begin{bmatrix} * & \lambda_{\psi_1 \to \psi_2} & \lambda_{\psi_1 \to \psi_3} & \lambda_{\sigma_1 \to \sigma_2} & 0 & 0 \\ \lambda_{\psi_2 \to \psi_1} & * & \lambda_{\psi_2 \to \psi_3}^{(1)} & 0 & \lambda_{\sigma_1 \to \sigma_2} & 0 \\ \lambda_{\psi_3 \to \psi_1} & \lambda_{\psi_3 \to \psi_2} & * & 0 & 0 & \lambda_{\sigma_1 \to \sigma_2} \\ \lambda_{\sigma_2 \to \sigma_1} & 0 & 0 & * & \lambda_{\psi_1 \to \psi_2} & \lambda_{\psi_1 \to \psi_3} \\ 0 & \lambda_{\sigma_2 \to \sigma_1} & 0 & \lambda_{\psi_2 \to \psi_1} & * & \lambda_{\psi_2 \to \psi_3}^{(2)} \\ 0 & 0 & \lambda_{\sigma_2 \to \sigma_1} & \lambda_{\psi_3 \to \psi_1} & \lambda_{\psi_3 \to \psi_2} & * \end{bmatrix}, \\
&= \begin{bmatrix} * & \lambda_{ex} & \lambda_{direct} & \lambda_{\sigma_1 \to \sigma_2} & 0 & 0 \\ \lambda_{d} & * & \lambda_{FRET}^{(1)} & 0 & \lambda_{\sigma_1 \to \sigma_2} & 0 \\ \lambda_{a} & 0 & * & 0 & 0 & \lambda_{\sigma_1 \to \sigma_2} \\ \lambda_{\sigma_2 \to \sigma_1} & 0 & 0 & * & \lambda_{ex} & \lambda_{direct} \\ 0 & \lambda_{\sigma_2 \to \sigma_1} & 0 & \lambda_{d} & * & \lambda_{FRET}^{(2)} \\ 0 & 0 & \lambda_{\sigma_2 \to \sigma_1} & \lambda_{a} & 0 & * \end{bmatrix},
\end{aligned}
\tag{27}
$$

The values for the rates in the matrix above are chosen based on physical timescales encountered in smFRET experiments. In particular, later, we will use values motivated from Refs. [61, 62] for continuous illumination and Ref. [63] for pulsed illumination. The synthetic data can now be generated using the classic Gillespie algorithm [64], as described in the next section.

2.7.1 Dynamics

To simulate how the system-FRET composite evolves over time, we generate a superstate trajectory. This superstate trajectory is a continuous quantity, but it is convenient to write it as a sequence of superstates occurring at the end of each transition (excluding self-transitions). Since these superstates do not coincide with superstates a_n defined earlier in Sec. 2, we introduce new notation to represent the sequence of these superstates as

$$\{s_1, s_2, s_3, \ldots, s_N\},$$

where $s_i \in \{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6\}$. We further emphasize here that consecutive superstates s_{n-1} and s_n can never be same unlike earlier where a_{n-1} and a_n were allowed to be the same.

Now, to generate such a trajectory, we first sample the initial superstate as

$$s_1 \sim \textbf{Categorical}_{\phi_{1:6}}(\boldsymbol{\rho}_{start}),$$

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.

Since the superstate trajectory is assumed to be Markovian, we can iteratively sample the rest of the trajectory from the Gillespie algorithm [64] as

$$s_{n+1}|s_n, \mathbf{G} \sim \textbf{Categorical}_{\phi_{1:6}}\left(\frac{\lambda_{s_n\to\phi_1}}{\lambda_{s_n}}, \frac{\lambda_{s_n\to\phi_2}}{\lambda_{s_n}}, \frac{\lambda_{s_n\to\phi_3}}{\lambda_{s_n}}, \frac{\lambda_{s_n\to\phi_4}}{\lambda_{s_n}}, \frac{\lambda_{s_n\to\phi_5}}{\lambda_{s_n}}, \frac{\lambda_{s_n\to\phi_6}}{\lambda_{s_n}}\right),$$

where $\lambda_{s_n} = \sum_i \lambda_{s_n\to\phi_i}$ is the escape rate for the superstate s_n and rates for self-transitions are zero. These escape rates indicate faster relaxation from superstate s_n due to multiplicity of competing escape transitions.

Once the next superstate is chosen, the time until the next transition (the time spent in the current state) is sampled as follows:

$$H_{n+1}|s_n, \mathbf{G} \sim \textbf{Exponential}(\lambda_{s_n}).$$

Now that we have selected a superstate trajectory, we can apply our measurement model to assign a detection channel for the photons emitted upon radiative transitions, that is,

$$\tilde{c}_n \sim \textbf{Categorical}_{\{\varnothing,1,2\}}(p^n_\varnothing, p^n_1, p^n_2), \tag{28}$$

where the probabilities in the Categorical distribution above depend on the detector effects (such as crosstalk and detection efficiency) and whether a radiative transition occurs or not.

Finally, we collect all of the events where $\tilde{c}_n \in \{1, 2\}$, that is, a photon is detected in one of the detection channels, as an smFRET trace (see Fig. 4) consisting of photon arrival times $T_{1:K}$ and detection channels $c_{1:K}$ as

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

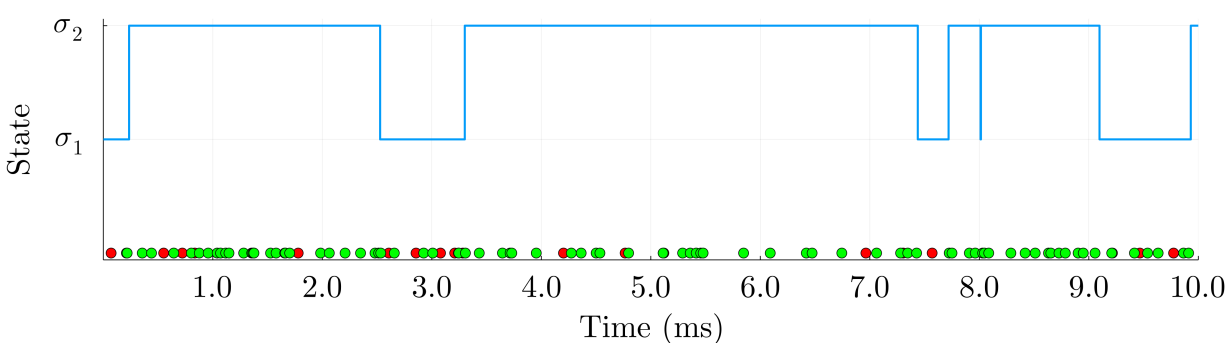

Figure 4: **Simulated data**. Artificially produced photon emission data for some chosen decay rates and excitation probabilities. The blue curve shows a molecule's trajectory transitioning between its system states σ_1 and σ_2. The green and red dots show the photon detections in the donor and acceptor channels, respectively.

3 Inverse Strategy

Now, armed with the likelihood for different experimental setups, we proceed to learn the parameters of interest including: transition rates that enter the generator matrix $\mathbf{G}$, and

probabilities for the system-FRET composite to be in a given superstate at the beginning of the experiment $\boldsymbol{\rho}_{start}$. However, accurate estimation of these parameters requires a stochastic inverse strategy capable of dealing with all the existing sources of uncertainty in the problem such as photon's stochasticity and detector noises. This naturally leads us to adopting 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 these parameters, termed posterior, as a product of the likelihood and a prior distribution using the Bayes' rule as

$$p(\mathbf{G}, \boldsymbol{\rho}_{start}|\boldsymbol{w}) \propto L(\boldsymbol{w}|\mathbf{G}, \boldsymbol{\rho}_{start})\, p(\mathbf{G}, \boldsymbol{\rho}_{start}), \tag{29}$$

where the last term $p(\mathbf{G}, \boldsymbol{\rho}_{start})$ is the prior distribution defined over the same domains as the parameters and 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, $1/p(\boldsymbol{w}_{1:N})$, and can be safely ignored as generation of samples using the Monte Carlo methods only involves ratios of posterior distributions or likelihoods.

Additionally, the ϵ^K factor in the likelihood first derived in Eq. 17 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):

$$p(\mathbf{G}, \boldsymbol{\rho}_{start}|\boldsymbol{w}) \propto \boldsymbol{\rho}_{start}\ \boldsymbol{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \boldsymbol{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \boldsymbol{\Pi}_{K-1}^{non}\mathbf{G}_{K-1}^{rad}\ \boldsymbol{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \boldsymbol{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{end}^T \times p(\mathbf{G}, \boldsymbol{\rho}_{start}). \tag{30}$$

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

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

where, motivated by computational reasons, we choose Dirichlet prior distribution for initial probabilities, that is,

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

where Dirichlet distribution is a multivariate generalization of the Beta distribution and $\boldsymbol{\zeta}$ is a vector of the same size as the superstate space. Additionally, we choose Gamma prior distributions for individual rates, that is,

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

to guarantee positive values. Here, α and λ_{ref} are the 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 expense, the number of parameters needed to learn can be reduced by reasonably assuming the system was at equilibrium immediately

preceding the time at which the experiment began. That is, instead of sampling 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. 30 now reduces to

$$p(\mathbf{G}|\boldsymbol{w}) \propto \boldsymbol{\rho}_{start}\ \boldsymbol{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \boldsymbol{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \boldsymbol{\Pi}_{K-1}^{non}\mathbf{G}_{K-1}^{rad}\ \boldsymbol{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \boldsymbol{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{end}^T \times p(\mathbf{G}). \tag{31}$$

In the following subsections, we first describe a parametric inverse strategy, *i.e.*, assuming a known number of system states, for sampling parameters from posterior distribution in Eq. 31 using Monte Carlo method. Next, we will generalize this inverse strategy to a nonparametric case where we also deduce the number of system states. Furthermore, we also describe strategies for practical computation of the likelihood.

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

Now, with the posterior at hand, in this section, we illustrate a sampling scheme to deduce the transition rates in the generator matrix $\mathbf{G}$. However, our posterior in Eq. 30 does not assume a standard form amenable to analytical calculations. Therefore, we must iteratively draw numerical samples using Markov Chain Monte Carlo (MCMC) techniques.

Specifically, we adopt the Gibbs algorithm to sequentially and separately generate samples for individual transition rates in each MCMC iteration. To do so, we first write the posterior in Eq. 31 using the chain rule as follows:

$$p(\mathbf{G}|\boldsymbol{w}) = p(\lambda_{\phi_i\to\phi_j}|\mathbf{G}\backslash\lambda_{\phi_i\to\phi_j}, \boldsymbol{w})p(\mathbf{G}\backslash\lambda_{\phi_i\to\phi_j}|\boldsymbol{w}), \tag{32}$$

where the backslash after $\mathbf{G}$ indicates exclusion of the following rate parameter. 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. 31) for the remaining rate parameters in the posterior on the left are also considered constant. Therefore, from Eqs. 31 & 32, we can write the conditional posterior for $\lambda_{\phi_i\to\phi_j}$ above as

$$p(\lambda_{\phi_i\to\phi_j}|\mathbf{G}\backslash\lambda_{\phi_i\to\phi_j}, \boldsymbol{w}) \propto \boldsymbol{\rho}_{start}\ \boldsymbol{\Pi}_1^{non}\mathbf{G}_1^{rad}\ \boldsymbol{\Pi}_2^{non}\mathbf{G}_2^{rad}\ \ldots\ \boldsymbol{\Pi}_K^{non}\mathbf{G}_K^{rad}\ \boldsymbol{\Pi}_{end}^{non}\ \boldsymbol{\rho}_{end}^T \times \mathbf{Gamma}\left(\lambda_{\phi_i\to\phi_j}; \alpha, \frac{\lambda_{ref}}{\alpha}\right). \tag{33}$$

Yet, the conditional posterior above does not take a closed form that allows direct sampling. As such, we turn to the Metropolis-Hastings (MH) algorithm [65], where new samples are drawn from a proposal distribution q and accepted with the probability

$$\alpha(\lambda^*_{\phi_i\to\phi_j}, \lambda_{\phi_i\to\phi_j}) = \mathbf{min}\left\{1, \frac{p(\lambda^*_{\phi_i\to\phi_j}|\boldsymbol{w}, \mathbf{G}\backslash\lambda_{\phi_i\to\phi_j})\ q(\lambda_{\phi_i\to\phi_j}|\lambda^*_{\phi_i\to\phi_j})}{p(\lambda_{\phi_i\to\phi_j}|\boldsymbol{w}, \mathbf{G}\backslash\lambda_{\phi_i\to\phi_j})\ q(\lambda^*_{\phi_i\to\phi_j}|\lambda_{\phi_i\to\phi_j})}\right\}, \tag{34}$$

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. 34. 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 which are naturally forbidden leading to rejection in the MH step and thus inefficient sampling. Therefore, it is more convenient to propose new samples using a Normal distribution in the 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 derivatives represent the Jacobian and the proposal distributions are canceled by virtue of using a Normal distribution.

The acceptance probability above depends on the difference of the current and proposed values for a given transitions rate. 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 optimum performance of the BNP-FRET sampler, or equivalently approximately one-fourth acceptance rate for the proposals [66].

This whole algorithm can now be summarized in the following pseudocode:

Initialize Chain of Samples
$j = 1$
for i in $1 : M_\sigma \times M_\sigma$

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

end

Iteratively sample from the posterior using Gibbs algorithm
for j in $2 :$ Draws
 for i in $1 : M_\sigma \times M_\sigma$

Propose new sample

$$\log(\lambda_i^*) \sim \textbf{Normal}\left(\log(\lambda_i^{(j-1)}), \sigma^2\right),$$

Compute acceptance probability

$$\alpha(\lambda_i^*, \lambda_i^{(j-1)}) = \textbf{min}\left\{1, \frac{p(\lambda_i^*|\boldsymbol{w}, \mathbf{G}\backslash\lambda_i)}{p(\lambda_i^{(j-1)}|\boldsymbol{w}, \mathbf{G}\backslash\lambda_i)} \frac{(\partial \log(\lambda_i)/\partial\lambda_i)^{(j-1)}}{(\partial \log(\lambda_i)/\partial\lambda_i)^*}\right\}$$

if $\alpha(\lambda_i^*, \lambda_i^{(j-1)}) > \textbf{rand}()$
 # Accept proposal
 $\lambda_i^{(j)} = \textbf{exp}(\lambda_i^*)$
else
 # Reject proposal
 $\lambda_i^{(j)} = \lambda_i^{(j-1)}$
end

 end
end

3.2 Nonparametrics: Predicting the Number of System States

After describing our parametric inverse strategy for the parametric case for a known number of system states, we turn to more realistic scenarios where we may not know the number of system states. In the following subsections, we first describe the BNP framework for the continuous illumination and then proceed to illustrate our BNP strategy for the pulsed illuminations case. Such BNP frameworks introduced herein, eventually, will provide us with distributions over the number of system states simultaneously with other model parameters.

3.2.1 Beta-Bernoulli Process for Continuous Illumination

In practice, the dimensionality of the system state space M_σ is not always known as molecules under study may exhibit complex and unexpected behaviors across conditions and timescales. Consequently, the dimensionality M_ϕ of the governing generator matrix $\mathbf{G}$ is also unknown, and must be determined by adopting a BNP framework [21, 34, 67] where we make the existence of a system state itself a random variable.

This can be achieved by first defining an infinite-dimensional version of the generator matrix in Eq. 27 and multiplying its elements with Bernoulli random variables b_i (also known as loads). These loads allow us to turn on/off a portion of the generator matrix associated with transitions between specific system states (including self-transitions). The resulting generator matrix now reads

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

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

where a load value of 1 represents an active system state, while inactive system states get a load value of 0, and the number of active loads is the number of system states estimated by the BNP-FRET sampler. As before, $*$ represents negative row-sums.

For using Bayesian inference methods, the previously defined posterior distribution (Eq. 31) should now incorporate the full set of loads, $\mathbf{b} = \{b_1, b_2, \dots, b_\infty\}$, as follows:

$$p(\mathbf{b}, \mathbf{G}, \boldsymbol{\rho}_{start}|\boldsymbol{w}) \propto L(\boldsymbol{w}|\mathbf{b}, \mathbf{G}, \boldsymbol{\rho}_{start})\, p(\mathbf{b}, \mathbf{G}, \boldsymbol{\rho}_{start}), \tag{35}$$

Here, we use a Bernoulli prior distribution for each load.

As in the case of the parametric BNP-FRET sampler presented in the previous subsection, 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 the rates using the MH procedure; then 2) loads by direct sampling, one-by-one from their corresponding conditional posteriors. Here, step (1) is identical to the parametric case in the previous subsection and we will only focus on the second step in what follows.

To sample the i-th load, the corresponding conditional posterior is given by [67]

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

where the backslash after $\mathbf{b}$ indicates exclusion of the following load. Here, we set the hyperparameters M_σ^{max}, the maximum allowed number of system states used in computations, and γ, expected number of system states based on empirically obtained intuition about the system under study.

Now, the conditional posterior in the equation above is discrete and describes the probabilities of the load being either active or inactive, that is, it is itself a Bernoulli distribution as follows:

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

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 simple standard form of this posterior allows direct sampling of the loads. In the end, the chain of generated samples can be used for subsequent statistical analysis.

We will show implementation of this method for synthetic and experimental data in the second manuscript of this series [68].

3.2.2 iHMM Methods for Pulsed Illumination

In this section, we describe the BNP framework for the analysis of smFRET data collected under pulsed illumination. To do so, we start by making a realistic approximations for this case. Given the timescale separation between the interpulse period and the system kinetics in a pulsed illumination experiment, it is possible to approximate the system state trajectory as being constant throughout the interpulse time. In essence, rather than treating the system state trajectory as a continuous time process, we treat it as a sequence of system states for each interpulse window and consider its transition probability matrix rather than its rate matrix. Analysis of such trajectories can be done through the use of HMM methods, and its nonparametric generalization, the infinite HMM (iHMM) [36, 42].

As the name suggests, the iHMM is fundamentally an infinite state space method, where, unlike the Bernoulli process described in the previous section which designates certain system states as fully inactive, all system states are possible throughout the experiment. As we are interested in learning an unknown but finite number of system states, we restrain the effect of the infinite state space through clever choice of priors. By using this prior to discourage overfitting, the iHMM is able to robustly learn the number of system states, not by learning the size of a generator matrix as in the Bernoulli process, but by accurately representing how many system states are visited in a system state trajectory.

To do this, we learn all elements of the infinite size transition probability vectors $\boldsymbol{\pi}_m$ (rows of propagators $\boldsymbol{\Pi}^{non}$ in Eq. 26) out of system state σ_m (*i.e.* the transition probability

matrix rows) from a common infinite probability vector $\boldsymbol{\beta}$, termed the *base distribution*, that is,

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

where the Dirichlet process is the infinite dimensional generalization of the Dirichlet distribution and α is a concentration parameter that affects how closely the transition probability vectors follow the base distribution. Here, the dominant elements of the base distribution determine the most visited system states.

Now, to learn this base distribution $\boldsymbol{\beta}$, we use the Griffiths–Engen–McCloskey process (denoted **GEM**) to avoid overfitting [36, 69–71], that is,

$$\boldsymbol{\beta} \sim \textbf{GEM}(\gamma).$$

This theoretical framework allows an infinite number of system states, but is constrained by the base distribution, which provides a robust way to model systems with an unknown number of system states.

In practice, we start by choosing a weak limit M_σ^{max}, that is, a large but finite number of system states that approximates an infinite system state space, as we did when using Bernoulli processes for continuous illumination. To learn the distributions over all the parameters through an MCMC algorithm, we choose our initial samples from the priors

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

where $\boldsymbol{\xi}$, a M_σ^{max} long vector with all elements equal to $1/M_\sigma^{max}$, reproduces the **GEM** process in finite dimensions. To allow for direct sampling of $\boldsymbol{\beta}$ and $\boldsymbol{\pi}_m$, we also learn the system state trajectory as follows:

$$s_n | s_{n-1} \sim \textbf{Categorical}(\boldsymbol{\pi}_{s_{n-1}}).$$

We can now generate samples for the transition probability vectors $\boldsymbol{\pi}_m$ of length M_σ^{max} via prior-likelihood conjugacy as

$$\boldsymbol{\pi}_m \sim \textbf{Dirichlet}(\mathbf{n}_m + \alpha\boldsymbol{\beta}), \qquad m = 1, 2, \ldots, M_\sigma^{max},$$

where the vector $\mathbf{n}_m$ collects the number of times each transition out of system state σ_m occurs.

Next, to sample the base distribution $\boldsymbol{\beta}$, we first sample an auxilary parameter $\mathbf{c}$ that effectively acts as a way to measure how often a system state is visited (in a sense, it is the analog of $\mathbf{n}_m$ for the transition probability vectors):

$$c_m = \sum_{i=1}^{M_\sigma^{max}} \sum_{j=0}^{(\mathbf{n}_m)_i - 1} \textbf{Bernoulli}\left(\frac{\alpha\beta_i}{j + \alpha\beta_i}\right), \tag{36}$$

where c_m, β_i, and $(\mathbf{n}_m)_i$ are elements of $\mathbf{c}$, $\boldsymbol{\beta}$, and $\mathbf{n}_m$, respectively. We can now generate new samples for $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta} \sim \textbf{Dirichlet}(\mathbf{c} + \gamma\boldsymbol{\xi}).$$

Other parameters are then sampled similarly to the parametric version of this sampler. We will show results obtained using this sampler in Sec. 4.2.

3.3 Likelihood Computation in Practice

As shown in the previous section, the likelihood typically takes the following generic form

$$L = \boldsymbol{\rho}_{start}\ \mathbf{Q}_1\ \mathbf{Q}_2\ \mathbf{Q}_3\ \ldots\ \ \mathbf{Q}_{K-1}\ \mathbf{Q}_K\ \boldsymbol{\rho}_{norm}^T, \tag{37}$$

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 the 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 in Eq. 37, we rewrite it as a product of multiple terms as follows:

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

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

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

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 (fixed number of system states) version of our sampler using synthetic data, while the two subsequent manuscripts [55, 68] focus on the nonparametric (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 continue to consider more realistic examples by incorporating: 1) crosstalk and detection efficiency; 2) background emission; 3) IRF; and then 4) end with a brief discussion on unknown number of system states. We demonstrate when these features become relevant, as well as the overall robustness and generality of the BNP-FRET package. Finally, we show that our BNP-FRET sampler can deduce the number of system states.

We assume continuous illumination for all the parametric examples and use the following priors for the analyses. The prior used for the FRET rates is

$$\lambda_{FRET}^{(i)} \sim \mathbf{Gamma}(1, 1\,\text{ns}),$$

and we use the following prior over the system transition rates

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

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

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

where θ_{ex}, θ_{FRET}, and θ_{sys} are placeholders for excitation, FRET, and system transition rates in nanosecond units. 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 sampler is quickly able to look for samples at different orders of magnitude and that about one-third of the samples are accepted in the Meteropolis-Hastings step. Intuitively, these variance values also represent the relative widths of the conditional posteriors for these parameters (in log-space). Furthermore, since a posterior width depends on the amount of data used, an increase or decrease in the number of photons used for analysis may require a corresponding decrease or increase in the covariances, respectively.

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 in 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. 2.7) 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 [61]

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

where the elements are specified in accordance with realistic cases. Using this generator matrix, we generated a superstate trajectory as described in Sec. 2.7. We analyzed 430000 photons from the generated data using our BNP-FRET package. The resulting posterior distribution for transitions between system states and FRET efficiencies (computed as $\varepsilon_{FRET}^{(i)} = \lambda_{FRET}^{(i)}/(\lambda_d + \lambda_{FRET}^{(i)})$ for the i-th system state) is shown in Fig. 4.1.1. However, limited data always leads to some error as evident from the slight offset of the peaks of the distribution away from the ground truth.

The effects of limited photon budget become significant especially when the system kinetics occur on varying timescales. Therefore, it would be useful to quantify these effects by defining a parameter that relates the number of photons analyzed by our BNP-FRET sampler and the resulting uncertainties over the estimated parameters. As such, we define a new dimensionless quantity called "photon budget index" that predicts the photon budget needed to estimate the transition rates in the model as

$$s = \frac{K\lambda_{ex}}{\lambda_{probe} M_\sigma}, \tag{38}$$

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 extensive experimentation, we have found a photon budget index of approximately 10^7 to be a safe lower threshold for keeping errors below 10% 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 $\lambda_{probe} = 2$ mss^{-1}, and $M_\sigma = 2$, which corresponds to a photon budget index of 1.08×10^7, that is, a safe number. 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 quantities such as quantiles.

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

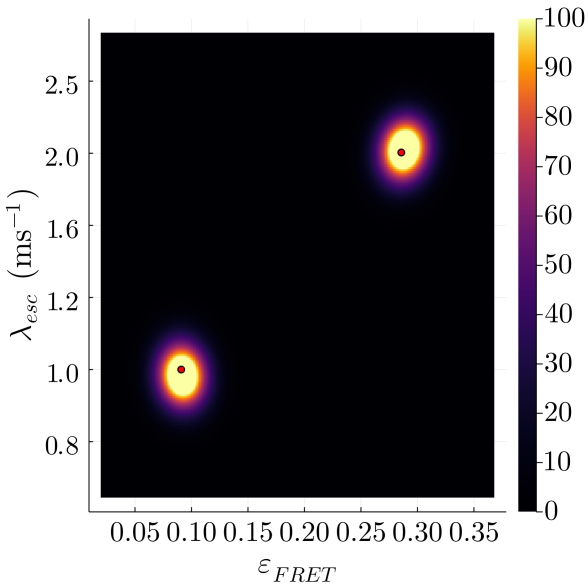

Figure 5: **Learned bivariate posterior for the system state escape rates λ_{esc} (log-scale) and FRET efficiencies ε_{FRET} given synthetic data.** To produce this plots, we analyzed synthetic data generated using an excitation rate of $\lambda_{ex} = 10$ ms^{-1}, and escape rates $\lambda_{esc} = 1$ & 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 small bias away from the ground truth is due to limited amount of data used. We have smoothed the distributions for illustrative purposes only using kernel density estimation (KDE) technique available through Julia Plots package.

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 dynamics 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 dynamics 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 learned posterior away from the ground truth increases as faster dynamics 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 complete deterioration of predictions, as the BNP-FRET sampler is unable to separate the two FRET efficiencies. The predicted transition rates are much closer to the excitation rate itself due to lack of temporal resolution available to probe such fast dynamics although FRET efficiencies are close to the average of ground truth values (≈ 0.185).

To conclude, the excitation rate used to collect smFRET data and the total number of photons in that data determine the amount of information available to resolve transitions among system states. The ability of the Bayesian methods to naturally propagate error from finiteness of information into parameter estimates make them indispensable tools for

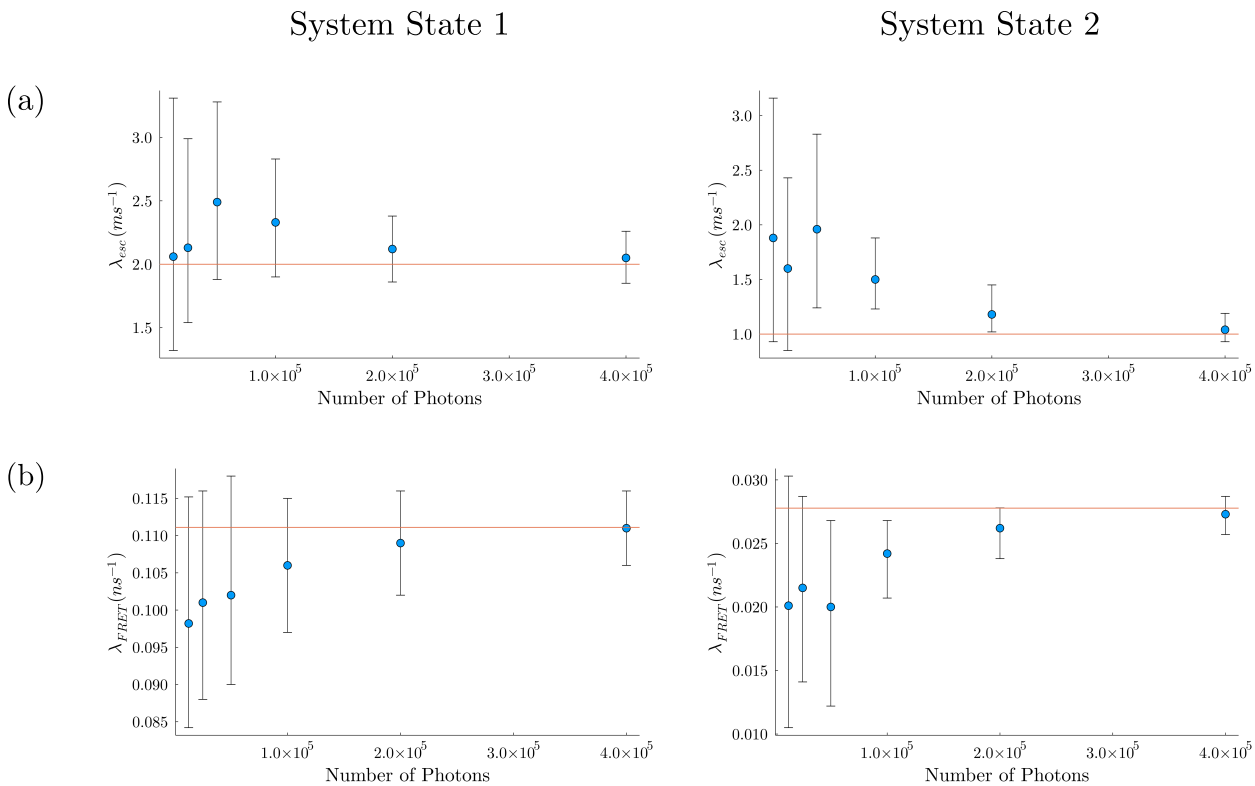

Figure 6: **System and FRET transition rates as functions of the number of photons used for analysis.** To produce these plots, we first generated synthetic data with an excitation rate of $\lambda_{ex} = 10$ ms^{-1}, and escape rates $\lambda_{esc} = 1$ and 2 ms^{-1} for the two system states, respectively. 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.

quantitative analysis of smFRET data. This is by contrast to maximum likelihood which only provides point estimates that are inaccurate 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 stochasti-

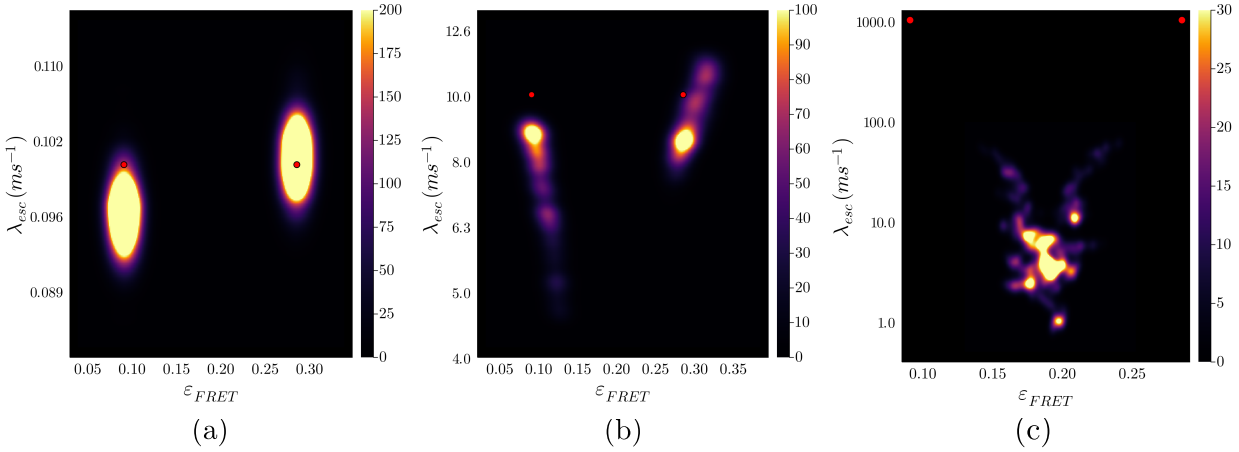

Figure 7: **Learned bivariate posterior for the system state escape rates λ_{esc} (log-scale) and FRET efficiencies ε_{FRET} from synthetic data**. For all the synthetic sm-FRET traces, we use an excitation rate of 10 ms^{-1} and FRET efficiences of 0.29 and 0.09 for the two system states, respectively. The three panels correspond to different timescales being probed with the transitions rates: (a) 0.1 ms^{-1}; (b) 10 ms^{-1}; and (c) 1000 ms^{-1}. These ground truth values are shown with red dots. The bias in the parameter estimates increases as faster and faster dynamics are probed, demonstrating deterioration of the BNP-FRET sampler's performance.

cally 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. As expected, incorporation of crosstalk into the likelihood as shown in Sec. 2.3.1 results in a smaller bias. The posterior distribution's mode for the corrected model migrates closer to the ground truth by 8 comparison to the uncorrected model.

4.1.3 An Example with Background Emissions

In section 2.5, we had shown a way to include the 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 previous examples but choose 10% of emissions to originate from background photons in each channel, that is, $\lambda_b^D = \lambda_b^A = \lambda_{ex}/10 = 1$ ms^{-1}. 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 9a. By comparison to the uncorrected method, our results migrate toward the ground truth when analyzed with the full method.

4.1.4 An Example with IRF

To demonstrate the effect of the detection instrument response function (IRF) as described in Sec. 2.3.3, we generated new synthetic data for a single fluorophore with an escape rate

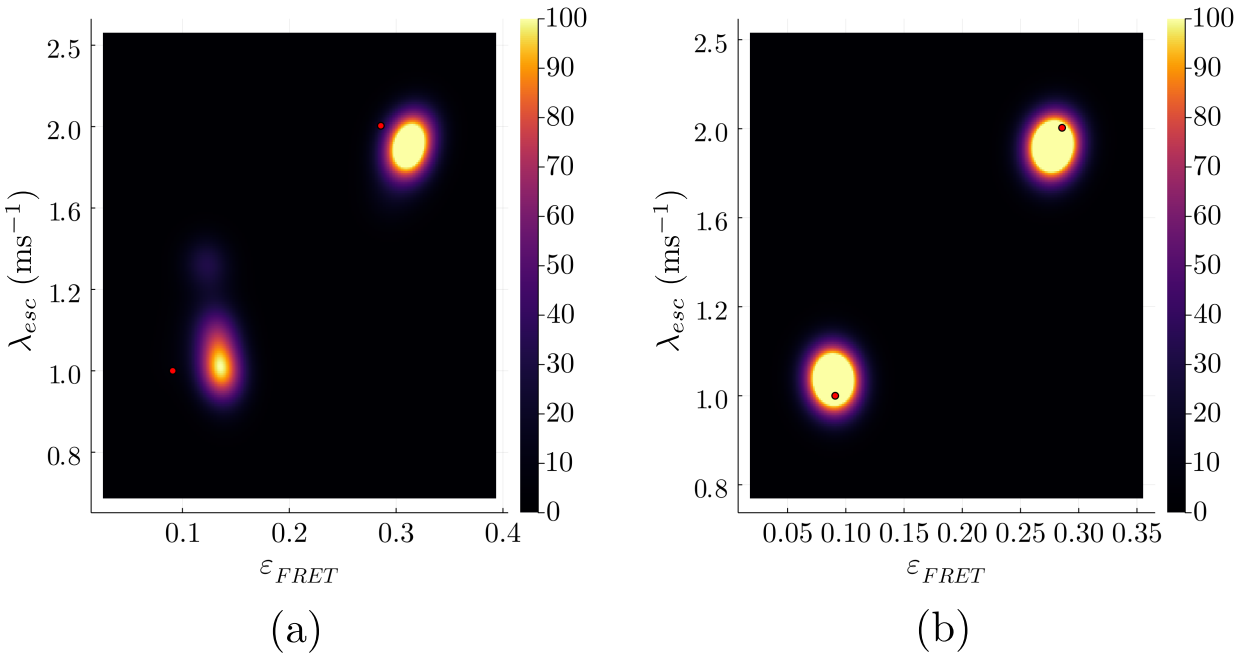

Figure 8: **Learned bivariate posterior for the system state escape rates λ_{esc} (log-scale) and FRET efficiencies ε_{FRET} for synthetic data with crosstalk**. The ground truth is shown with red dots. The learned posterior using the model that doesn't correct for cross talk (panel (a)) consistently shows bias away from the ground truth with higher FRET efficiency estimates on account of more donor photons being detected as acceptor photons. The corrected model mitigates this bias as demonstrated by the posterior in panel (b).

of $\lambda_d = 2.0$ ns^{-1} (similar to that of Cy3 dye [72]) being excited by a continuous-wave laser at a very high excitation rate $\lambda_{ex} = 0.01$ ns^{-1}. Here, we incorporate the detection IRF. 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 not incorporating the IRF. The results can be seen in Fig. 10, where the learned posterior under the model where the IRF is taken into account narrows. This is especially helpful when accurate lifetime determination is important in discriminating between different system states of molecules. By contrast, determining accurate lifetimes (which span ns timescales) do not impact the determination of much slower timescale dynamics of systems from one system state to the next.

4.2 A Nonparametric Example

Here, we demonstrate our method in learning number of system states by analyzing about 600 ms (about 120,000 photons) of synthetic smFRET time trace data with three system states (see Fig. 11) under pulsed illumination with 25 ns interpulse window. This example utilizes the iHMM method described earlier and discussed in greater depth in the third companion manuscript of the series. The excitation probability per pulse of 0.005, to be consistent with the data obtained in the third manuscript. Dynamics are set at 1.2 ms^{-1} es-

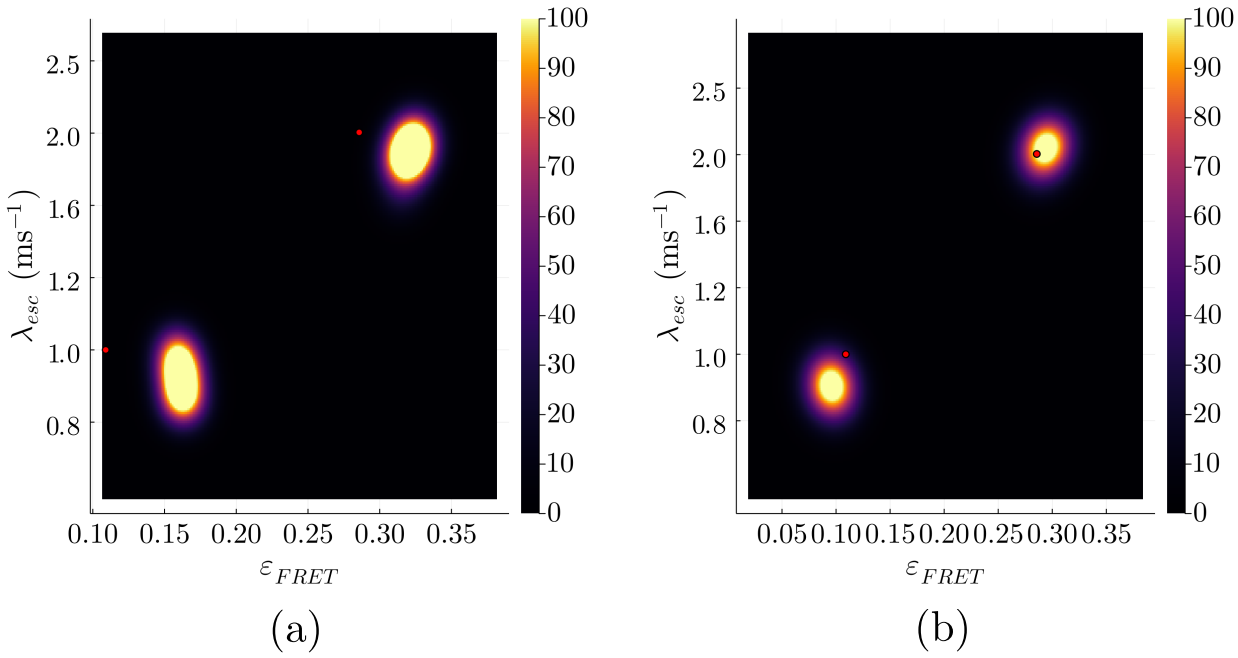

Figure 9: **Learned bivariate posterior for the system state escape rates λ_{esc} (log-scale) and FRET efficiencies ε_{FRET} given synthetic data with background emissions**. The ground truth is shown with red dots. The learned posterior distribution using the model that doesn't correct for background emissions (panel (a)) consistently shows bias away from the ground truth with higher estimates for FRET efficiencies on account of extra background photons. The corrected model mitigates this bias as demonstrated by the posterior in panel (b).

cape rates for the highest and lowest FRET system states, and an escape rate of $2.4\,\mathrm{ms}^{-1}$ for an intermediate system state between them. Our BNP method recovers these system states and their dynamics simultaneously alongside 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 states being combined together into one system state with a FRET efficiency of 0.63 and a lifetime of about $0.6\,\mathrm{ms}^{-1}$.

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 mod-

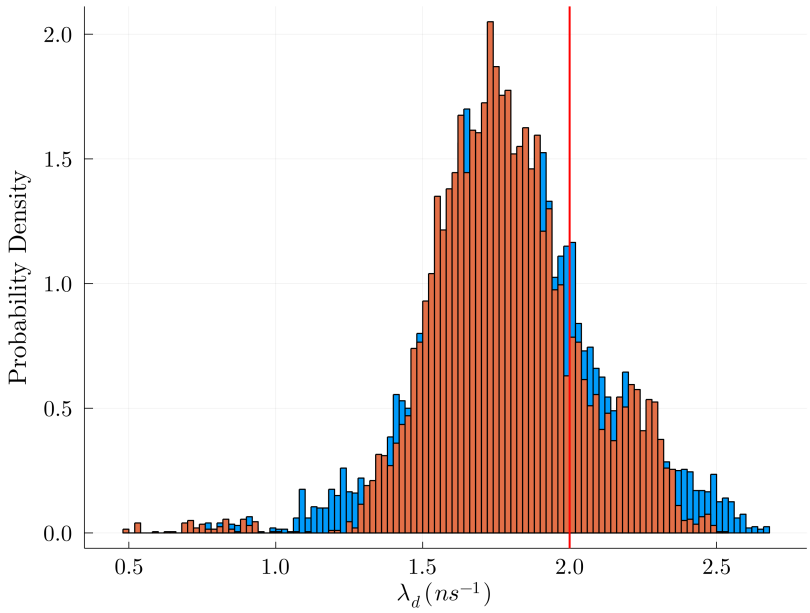

Figure 10: **Effects of IRF**s Both histograms show the fluorophore's inverse lifetime with ground truth shown by a red line. The bias in the peaks away from the ground truth arises from the limited amounts of data being used to learn the posterior shown. Corrected model (orange) reduces the histogram's breadth as compared with the uncorrected model (blue).

ification of generator matrices included in Eq. 17. This is useful in experiments employing pulsed illumination, [Sec. 2.4], 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 [54, 73], to reduce photobleaching [54, 73], or to study intermolecular interactions [12]. Similarly, the generator matrix in Eq. 17 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 [74], monitored conformational subpopulations of molecules [75], and improved our understanding of protein folding and interactions [2, 12, 76].

As the likelihood (Eq. 17) 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. Additions to our model which increase computational cost include: (1) IRF; (2) pulsed illumination; and (3) BNPs. The computational cost associated to IRF is attributed to the integral required (see Sec. 2.3.3). The cost likelihood computation in the pulsed illumination case depends linearly with pulse number, rather than with photon number (see Eq. 23). 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. 17) 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 just a calculation over a single data acquisition period. In our third

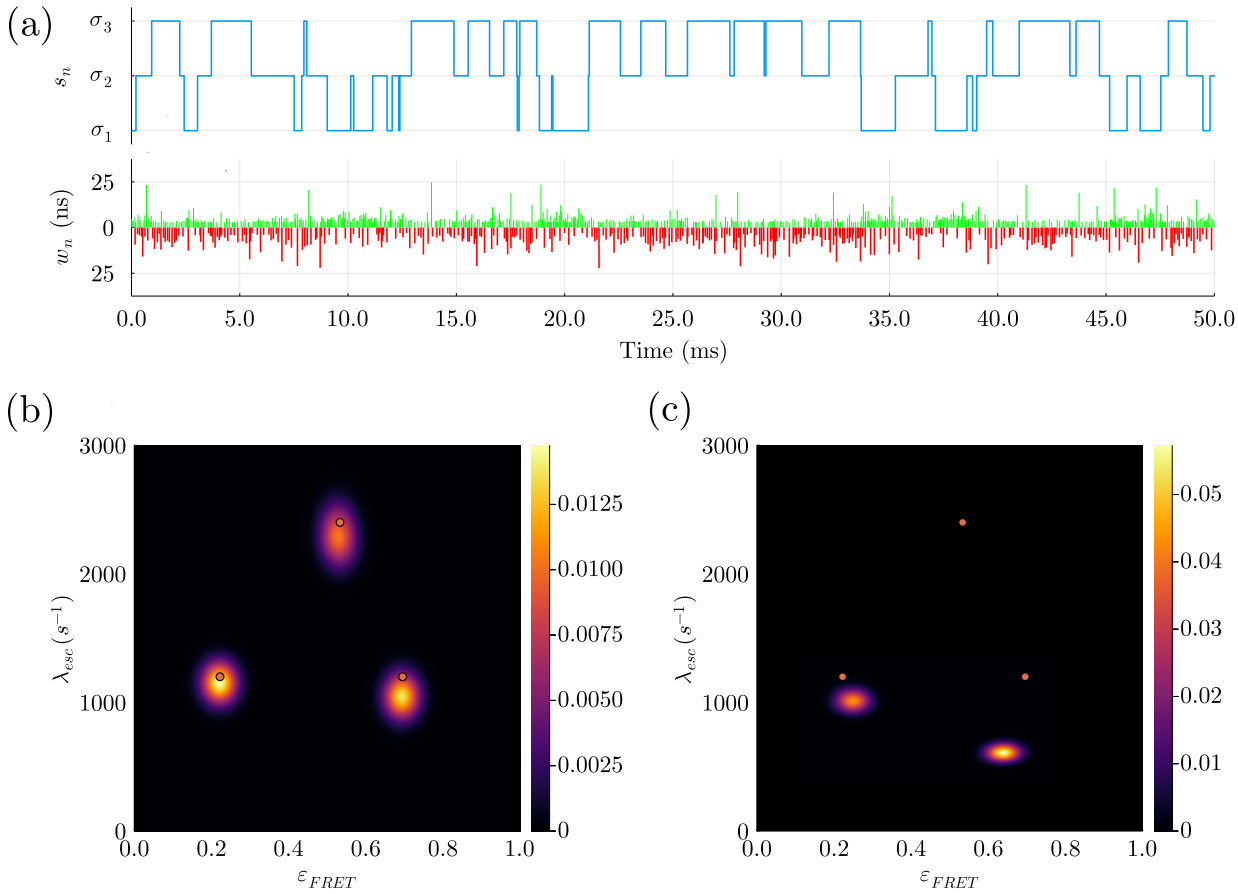

Figure 11: **Demonstration of nonparametric analysis on synthetic pulsed data**. In (a) we show simulated data for a pulsed illumination experiment. It illustrates a trajectory with three system states labelled s_n in blue and the corresponding photon arrivals with the vertical length of the photon marker 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).

companion paper [55], 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 formulation herein to a first order HMM, and allowing for faster computation of the likelihood in pulses where no photon is detected. In [55], 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 [13], protein-protein interactions [13, 54], and unhindered conformational dynamics of freely diffusing proteins [54, 77]. Photon-by-photon analysis of such data is often based on correlation methods which suffer from bulk averaging [35, 37, 38]. We believe our framework has the potential to extend Refs. [44, 78, 79] to learn both the system state dynamics 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 [13, 80]. Adopting an adaptation of [49] should allow us to generalize this framework and instead infer energy landscapes, perhaps relevant to protein folding [81, 82], continuum ratchets as applied to motor protein dynamics [83], and the stress modified potential landscapes of mechano-sensitive molecules [84].

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 papers [55, 68], 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 [68] under continuous illumination, and the dynamics of the Holliday junction under pulsed illumination [55].

6 Acknowledgments

We thank Weiqing Xu and Dr Zeliha Kilic for regular feedback and help, especially during the development of the nonparametrics samplers. S. P. acknowledges support from the NIH NIGMS (R01GM130745) for supporting early efforts in nonparametrics and NIH NIGMS (R01GM134426) for supporting single photon efforts.

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*, 49:5110–5139, 2020.

[2] Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule FRET. *Nature Methods*, 5(6):507–516, 2008.

[3] Alexander P. Demchenko. *Fluorescence Detection Techniques*, pages 69–132. Springer International Publishing, Cham, 2015.

[4] Ammasi. Periasamy and Richard N. Day. *Molecular imaging : FRET microscopy and spectroscopy*. The American Physiological Society methods in physiology series. Published for the American Physiological Society by Oxford University Press, Oxford, 2005.

[5] Boris Martinac. Single-molecule FRET studies of ion channels. *Progress in Biophysics and Molecular Biology*, 130:192–197, 2017.

[6] 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):981–984, 2012.

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

[8] 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):1188–1191, 2011. Publisher: American Chemical Society.

[9] 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–E3621, 2017.

[10] 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):16020–16023, 2015. Publisher: American Chemical Society.

[11] 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.

[12] 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.

[13] Benjamin Schuler and William A Eaton. Protein folding studied by single-molecule FRET. *Current Opinion in Structural Biology*, 18(1):16–26, 2008.

[14] 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 HP1[alpha]. *Nature Communications*, 9(1):235, 2018.

[15] 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. Publisher: American Association for the Advancement of Science.

[16] Th. Förster. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, 437(1-2):55–75, 1948.

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

[18] 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.

[19] Sean A. McKinney, Chirlmin Joo, and Taekjip Ha. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal*, 91(5):1941–1951, 2006.

[20] 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):3196–3205, 2009.

[21] 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):675–688, 2019.

[22] Ammasi Periasamy, Nirmal Mazumder, Yuansheng Sun, Kathryn G. Christopher, and Richard N. Day. *FRET Microscopy: Basics, Issues and Advantages of FLIM-FRET Imaging*, pages 249–276. Springer International Publishing, Cham, 2015.

[23] 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. *Opt. Express*, 24(9):9429–9445, 2016.

[24] Ivan Rasnik, Sean A McKinney, and Taekjip Ha. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Methods*, 3(11):891–893, 2006.

[25] 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):9619–9622, 2001.

[26] 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):161–193, 1979.

[27] Benjamin Schuler. Single-molecule fluorescence spectroscopy of protein folding. *ChemPhysChem*, 6(7):1206–1220, 2005.

[28] 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.

[29] Irina V. Gopich and Attila Szabo. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *The Journal of Physical Chemistry B*, 107(21):5058–5063, 2003.

[30] 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):7454–7464, 2003.

[31] 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):4534–4543, 2014.

[32] 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.

[33] Irina V. Gopich and Attila Szabo. Decoding the pattern of photon colors in single-molecule FRET. *The Journal of Physical Chemistry B*, 113(31):10965–10973, 2009.

[34] Zeliha Kilic, Ioannis Sgouralis, and Steve Pressé. Generalizing HMMs to continuous time for fast kinetics: Hidden Markov jump processes. *Biophysical Journal*, 120(3):409–423, 2021.

[35] 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.

[36] Ioannis Sgouralis and Steve Pressé. An introduction to infinite HMMs for single-molecule data analysis. *Biophysical Journal*, 112(10):2021–2029, 2017.

[37] Irina V. Gopich and Attila Szabo. Single-molecule FRET with diffusion and conformational dynamics. *The Journal of Physical Chemistry B*, 111(44):12925–12932, 2007.

[38] 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):13065–13075, 2016.

[39] Jake Lever, Martin Krzywinski, and Naomi Altman. Model selection and overfitting. *Nature Methods*, 13(9):703–704, 2016.

[40] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973. Publisher: Institute of Mathematical Statistics.

[41] Samuel J. Gershman and David M. Blei. A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1–12, 2012.

[42] 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.

[43] Meysam Tavakoli, J. Nicholas Taylor, Chun-Biu Li, Tamiki Komatsuzaki, and Steve Pressé. Single molecule data analysis: An introduction. *arXiv:1606.00403 [physics, q-bio]*, 2016.

[44] 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. Publisher: American Physical Society.

[45] 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.

[46] J. Shepard Bryan IV, Ioannis Sgouralis, and Steve Pressé. Diffraction-limited molecular cluster quantification with Bayesian nonparametrics. *Nature Computational Science*, 2(2):102–111, 2022. Number: 2 Publisher: Nature Publishing Group.

[47] 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):1015–1025, 2022. Publisher: American Chemical Society.

[48] Ayush Saurabh, Matthew Safar, and Steve Pressé. BNP-FRET: A software suite to analyze smFRET data using bayesian nonparametrics. *In preparation. Github link to appear.*

[49] J. Shepard Bryan IV, Prithviraj Basak, John Bechhoefer, and Steve Presse. Inferring potential landscapes from noisy trajectories. *arXiv:2202.02347 [physics]*, 2022. arXiv: 2202.02347.

[50] 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. Publisher: American Institute of Physics.

[51] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[52] 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):2702–2713, 1998.

[53] 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):6617–6628, 2021. PMID: 34125545.

[54] 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):523–533, 2005.

[55] 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. *In preparation.*

[56] Katsumasa Fujita, Minoru Kobayashi, Shogo Kawano, Masahito Yamanaka, and Satoshi Kawata. High-resolution confocal microscopy by saturated excitation of fluorescence. *Physical Review Letters*, 99:228105, 2007.

[57] Alex Dytso and H. Vincent Poor. Estimation in poisson noise: Properties of the conditional mean estimator. *IEEE Transactions on Information Theory*, 66(7):4304–4323, 2020.

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

[59] J. Brewer. Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, 25(9):772–781, 1978.

[60] 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–E118, 2015. Publisher: Proceedings of the National Academy of Sciences.

[61] 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.

[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):13480–13489, 2020.

[63] Wooseok Heo, Kazuto Hasegawa, Kenji Okamoto, Yasushi Sako, Kunihiko Ishii, and Tahei Tahara. Scanning two-dimensional fluorescence lifetime correlation spectroscopy: Conformational dynamics of DNA Holliday junction from microsecond to subsecond. *The Journal of Physical Chemistry Letters*, 13(5):1249–1257, 2022.

[64] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

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

[66] 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):110–120, 1997.

[67] Ioannis Sgouralis and Steve Pressé. An introduction to infinite HMMs for single-molecule data analysis. *Biophysical Journal*, 112(10):2021–2029, 2017.

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

[69] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[70] Jim Pitman. Poisson–dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. *Combinatorics, Probability and Computing*, 11(5):501–514, 2002.

[71] Jayaram Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.

[72] 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):11064–11074, 2007. PMID: 17718469.

[73] Johannes Hohlbein, Timothy D. Craggs, and Thorben Cordes. Alternating-laser excitation: single-molecule FRET and beyond. *Chemical Society Reviews*, 43:1156–1171, 2014.

[74] 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):1274–1283, 2010.

[75] Jean-Pierre Clamme and Ashok A. Deniz. Three-color single-molecule fluorescence resonance energy transfer. *ChemPhysChem*, 6(1):74–77, 2005.

[76] Sungchul Hohng, Chirlmin Joo, and Taekjip Ha. Single-molecule three-color FRET. *Biophysical Journal*, 87(2):1328–1337, 2004.

[77] 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):3670–3675, 1999.

[78] 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. Publisher: American Institute of Physics.

[79] 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. Number: 1 Publisher: Nature Publishing Group.

[80] Benjamin Schuler and Hagen Hofmann. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Current Opinion in Structural Biology*, 23(1):36–47, 2013.

[81] 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.

[82] Benjamin A. Shoemaker, Jin Wang, and Peter G. Wolynes. Structural correlations in protein folding funnels. *Proceedings of the National Academy of Sciences*, 94(3):777–782, 1997.

[83] Anatoly B. Kolomeisky and Michael E. Fisher. Molecular motors: A theorist's perspective. *Annual Review of Physical Chemistry*, 58(1):675–695, 2007.

[84] 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.