Effect of Pixelation on the Parameter Estimation of Single Molecule Trajectories

Milad R. Vahid , Bernard Hanzon, and Raimund J. Ober, *Senior Member, IEEE*

Abstract—The advent of single molecule microscopy has revolutionized biological investigations by providing a powerful tool for the study of intercellular and intracellular trafficking processes of protein molecules which was not available before through conventional microscopy. In practice, pixelated detectors are used to acquire the images of fluorescently labeled objects moving in cellular environments. Then, the acquired fluorescence microscopy images contain the numbers of the photons detected in each pixel, during an exposure time interval. Moreover, instead of having the exact locations of detection of the photons, we only know the pixel areas in which the photons impact the detector. These challenges make the analysis of single molecule trajectories, from pixelated images, a complex problem. Here, we investigate the effect of pixelation on the parameter estimation of single molecule trajectories. In particular, we develop a stochastic framework to calculate the maximum likelihood estimates of the parameters of a stochastic differential equation that describes the motion of the molecule in living cells. We also calculate the Fisher information matrix for this parameter estimation problem. The analytical results are complicated through the fact that the observation process in a microscope prohibits the use of standard Kalman filter type approaches. The analytical framework presented here is illustrated with examples of low photon count scenarios for which we rely on Monte Carlo methods to compute the associated probability distributions.

Index Terms—Cramér-Rao lower bound, Fisher information matrix, maximum likelihood estimation, Monte Carlo, pixelated detectors, single molecule tracking, stochastic differential equations.

I. INTRODUCTION

THE STUDY of intercellular and intracellular trafficking processes of objects of interest has been the subject of many research projects during the past few decades. The advent of single molecule microscopy made it possible to observe and track single molecules in living cells, which were not achievable before using conventional microscopes [1]–[7].

Milad R. Vahid was with the Department of Biomedical Engineering, Texas A&M University, College Station, TX 77843 USA. He is now with the Department of Biomedical Data Science, Stanford University, Stanford, CA 94305 USA (e-mail: milad.mrv@gmail.com).

Bernard Hanzon is with the Department of Mathematics, University College Cork, T12YX86 Cork, Ireland (e-mail: b.hanzon@ucc.ie).

Raimund J. Ober is with the Centre for Cancer Immunology, Faculty of Medicine, University of Southampton, Southampton SO16 6YD, U.K. (e-mail: r.ober@soton.ac.uk).

In fluorescence microscopy, the photons emitted by a fluorescently labeled object located in the object space are detected by a planar detector in the image space. In the fundamental data model, we assume that the time points and locations of the photons emitted by the object are detected by an ideal unpixelated detector. However, in practice, pixelated detectors, such as charge-coupled device (CCD) and electron multiplying CCD (EMCCD) cameras, are commonly used for acquiring the image of the object. In this case, referred to as the practical data model, the measurements, i.e., the fluorescence microscopy images, consist of the numbers of the photons detected in each pixel. Moreover, instead of having the exact locations of detection of the photons, we only know the pixel areas in which the photons impact the detector. These challenges make the analysis of single molecule trajectories from pixelated images a complex problem.

In the literature, there are several methods available concerning the problem of the parameter estimation of single molecule trajectories in cellular contexts. The majority of these methods model the effect of pixelation by using an additive noise in the fundamental data model. However, in general, this approximation does not describe the underlying stochastic model precisely. For example, in [8]–[10], by encapsulating the effect of pixelation in a Gaussian additive random variable, referred to as the localization uncertainty, Berglund and Michalet have proposed methods for the estimation of diffusion coefficients based on mean square displacement of the observed locations of the molecule. For a similar observation model, Relich *et al.* [11] have proposed a method for the maximum likelihood estimation of the diffusion coefficient, with an information-based confidence interval, from Gaussian measurements. Although using these approximate observation models makes all corresponding computations simpler, it does not model the effect of the pixelated camera accurately. Calderon has extended Berglund's motion blur model to handle confined dynamics [12]–[14]. His proposed approach enables the estimation of the parameters of the motion model of the molecule by considering confinement and motion blur within a time domain maximum likelihood estimation framework. In [15], for the single molecule trajectory parameter estimation problem, a more accurate model has been used to describe the image of pixelated detectors. In this model, the expected intensity measured in each pixel is obtained by integrating the image profile, which is expressed in terms of a scaled and shifted version of the point spread function, over the pixel area. Here, we use a similar approach to model pixelated data more accurately.

In [16], we developed a stochastic framework in which we calculate the maximum likelihood estimates of the parameters of the model that describes the motion of the molecule in cellular environments. More importantly, we proposed a general framework to calculate the Cramér-Rao lower bound (CRLB), given by the inverse of the Fisher information matrix, for the estimation of unknown parameters and use it as a benchmark in the evaluation of the standard deviation of the estimates. In [16], we focused on the fundamental microscopy data model, in which the image of a molecule is acquired by an unpixelated detector.

In this paper, we propose a general framework to investigate the effect of pixelation of the detector on the parameter estimation of single molecule trajectories accurately. We extend our previous results obtained for deterministic trajectories [17] and show examples of low photon count scenarios. We also include an example analysis in which we show how the standard deviation of parameter estimates depends on the pixel size of the detector. We consider complex relationships between the single molecule motion model, the photon emission process and the underlying statistical model of photon detection in the pixels of the detector in deriving all analytical expressions, e.g., likelihood function and Fisher information matrix, and no approximations are made.

This paper is organized as follows. In Section II, we introduce important notations used to define different data models in the paper. In Section III, we define fundamental and practical data models and present mathematical descriptions of them. In Section IV, we introduce continuous-time stochastic differential equations, which are used to model the motion of single molecules in cellular environments, and calculate their solutions at discrete time points. Section V is devoted to calculation of the maximum likelihood estimates of the parameters of the system based on the introduced motion and data models in the previous sections. In Section VI, we calculate the general expressions for the Fisher information matrix for both of the fundamental and practical data model. In Section VII, we investigate the effect of noise on the expressions derived in the previous sections. Finally, a summary of the paper and conclusions are provided in Section VIII.

II. NOTATIONS

In this section, we introduce the following notations that will be used throughout the paper.

Let $\mathcal{C}^p$ be a pixelated detector defined as the union of a collection $\{C_1, \ldots, C_K\}$ of connected open and disjoint subsets of a region within $\mathbb{R}^2$ corresponding to the photon detection area of the detector. Let $L \geq 1$. We use the notation $A_K^L = \{1, \ldots, K\}^L$ to denote the set of vectors of all possible pixel labels of L photons detected in the pixelated image (see Section III-B for more detailed definitions). We also denote the Cartesian product of the pixel sets $C_{v_1}, \ldots, C_{v_L}$, where $v_1, v_2, \ldots, v_L \in \{1, 2, \ldots, K\}$ by $C_{v_{1:L}} = C_{v_1} \times C_{v_2} \times \ldots C_{v_L}$.

In this paper, for random vectors X and Y, the conditional probability density function of X, given Y, is denoted by $p_{X|Y}$. For example, let $X(\tau_1), \ldots, X(\tau_L) \in \mathbb{R}^3$ denote the locations of the molecule at a sequence of time points $(\tau_1, \ldots, \tau_L) \in$

$\Delta^L := \{(\tau_1, \ldots, \tau_L) : t_0 \leq \tau_1 < \tau_2 < \ldots < \tau_L \leq t\}$. Then, denoting the locations of the detected photons on the detector plane by $R_1 := U(X(T_1)), R_2 := U(X(T_2)), \ldots, R_L := U(X(T_L)) \in \mathbb{R}^2$, where U is a random function that maps the object space into the image space, the conditional probability density function of R_l, given $R_{l-1}, l = 2, \ldots, L$, where we assume $L \geq 2$, is denoted by $p_{R_l|R_{l-1}}$. Note that in this paper, we use uppercase letters, e.g., T_i, to denote random variables, and lowercase letters, e.g., τ_i, to denote particular values that the random variables can assume. We also denote the Poisson-distributed probability, with nonnegative intensity function $\Lambda(\tau), \tau \geq t_0$, of detecting $L \geq 0$ photons in the time interval $[t_0, t]$ by $p_L := \frac{e^{-\int_{t_0}^{t} \Lambda(\psi)d\psi}(\int_{t_0}^{t} \Lambda(\psi)d\psi)^L}{L!}$ (see Section III-A for more details).

Given observed data $r \in \mathbb{R}^2$ with probability distribution p_R^θ, where $\theta \in \mathbb{R}^n$ denotes the row vector of parameters, the Fisher information matrix $I(\theta)$ is given by

$$I(\theta) = E\left\{\left(\frac{\partial}{\partial \theta} \log p_R^\theta(r)\right)^T \left(\frac{\partial}{\partial \theta} \log p_R^\theta(r)\right)\right\}. \quad (1)$$

See Section VI for the detailed description of Fisher information matrix.

Also, in order to simplify vector representations of complex formulas in this paper, for arbitrary n-dimensional vectors $x_1, \ldots, x_L \in \mathbb{R}^n$ (or n-dimensional random variables $X_1, \ldots, X_L$), we define the $n \times L$ vector $x_{1:L} := (x_1, \ldots, x_L)$ (or $X_{1:L} := (X_1, \ldots, X_L)$). The integral of $f : \mathbb{R}^{n \times L} \mapsto \mathbb{R}$ over $\mathbb{R}^{n \times L}$ is denoted by $\int_{\mathbb{R}^{n \times L}} f(x_{1:L})dx_{1:L}$.

III. DATA MODEL

In a standard optical microscope, the image of an object, which is in general moving in the object space, is captured by a detector in the image space. In order to be able to model the photon detection process in a pixelated detector we firstly require a model of the precise impact locations on the detector of photons emitted by the imaged object. Such a model, the *fundamental data model*, has been developed in [16], [18], [19]. In this fundamental data model, we consider ideal conditions for the data acquisition procedure, in which it is assumed that we have an unpixelated image detector. We briefly summarize this model here before proceeding to the development of the *practical data model*. For this we use the fundamental data model to obtain a probabilistic description for the number of photons detected in each pixel of a pixelated detector.

A. Fundamental Data Model

In the fundamental data model, the acquired data are the time points and locations of detection of the photons emitted from the object, where we have an unpixelated image detector. These time points and locations are intrinsically random. In general, the time points of detection of the emitted photons can be modeled as a counting process. The locations of detection of the photons emitted by the object are described by a random function that maps the object space into the image space. In this paper, for

the fundamental data model, we assume that the locations of the photons, with the correct chronological ordering, emitted by the object can be detected by the detector.

We introduce the following notation. For $t_0 \in \mathbb{R}$, let the random process $X(\tau), \tau \geq t_0$, describe the location of an object of interest, which emits photons, in the object space at time τ. Let $\mathcal{C} := \mathbb{R}^2$ denote a non-pixelated detector. Let $\{N(\tau), \tau \geq t_0\}$ be a Poisson process with non-negative and piecewise continuous intensity function $\Lambda(\tau), \tau \geq t_0$, that describes the time points of detection of the photons emitted by the object that impact the detector $\mathcal{C}$. These ordered time points, which are the events (jump points) of $\{N(\tau), \tau \geq t_0\}$, are denoted by one-dimensional (1D) random variables $t_0 \leq T_1 < T_2 < \cdots$. The location of detection of the photon emitted by the object, at time $\tau \geq t_0$, that impacts the detector $\mathcal{C}$ is described by $U(X(\tau))$, where U is a random function that maps the object space into the image space. For $x \in \mathbb{R}^3$, let f_x denote the probability density function of $U(x)$, referred to as the *image profile* of an object located at $x \in \mathbb{R}^3$ in the object space. In many practical scenarios, the image profile can be described as a scaled and shifted version of a function, referred to as the *image function*, that describes the image of an object on the detector plane at unit lateral magnification. Assume that there exists a function $q_{z_0}: \mathbb{R}^2 \mapsto \mathbb{R}, z_0 \in \mathbb{R}$, such that for an invertible matrix $M \in \mathbb{R}^{2 \times 2}$ and $x := (x_0, y_0, z_0) \in \mathbb{R}^3$,

$$f_x(r) := \frac{1}{|\det(M)|} q_{z_0} \left(M^{-1} r - (x_0, y_0)^T \right), \quad r \in \mathcal{C}. \quad (2)$$

In particular, when the object is a point source and is in-focus with respect to the detector, according to optical diffraction theory, its image can be modeled by the continuous function q, called an Airy profile given by

$$q(x_0, y_0) = \frac{J_1^2 \left(\frac{2\pi n_a}{\lambda} \sqrt{x_0^2 + y_0^2} \right)}{\pi \left(x_0^2 + y_0^2 \right)}, \quad (x_0, y_0) \in \mathbb{R}^2, \quad (3)$$

where n_a denotes the numerical aperture of the objective lens, λ denotes the emission wavelength of the molecule, and J_1 denotes the first order Bessel function of the first kind. In some applications, it is computationally more convenient to approximate the Airy profile by a Gaussian distribution given by

$$q(x_0, y_0) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \left(\frac{x_0^2 + y_0^2}{\sigma^2} \right)}, \quad (x_0, y_0) \in \mathbb{R}^2, \quad (4)$$

where $\sigma > 0$.

For an out-of-focus point source, the image function can be modeled by the classical Born and Wolf model given by, for $(x_0, y_0) \in \mathbb{R}^2$, [20]

$$
\begin{aligned}
&q_{z_0}(x_0, y_0) \\
&= \frac{4\pi n_a^2}{\lambda^2} \left| \int_0^1 J_0 \left(\frac{2\pi n_a}{\lambda} \rho \sqrt{x_0^2 + y_0^2} \right) e^{\frac{j\pi n_a^2 z_0}{n_o \lambda} \rho^2} \rho d\rho \right|^2,
\end{aligned}
$$
$$(5)$$

where J_0 is the zeroth-order Bessel function of the first kind, n_o is the refractive index of the objective lens immersion medium, and $z_0 \in \mathbb{R}$ is the z-location of the point source on the optical axis in the object space.

The above mentioned image functions are only examples of the myriad of image functions that have been proposed and are used to describe image formation in a microscope and are themselves in fact approximations [18]. The subsequent developments are therefore carried out with a high level of generality to allow for the use of the most appropriate image function in a circumstance.

In [16], we use the fundamental data model for the maximum likelihood estimation of biophysical parameters such as diffusion and drift coefficients, from images acquired by an unpixelated detector. The joint probability density of the acquired data points, i.e. the impact locations of the detected photons on the detector, is the key probabilistic concept for our later developments. Specifically, in Theorem 3.1, we calculate the conditional probability density function $p_{R_{1:L}|N(t)}$ of $R_1 = U(X(T_1)), \ldots, R_L = U(X(T_L))$, given $N(t)$ for the fundamental data model. In the next section, we will use these results to characterize the acquired data from pixelated detectors.

Theorem 3.1: Let $L \geq 1$. The conditional probability density function $p_{R_{1:L}|N(t)}$ of $R_1, \ldots, R_L$, given $N(t) = L$, can be calculated as

$$p_{R_{1:L}|N(t)} (r_{1:L}|L)$$
$$= \int_{\mathbb{R}^{3 \times L}} \left(\int_{\Delta^L} F(r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L}, \quad (6)$$

where

$$F(r_{1:L}, x_{1:L}, \tau_{1:L}) := \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi) d\psi \right)^L}$$
$$\times \left(\prod_{i=1}^L f_{x_i}(r_i) \Lambda(\tau_i) \right) p_{X(\tau_1), \ldots, X(\tau_L)}(x_{1:L}), \quad (7)$$

is the conditional probability density function over the set of all (r, x, τ), given $N(t) = L$, and $p_{X(\tau_1), \ldots, X(\tau_L)}$ is the joint probability density function of $X(\tau_1), \ldots, X(\tau_L)$.

If $L \geq 2$ and $\{X(\tau_1), \ldots, X(\tau_L)\}$ is a Markov sequence, then,

$$p_{X(\tau_1), \ldots, X(\tau_L)} (x_{1:L})$$
$$= p_{X(\tau_1)} (x_1) \prod_{l=2}^L p_{X(\tau_l)|X(\tau_{l-1})} (x_l|x_{l-1}),$$

where $p_{X(\tau_l)|X(\tau_{l-1})}, l = 2, \ldots, L$, is the conditional probability density function of $X(\tau_l)$, given $X(\tau_{l-1})$, and $p_{X(\tau_1)}$ is the probability density function of $X(\tau_1)$.

Proof: See Appendix IX-B. ∎

B. Practical Data Model

In practice, pixelated detectors, e.g. CCD and EMCCD cameras, are commonly used for acquiring images of fluorescently labeled objects. In this subsection, we describe the practical data model.

In the practical data model, the data acquired by a pixelated detector are the number of detected photons at each pixel (Fig. 1). Let the pixelated detector $\mathcal{C}^p$ be defined as the union of a
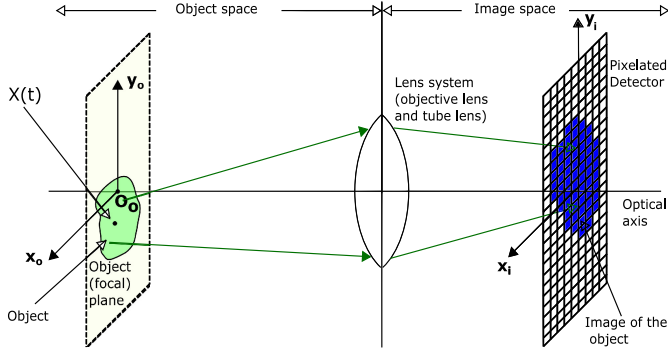
Fig. 1. Schematic of an optical microscope. An object located in the object (focal) plane is imaged by an optical lens system and the image of the object is acquired by a pixelated detector in the image space. A random variable $X(t), t \geq t_0$, describes the location of the object in the object plane at time t.

collection $\{C_1, \ldots, C_K\}$ of connected open and disjoint subsets of a region within $\mathbb{R}^2$ corresponding to the photon detection area of the detector, such that $\bigcup_{k=1}^{K} C_k = \mathcal{C}^p$. We use the random variable $S_k, k = 1, \ldots, K$, to describe the number of photons in the pixel C_k that result from the detection of photons from the object of interest. Let $L = \sum_{k=1}^{K} S_k$ denote the total number of photons that impact the detector in a given experiment. We now need a notation that allows us to translate the information of the pixel labels of the individual photons to the number of photons in each pixel (i.e. the S_k values). Note that $L = 0$ if and only if $S_k = 0$ for all $k = 1, 2, \ldots, K$, which is a trivial case. Now assume $L \geq 1$.

Let $A_K^L = \{1, \ldots, K\}^L$ denote the set of vectors of all possible pixel labels of L photons. This means that each L element vector in A_K^L contains the pixel numbers in which each of the detected photons is captured. For the simple example where we have $K = 2$ pixels and $L = 3$ photons, for each vector in the set A_2^3, the first component denotes the pixel in which the first detected photon is captured, etc. The vector $v = (2, 1, 2) \in A_2^3$ implies that the first photon is captured by pixel 2, the second photon by pixel 1 and the third photon again by pixel 2. The set of all such vectors for this example is given by

$$A_2^3 : = \{(1,1,1),(1,1,2),(1,2,1),(1,2,2),(2,1,1),$$
$$(2,1,2),(2,2,1),(2,2,2)\}.$$

Note that for a vector $v = (v_1, v_2, \ldots, v_L)$,

$$Pr\left[\bigcap_{l=1}^{L} (R_l \in C_{v_l}) \,|\, N(t) = L\right]$$
$$= \int_{C_{v_{1:L}}} p_{R_{1:L}|N(t)} (r_{1:L}|L)\, dr_{1:L}, \qquad (8)$$

or for our small example, for $v = (v_1, v_2, v_3) = (2, 1, 2)$,

$$Pr\left[(R_1 \in C_2) \cap (R_2 \in C_1) \cap (R_3 \in C_2) \,|\, N(t) = 3\right]$$
$$= \int_{C_2}\int_{C_1}\int_{C_2} p_{R_1,R_2,R_3|N(t)} (r_1, r_2, r_3|3)\, dr_3 dr_2 dr_1. \qquad (9)$$

In order to be able to compute the probability $Pr[S_1 = z_1, S_2 = z_2, \ldots, S_K = z_K]$, we need to know all the events that lead to the photon count $S_1 = z_1, S_2 = z_2, \ldots, S_K = z_K$, or we need to know the elements of A_K^L that can lead to this photon count. For our example, as we have $K = 2$ pixels we are interested to determine $Pr[S_1 = z_1, S_2 = z_2]$. If as above we know that $L = z_1 + z_2 = 3$, i.e. that sum of the pixel counts is 3, we need to determine the elements of A_2^3 that are consistent with this photon count.

We therefore, for a vector $v \in A_K^L$, denote by $\|v\|_{=k}, k = 1, \ldots, K$, the number of the entries of v which are equal to k. So as v denotes a vector of pixel labels of L photons, then $\|v\|_{=k}, k = 1, \ldots, K$ denotes the number of photons that have ended up in pixel k. For example, for $v = (1, 1, 2) \in A_2^3$, we have $\|v\|_{=1} = 2$, $\|v\|_{=2} = 1$.

This leads us to the following notation, which, for given photon numbers $z_1, \ldots, z_K \in \{0, 1, \cdots\}$, and $\sum_{k=1}^{K} z_k = L$, lists the elements in A_K^L that will produce these photon counts in the pixels $C_1, C_2, \ldots, C_K$, i.e.

$$A_K^L (z_1, \ldots, z_K) := \left\{ v \in A_K^L |\, \|v\|_{=k} = z_k, k = 1, \ldots, K \right\}. \qquad (10)$$

So this set contains all vectors of pixel labels of L photons that correspond to a configuration in which z_k photons have landed in pixel C_k, for $k = 1, 2, \ldots, K$.

Continuing our example we want to determine $A_2^3(1,2)$, i.e. the elements in A_2^3 that are such that there is one photon detected in pixel C_1 and 2 photons in pixel C_2. Examining the set A_2^3 we see that $(1,2,2)$ is an element of $A_2^3(1,2)$ as one photon, the first detected, is captured in pixel 1 and two photons, the second and third detected photon, are captured in pixel 2. Proceeding in this manner we obtain,

$$A_2^3(1,2) := \{(1,2,2),(2,1,2),(2,2,1)\},$$
$$A_2^3(2,1) := \{(1,1,2),(1,2,1),(2,1,1)\},$$
$$A_2^3(0,3) := \{(2,2,2)\},$$
$$A_2^3(3,0) := \{(1,1,1)\}.$$

With this notation we can now immediately see what one needs to do to, for example, compute the probability $Pr[S_1 = 2, S_2 = 1]$. The set of events that lead to $S_1 = 2$ and $S_2 = 1$ is given by $A_2^3(2,1)$, i.e. $\{(1,1,2),(1,2,1),(2,1,1)\}$. Therefore,

$$Pr[S_1 = 2, S_2 = 1]$$
$$= Pr\left[((R_1 \in C_1) \cap (R_2 \in C_1) \cap (R_3 \in C_2))\right.$$
$$\cup ((R_1 \in C_1) \cap (R_2 \in C_2) \cap (R_3 \in C_1))$$
$$\cup ((R_1 \in C_2) \cap (R_2 \in C_1) \cap (R_3 \in C_1)) \,|\, N(t)$$
$$= L]\, Pr[N(t) = L]$$
$$= Pr\left[\bigcup_{v_{1:3} \in A_2^3(2,1)} \left(\bigcap_{l=1}^{3} (R_l \in C_{v_l})\right) \,\middle|\, N(t) = 3\right] Pr[N(t) = 3]$$

$$= \sum_{v_{1:3} \in A_2^3(2,1)} Pr\left[\left(\bigcap_{l=1}^3 (R_l \in C_{v_l})\right) | N(t) = 3\right] Pr\left[N(t) = 3\right].$$

The expressions on the right hand side can be computed using the expression above (Eq. (8)).

In the following theorem we summarize the above derivations and state the main result of this section. It provides the desired expression for the discrete multivariable probability distribution $Pr[S_1 = z_1, S_2 = z_2, \ldots, S_K = z_K]$ that $z_1, z_2, \ldots, z_K$ photons are detected in pixels $C_1, C_2, \ldots, C_K$. Assuming that we have acquired L photons in total on the detector, this expression is given by Eq. (11) below. It shows how we can compute the probability that the detected photons during one trajectory impact the various pixels. In this theorem we need to make a technical distinction between two different versions of the detector. If $\bar{C}^p = \mathbb{R}^2$, the model is referred to as the *infinite practical data model*. We will see later that the infinite practical model will lead to somewhat simplified expressions, for example of the likelihood function. Obviously, in reality no practical detector will be infinite. But this model can serve as a useful approximation in the cases where we can assume that all photons that impact the infinite detector plane that is spanned by the detector, are in fact captured by the detector itself.

Theorem 3.2: Let t_0 and $t, t_0 < t$, be given. 1. In the infinite practical data model, we have, for $z_1, \ldots, z_K = 0, 1, \cdots,$ and $L = \sum_{k=1}^K z_k$: $Pr(L = 0) = p_0$, and for $L \geq 1$:

$$Pr\left[S_1 = z_1, \ldots, S_K = z_K\right] = p_L \sum_{v_{1:L} \in A_K^L(z_{1:K})} \int_{C_{v_{1:L}}} \left[\int_{\mathbb{R}^{3 \times L}}\right.$$

$$\left.\left(\int_{\Delta^L} F\left(r_{1:L}, x_{1:L}, \tau_{1:L}\right) d\tau_{1:L}\right) dx_{1:L}\right] dr_{1:L}, \quad (11)$$

where $F(.)$ is given by Eq. (7).

2. In the practical data model, we have

$$Pr[L = 0] = p_0 + \sum_{z=1}^\infty p_z \sum_{v_{1:z} \in A_{K+1}^z(z_{1:K}, z)}$$

$$\int_{C_{v_{1:z}}} \left[\int_{\mathbb{R}^{3 \times z}} \left(\int_{\Delta^z} F\left(r_{1:z}, x_{1:z}, \tau_{1:z}\right) d\tau_{1:z}\right) dx_{1:z}\right] dr_{1:z},$$

and

$$Pr\left[S_1 = z_1, \cdots, S_K = z_K\right] = \sum_{z=0}^\infty p_{L+z} \sum_{v_{1:L+z} \in A_{K+1}^{L+z}(z_{1:K}, z)}$$

$$\int_{C_{v_{1:L+z}}} \left[\int_{\mathbb{R}^{3 \times (L+z)}} \left(\int_{\Delta^{L+z}} F(r_{1:L+z}, x_{1:L+z}, \tau_{1:L+z}) d\tau_{1:L+z}\right)\right.$$

$$\left.\times dx_{1:L+z}\right] dr_{1:L+z}, \quad L \geq 1. \quad (12)$$

Here C_{K+1} denotes the complement (in the detector plane) of the closure of the union of the pixel sets that form the detector, so it accounts for all the photons that have gone through the detector plane, but have missed the detector. (Note that C_{K+1} will be open, but we do not claim C_{K+1} is connected).

Proof: See Appendix IX-C. ∎

The result states that in order to compute the discrete probability distribution for a set of photon counts we need to carry out a summation of integrals. The number of summands is given by the size $|A_K^L(z_1, \ldots, z_K)|$ of the set $A_K^L(z_1, \ldots, z_K)$ which is equal to $\frac{L!}{z_1! \cdots z_K!}$.

In the above theorem, we account for all photons that cross the detection plane, and therefore the time points of detection are Poisson distributed with intensity function $\Lambda(\psi), \psi \in [t_0, t]$. Another approach would be to only use the time points at which a photon is detected on the detector. These time points still form a Poisson process, but with a location-sensitive intensity function. This approach leads to a very complicated analysis, which is why we take the alternative approach as in Theorem 3.2.

IV. LINEAR STOCHASTIC SYSTEMS

In general, the motion of an object in cellular environments is subject to different types of forces, e.g., deterministic forces due to the environment and random forces due to random collisions with other objects [21], [22]. The 3D random variable $X(\tau)$ denotes the location of the object at time $\tau \geq t_0$. Then, the motion of the object is assumed to be modeled through a general state space system with state $\tilde{X}(\tau) \in \mathbb{R}^k, \tau \geq t_0$, as, for $\tau_0 := t_0 \leq \tau_1 < \cdots < \tau_{l+1} < \cdots,$

$$\tilde{X}(\tau_{l+1}) = \tilde{\phi}(\tau_l, \tau_{l+1})\tilde{X}(\tau_l) + \tilde{W}(\tau_l, \tau_{l+1}), \quad (13)$$

where we assume that there exists a matrix $H \in \mathbb{R}^{3 \times k}$ such that $X(\tau) = H\tilde{X}(\tau), \tau \geq t_0, \tilde{\phi}(\tau_l, \tau_{l+1}) \in \mathbb{R}^{k \times k}$ is a state transition matrix, and $\{\tilde{W}(\tau_l, \tau_{l+1}), l = 1, 2, \cdots\}$ is a sequence of k-dimensional random variables. We also assume that the initial state $\tilde{X}(t_0)$ is independent of $\tilde{W}$ and its probability density function is given by $p_{\tilde{X}(t_0)}$. The framework we employ here is very general in that, depending on the specific problem we are considering, the system matrices are assumed to be known, unknown or partially known. The unknown elements of the system matrices would form part of the parameter vector that is to be estimated.

The general system of discrete evolution equations described by Eq. (13) can arise, for example, from stochastic differential equations [23]. In particular, in many biological applications, solutions of linear stochastic differential equations are good fits to experimental single-molecule trajectories [23]. As an example, we assume that the motion of the object of interest, e.g., a single molecule, is described by the following linear vector stochastic differential equation [14]

$$dX(\tau) = (V + F(\tau)X(\tau)) d\tau + G(\tau)dB(\tau), \quad \tau \geq t_0, \quad (14)$$

where the 3D random process $X(\tau)$ describes the location of the object at time $\tau \geq t_0$, $F \in \mathbb{R}^{3 \times 3}$ and $G \in \mathbb{R}^{3 \times r}$ are continuous matrix time-functions related to the first order drift and diffusion coefficients, respectively, $V \in \mathbb{R}^3$ is the zero order drift coefficient, and $\{B(\tau) \in \mathbb{R}^r, \tau \geq t_0\}$ is an r-vector Brownian motion (Wiener) process with $E\{dB(\tau)dB(\tau)^T\} = I_{r \times r}, \tau \geq t_0$, where $I_{r \times r}$ is the $r \times r$ identity matrix [12]–[14]. Then, the solution of Eq. (14) at discrete time points $\tau_0 := t_0 \leq \tau_1 <$

$\cdots < \tau_{l+1} < \cdots$ is given by [24]

$$X(\tau_{l+1}) = \phi(\tau_l, \tau_{l+1})X(\tau_l) + a(\tau_l, \tau_{l+1}) + W(\tau_l, \tau_{l+1}), \quad (15)$$

where the continuous matrix time-function $\phi \in \mathbb{R}^{3\times3}$ is given by

$$\frac{d\phi(t, \tau)}{dt} = F(t)\phi(t, \tau), \quad \phi(\tau, \tau) = I_{3\times3}, \quad \text{for all } t, \tau \geq t_0,$$

$$\phi(t, \tau)\phi(\tau, \psi) = \phi(t, \psi), \quad \text{for all } t, \tau, \psi \geq t_0,$$

and the vector $a(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3\times1}$ is given by

$$a(\tau_l, \tau_{l+1}) := \left(\int_{\tau_l}^{\tau_{l+1}} \phi(\tau, \tau_{l+1})d\tau \right) V.$$

Also, in this case,

$$\left\{ W(\tau_l, \tau_{l+1}) := \int_{\tau_l}^{\tau_{l+1}} \phi(\tau, \tau_{l+1})G(\tau)dB(\tau), l = 1, 2, \cdots \right\}$$

is a zero mean white Gaussian sequence with covariance $Q(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3\times3}$ given by

$$Q(\tau_l, \tau_{l+1}) = \int_{\tau_l}^{\tau_{l+1}} \phi(\tau, \tau_{l+1})G(\tau)G^T(\tau)\phi^T(\tau, \tau_{l+1})d\tau.$$

By letting $X(\tau) = H\tilde{X}(\tau) = I_{3\times3}\tilde{X}(\tau) = \tilde{X}(\tau), \tau \geq t_0$, and $\phi(\tau_l, \tau_{l+1}) = \tilde{\phi}(\tau_l, \tau_{l+1})$, we obtain expressions of the form of Eq. (13), where we assume that

$$\left\{ \tilde{W}(\tau_l, \tau_{l+1}) = a(\tau_l, \tau_{l+1}) + W(\tau_l, \tau_{l+1}), l = 1, 2, \cdots \right\}$$

is a white Gaussian sequence with mean $a(\tau_l, \tau_{l+1})$ and covariance $Q(\tau_l, \tau_{l+1})$.

V. Maximum Likelihood Estimation

In this section, we provide a general framework to calculate the maximum likelihood estimates of the parameters of interest for both fundamental and practical data models. In general, these parameters can include the ones that describe the motion of the object, such as drift and diffusion coefficients, or the ones related to the image formation of the object on the detector, such as the intensity function. In the following, we briefly explain the basis of the maximum likelihood estimation.

A. Maximum Likelihood Estimation for Fundamental Data Model

Let Θ denote the parameter space that is an open subset of $\mathbb{R}^{1\times n}$.[*] The maximum likelihood estimate $\hat{\theta}_{mle}$ of $\theta \in \Theta$ for the fundamental data model is given by

$$\hat{\theta}_{mle} = \text{argmin}_{\theta \in \Theta} \left(-\log \mathcal{L}_f(\theta|r_{1:L}) \right),$$

where $r_1, \ldots, r_L \in \mathbb{R}^2$ denote the acquired data and $\mathcal{L}_f(\theta|r_{1:L}) = p^\theta_{R_{1:L}|N(t)}(r_{1:L}|L)$ denotes the likelihood function for the fundamental data model given by Eq. (6).

In the rest of this paper, we only focus on the estimation of the parameters of the motion model, such as drift coefficient, diffusion coefficient and initial location of the molecule, i.e.,

---

*This assumption is made for ease of exposition.

we assume that Λ and f_x are independent of θ. All parameters of the trajectory of the molecule have been encapsulated in the parameter vector θ and our approach does not have significant restrictions on which parameters can be included in this parameter vector. We also denote the dependence of the variable/function on the parameter vector θ, by adding θ to its symbol as a superscript or subscript. For example, $p_{R_{1:L}|N(t)}$ and F in Eq. (6) are denoted by $p^\theta_{R_{1:L}|N(t)}$ and F_θ, respectively.

B. Maximum Likelihood Estimation for Practical Data Model

The maximum likelihood estimate $\hat{\theta}_{mle}$ of $\theta \in \Theta$ for the practical data model is given by

$$\hat{\theta}_{mle} = \text{argmin}_{\theta \in \Theta} \left(-\log \mathcal{L}_p(\theta|z_1, \ldots, z_K) \right),$$

where $\{z_1, \ldots, z_K\}, z_1, \ldots, z_K = 0, 1, \cdots; L = \sum_{k=1}^{K} z_k$, denotes an image with K pixels and $\mathcal{L}_p$ denotes the likelihood function for the infinite practical data model given by, according to Theorem 3.2,

$$\mathcal{L}_p(\theta|z_1, \ldots, z_K)$$
$$= Pr^\theta[S_1 = z_1, \ldots, S_K = z_K]$$
$$= p_L Pr^\theta[S_1 = z_1, \ldots, S_K = z_K|N(t) = L]$$
$$= p_L \int_{\Delta^L} \left[\sum_{v_{1:L} \in A_K^L(z_{1:K})} \int_{\mathbb{R}^{3\times L}} \prod_{l=1}^{L} I_{C_{v_l}}(x_l) \right.$$
$$\left. \times p^\theta_{X(\tau_1), \ldots, X(\tau_L)}(x_{1:L}) \, dx_{1:L} \right] p_{T_{1:L}|N(t)}(\tau_{1:L}|L) \, d\tau_{1:L},$$
$$(16)$$

where

$$I_{C_{v_l}}(x_l) := \int_{C_{v_l}} f_{x_l}(r)dr, \quad l = 1, \ldots, L, \quad (17)$$

is the probability that, given that the systems state is x_l at time τ_l, the photon that arrives at time τ_l hits C_{v_l}. For the practical data model, we have

$$\mathcal{L}_p(\theta|z_1, \ldots, z_K)$$
$$= \sum_{z=0}^{\infty} p_{L+z} \int_{\Delta^{L+z}} \left[\sum_{v_{1:L+z} \in A_{K+1}^{L+z}(z_{1:K}, z)} \int_{\mathbb{R}^{3\times(L+z)}} \right.$$
$$\left. \times \prod_{l=1}^{L+z} I_{C_{v_l}}(x_l) p^\theta_{X(\tau_1), \ldots, X(\tau_{L+z})}(x_{1:L+z}) \, dx_{1:L+z} \right]$$
$$\times p_{T_{1:L+z}|N(t)}(\tau_{1:L+z}|L+z) \, d\tau_{1:L+z}. \quad (18)$$

In general, computing the integrals of the likelihood function is not a trivial task. Here, based on the Monte Carlo approach provided in [25]–[27], we develop an algorithm to approximate these integrals. The basis of our algorithm is the law of large numbers, which can be stated as follows.

Let $\{\tau_1^n, \ldots, \tau_L^n\}_{n=1}^N$ be N sequences of L time points in the interval $[t_0, t]$, each are arrivals of a Poisson process with intensity function $\Lambda(\psi)$ and such that precisely L time points are in the exposure time interval $[t_0, t]$. Let $X(\tau_1^n), \ldots, X(\tau_L^n), n =$

$1, \ldots, N$, be 3D random variables that describe the locations of the object at time points $\tau_0 := t_0 \leq \tau_1^n < \cdots < \tau_L^n \leq t$. Let $p_{X(\tau_1^n), \ldots, X(\tau_L^n)}$ be the joint distribution of $X(\tau_1^n), \ldots, X(\tau_L^n)$. Then, the likelihood function (Eq. (16)) can be approximated as

$$\mathcal{L}_p(\theta | z_1, \ldots, z_K)$$

$$\approx p_L \frac{1}{N} \sum_{n=1}^{N} \sum_{v_{1:L} \in A_K^L(z_{1:K})} \left[\frac{1}{M_c} \sum_{m=1}^{M_c} \left(\prod_{l=1}^{L} I_{C_{v_l}}(x_l^{m,n}) \right) \right], \tag{19}$$

where according to the law of large numbers

$$\lim_{M_c \to \infty} \frac{1}{M_c} \sum_{m=1}^{M_c} \left(\prod_{l=1}^{L} I_{C_{v_l}}(x_l^{m,n}) \right)$$

$$= E \left\{ \prod_{l=1}^{L} I_{C_{v_l}}(X(\tau_l)) \right\}$$

$$= \int_{\mathbb{R}^{2 \times L}} \left(\prod_{l=1}^{L} I_{C_{v_l}}(x_l) \right) p_{X(\tau_1^n), \ldots, X(\tau_L^n)}(x_{1:L}) \, dx_{1:L}, \tag{20}$$

in which $\{x^{m,n} := (x_1^{m,n}, \ldots, x_L^{m,n})\}_{m=1}^{M_c}, x_l^{m,n} \in \mathbb{R}^3, l = 1, \ldots, L, m = 1, \ldots, M_c$, is a sequence of M_c independent and identically distributed trajectories drawn from the distribution $p_{X(\tau_1^n), \ldots, X(\tau_L^n)}$. As the state process $\{\tilde{X}(\tau)\}_{\tau \geq t_0}$ is a Markov process (see Eq. (13)) we carry out the Monte Carlo simulations for this state process and then obtain the simulations for $\{X(\tau)\}_{\tau \geq t_0}$ as $X(\tau) = H\tilde{X}(\tau), \tau \geq t_0$. As $\{\tilde{X}(\tau)\}_{\tau \geq t_0}$ is Markov, the joint probability density function $p_{\tilde{X}(\tau_1), \ldots, \tilde{X}(\tau_L)}$ of $\tilde{X}(\tau_1), \ldots, \tilde{X}(\tau_L)$ is given by

$$p_{\tilde{X}(\tau_1), \ldots, \tilde{X}(\tau_L)}(\tilde{x}_{1:L}) = p_{\tilde{X}(\tau_1)}(\tilde{x}_1)$$

$$\times \prod_{l=2}^{L} p_{\tilde{X}(\tau_l) | \tilde{X}(\tau_{l-1})}(\tilde{x}_l | \tilde{x}_{l-1}). \tag{21}$$

We draw M_c trajectories $x^{m,n}, m = 1, \ldots, M_c, n = 1, \ldots, N$, through the following Monte Carlo algorithm [25]–[27]:

Algorithm 5.1 (Monte Carlo method): **Step 0:** Suppose we have observed the arrival of L photons on the detector during the exposure time interval. Let $\{\tau_1^n, \ldots, \tau_L^n\}_{n=1}^{N}$ be N sequences of L time points, each are arrival times of a Poisson process with intensity function $\Lambda(\psi)$ and with precisely L time points in the exposure time interval $[t_0, t]$.

Step 1: For each n, draw independent and identically distributed (i.i.d.) samples $\{\tilde{x}_1^{m,n}\}_{m=1}^{M_c}$ according to $p_{\tilde{X}(\tau_1^n)}(\tilde{x}), \tilde{x} \in \mathbb{R}^k$, i.e., $\tilde{x}_1^{m,n} \sim p_{\tilde{X}(\tau_1^n)}(\tilde{x}), m = 1, \ldots, M_c$.

Step 2: For each n, draw i.i.d. samples $\{\tilde{x}_2^{m,n}\}_{m=1}^{M_c}$ according to $p_{\tilde{X}(\tau_2^n) | \tilde{X}(\tau_1^n)}(\tilde{x} | \tilde{x}_1^{m,n}), \tilde{x} \in \mathbb{R}^k$, i.e., $\tilde{x}_2^{m,n} \sim p_{\tilde{X}(\tau_2^n) | \tilde{X}(\tau_1^n)}(\tilde{x} | \tilde{x}_1^{m,n}), m = 1, \ldots, M_c$.

$\vdots$

Step L. For each n, draw i.i.d. samples $\{\tilde{x}_L^{m,n}\}_{m=1}^{M_c}$ according to $p_{\tilde{X}(\tau_L^n) | \tilde{X}(\tau_{L-1}^n)}(\tilde{x} | \tilde{x}_{L-1}^{m,n}), \tilde{x} \in \mathbb{R}^k$, i.e., $\tilde{x}_L^{m,n} \sim p_{\tilde{X}(\tau_L^n) | \tilde{X}(\tau_{L-1}^n)}(\tilde{x} | \tilde{x}_{L-1}^{m,n}), m = 1, \ldots, M_c$.

Step $L + 1$. Repeat Steps 0 to L, N times and approximate the likelihood function by Eqs. (19) and (20) by setting $x_l^{m,n} := H\tilde{x}_l^{m,n}, m = 1, \ldots, M_c, l = 1, \ldots, L, n = 1, \ldots, N$.

Note that with the same technique we can approximate Eq. (18). The infinite sum in Eq. (18) can be approximated by truncating the summation. As the denominator of a term with index z is $(L + z)!$ the summation can be expected to converge quickly.

We further assess the performance of the above algorithm in the computation of the likelihood function for the cases that we have a small number of photons detected in simulated pixelated images in Examples 5.1, 5.2, 5.3 and 5.4. In Example 5.1, we investigate the case that we have only one photon, and evaluate the convergence of the Monte Carlo approach with different numbers of samples. Example 5.2 is focused on the computation of maximum likelihood estimates of the unknown parameters of the motion model, and the evaluation of the mean and standard deviation of the estimates, for the cases that the number of photons detected in the simulated pixelated images is equal to four. In Example 5.3, for data sets of simulated images of a stationary molecule, we show that the standard deviations of location estimates match well with the CRLB. Finally, in Example 5.4, we examine the effect of pixel size of the detector on the performance of the estimation method, in terms of standard deviation, a fact that has not been considered in most available methods.

Example 5.1: Assume that we have a typical two-dimensional single molecule trajectory $X(\tau)$ in the object space (Eq. (14)), where the time point $\tau = 0.01$ ms is fixed, with the first order drift coefficient $F = -10/s$ (we assume that there is no zero order drift, i.e., $V = 0$) and the diffusion coefficient $D = 1$ μ^2/s ($G := \sqrt{2D}$). Also, we assume that the initial location of the molecule is $x(t_0) = (2.4, 2.4)^T$ μm. In the fundamental data model, detected locations of the photons emitted from the molecule in the image space are simulated using a zero-mean Gaussian profile with covariance matrix $\Sigma = 0.01 I_{2 \times 2}$ $\mu^2 m$. In the practical data model, a 60×60 pixelated detector with square pixels of width $W = 16$ μm is used to acquire the pixelated image of the molecule trajectory. Assume that the photon emitted from the object hits the pixel C_1 centered at $(c_x^1, c_y^1) = (230.75, 237.25)^T$ μm on the image plane. Then, using Algorithm 5.1, we calculate the probability that this event takes place as

$$Pr\left[U((X(\tau)) \in C_1 \right] = \int_{\mathbb{R}^2} I_{C_1}(x) p_{X(\tau)}(x) dx$$

$$\approx \frac{1}{M_c} \sum_{m=1}^{M_c} I_{C_1}(x^m),$$

where, for an invertible magnification matrix $M \in \mathbb{R}^{2 \times 2}$ and $x \in \mathbb{R}^2$,

$$I_{C_1}(x) = \frac{1}{|\det(M)|} \int_{C_1} q\left(M^{-1} r - x\right) dr$$

$$= \frac{1}{|\det(M)|} \int_{c_x^1 - \frac{W}{2}}^{c_x^1 + \frac{W}{2}} \int_{c_y^1 - \frac{W}{2}}^{c_y^1 + \frac{W}{2}} q\left(M^{-1}(r_x, r_y)^T - x\right) dr_y dr_x, \tag{22}$$
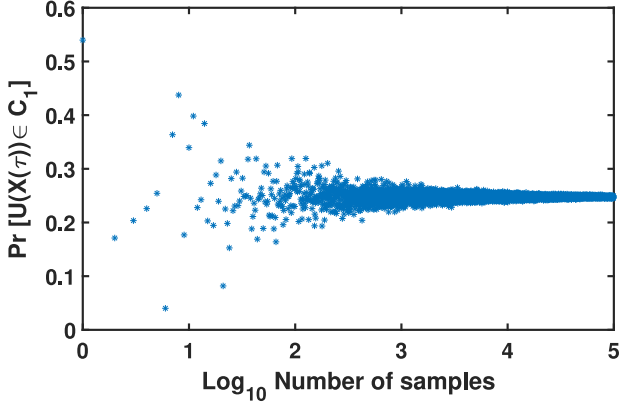
Fig. 2. Convergence of the Monte Carlo method. The probabilities $Pr[U((X(\tau)) \in C_1] = 0.24735$ for different number M_c of Monte Carlo samples, which ranges from 1 to 100,000, where $X(\tau)$ is a two-dimensional single molecule trajectory (Eq. (14)), are shown in which the time point $\tau = 0.01$ ms is fixed, with the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1$ μ^2/s ($G := \sqrt{2D}$). Also, assume that the initial location of the molecule is $x(t_0) = (2.4, 2.4)^T$ μm. Detected locations of the photons emitted from the molecule in the image space are simulated using a zero-mean Gaussian model with covariance matrix $\Sigma = 0.01 I_{2 \times 2}$ μ^2m. A 60×60 pixelated detector with square pixels of width $W = 16$ μm is used to acquire the pixelated image of the molecule trajectory.
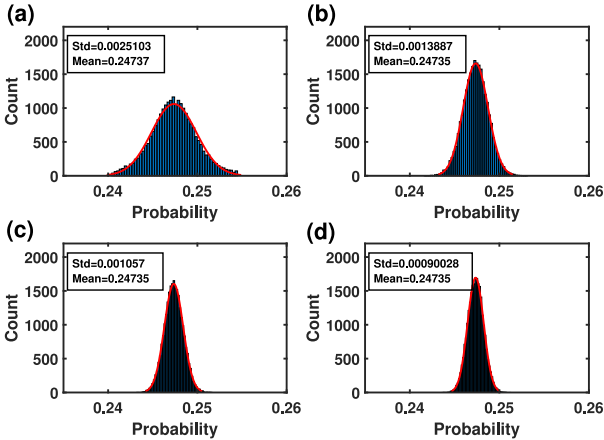

Fig. 3. Histograms, means and standard deviations of the probabilities computed using the Monte Carlo method with different number of samples. Gaussian models fitted to the histograms of the probabilities computed using the Monte Carlo method, where the number of samples ranges between (a) 1 to 25,000, (b) 25,001 to 50,000, (c) 50,001 to 75,000, and (d) 75,001 to 100,000 ((a) first, (b) second, (c) third, and (d) fourth quarters of the data shown in Fig. 2).

and $\{x^m\}_{m=1}^{M_c}, x^m \in \mathbb{R}^2, m = 1, \ldots, M_c$, is a sequence of independent and identically distributed samples drawn from the distribution $p_{X(\tau)}$ using Algorithm 5.1. In Fig. 2, we have shown the probabilities $Pr[U((X(\tau)) \in C_1]$ computed for different number M_c of Monte Carlo samples. As can be seen in Fig. 3, the standard deviation of the probabilities decreases by increasing the number of samples, where the number of samples ranges between (a) 1 to 25,000, (b) 25,001 to 50,000, (c) 50,001 to 75,000, and (d) 75,001 to 100,000 ((a) first, (b) second, (c) third, and (d) fourth quarters of the data. This suggests the convergence of these probabilities.
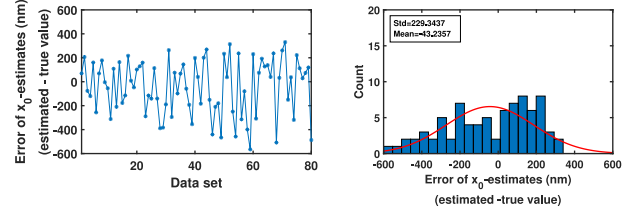

Fig. 4. Analysis of the error of initial location estimates from pixelated image data sets of single molecule trajectories for the Gaussian measurement noise case. Differences between the estimates of the initial x_0- and y_0-location of the molecule and their true values from the images of the molecule trajectories simulated using Eqs. (14) with four time points, where the time points are drawn from a Poisson process, and the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1.5$ μm^2/s ($G := \sqrt{2D}$). The initial location of the molecule is $x(t_0) = (2.3, 2.3)^T$ μm. The locations of the photons emitted from the molecule trajectories, in the image space, are simulated with the Gaussian measurement noise (Eq. (4)) and $\sigma = 0.1$ μm. These photons are detected using a pixelated detector of pixel size and image size of 6.5×6.5 μm and 60×60 pixels, respectively.

Example 5.2: We next examine the performance of our proposed parameter estimation method. For this purpose, we simulated 100 pixelated images of single molecule trajectories. These trajectories were simulated using Eq. (14) with four time points, where the time points were drawn from a Poisson process in the exposure time interval $[0,20]$ ms, and the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1.5$ μm^2/s ($G := \sqrt{2D}$). Also, we assumed that the initial location of the molecule was $(2.3, 2.3)^T$ μm. The locations of the photons emitted from the molecule trajectories, in the image space, were simulated with the Gaussian measurement noise (Eq. (4)) and $\sigma = 0.1$ μm. We assumed that these photons were detected using a pixelated detector of pixel size and image size of 6.5×6.5 μm and 60×60 pixels, respectively. For example, in the 50th image, we have four photons which are detected in the pixels centered at $(217.75, 230.75)^T, (172.25, 217.75)^T, (146.25, 204.75)^T$ and $(139.75, 198.25)^T$ μm positions on the image plane, denoted by C_1, C_2, C_3 and C_4. Then, the summation of the likelihood function (19) is performed over $A_4^4(1, 1, 1, 1) = \{(1, 2, 3, 4), (1, 2, 4, 3), \ldots, (4, 3, 2, 1)\}$ (the size of $A_4^4(1, 1, 1, 1)$ in this case is equal to $4! = 24$).

We then estimated all parameters of the trajectories, e.g., initial location of the molecule, drift and diffusion coefficients, together using Algorithm 5.1, where the number of spatial M_c and temporal N Monte Carlo samples equal to 1000 and 100, respectively. The errors (estimate - true value) of the estimation are shown in Figs. 4 and 5. As can be seen in these figures, the spreads of the errors are around zero and there is no systematic bias associated with the estimates.

We also applied the algorithm to the pixelated images of single molecule trajectories simulated using an Airy point spread functions with $\alpha = \frac{2\pi n_a}{\lambda} = 13.23$, which corresponds to a Gaussian profile with $\sigma = 0.1$ μm. The parameters of the molecule trajectories were the same as the parameters of the data set of Fig. 4. As can be seen in Figs. 6 and 7, we have obtained similar results as in the Gaussian case.

Example 5.3: We further evaluate the performance of the proposed method in terms of the standard deviation of the
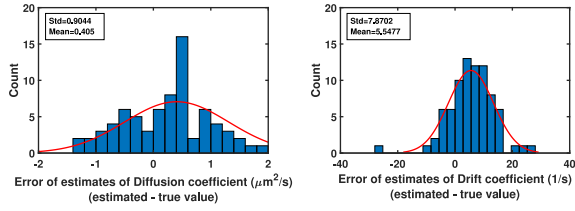
Fig. 5. Analysis of the error of diffusion coefficient and drift coefficient estimates from pixelated image data sets of single molecule trajectories for the Gaussian measurement noise case. Differences between the diffusion (first order drift) coefficient estimates and the true diffusion (first order drift) coefficient value for data sets of Fig. 4.
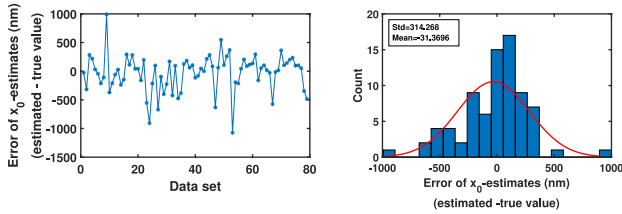

Fig. 6. Analysis of the error of initial location estimates from pixelated image data sets of single molecule trajectories for the Airy measurement noise case. Differences between the estimates of the initial x_0- and y_0-location of the molecule and their true values from the images of the molecule trajectories simulated using the parameters of the data set of Fig. 4. The locations of the photons emitted from the molecule trajectories, in the image space, are simulated using an Airy model with $\alpha = \frac{2\pi n_a}{\lambda} = 13.23$. These photons are detected using a pixelated detector of pixel size and image size of $6.5 \times 6.5 \ \mu m$ and 60×60 pixels, respectively.
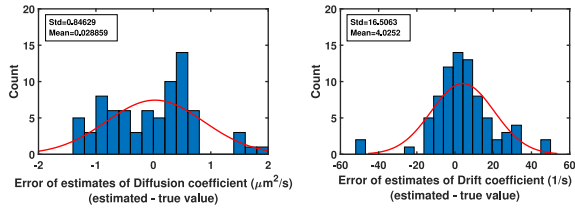

Fig. 7. Analysis of the error of diffusion coefficient and drift coefficient estimates from pixelated image data sets of single molecule trajectories for the Airy measurement noise case. Differences between the diffusion (first order drift) coefficient estimates and the true diffusion (first order drift) coefficient value for data sets of Fig. 6.

estimates. In order to do this, we simulated the pixelated images of a stationary object using a pixelated detector of pixel size and image size of $6.5 \times 6.5 \ \mu m$ and 60×60 pixels, respectively, assuming that the number of photons detected by the detector is equal to three. The locations of the photons in the image space were simulated with the Gaussian measurement noise (Eq. (4)) and $\sigma = 0.1 \ \mu m$. We then estimated the location of the molecule using Algorithm 5.1, where the number of Monte Carlo samples is equal to 10000. The errors of the location estimates are shown in Fig. 9. As before, the errors are spreading around zero and no systematic bias can be seen. We also calculated the standard deviations of the estimates. These standard deviations, which are computed as 57.4 nm and 59.6 nm for the x_0- and y_0-locations of the molecule, respectively, are close to the localization accuracy, i.e., the positive definite square root of the CRLB, which is
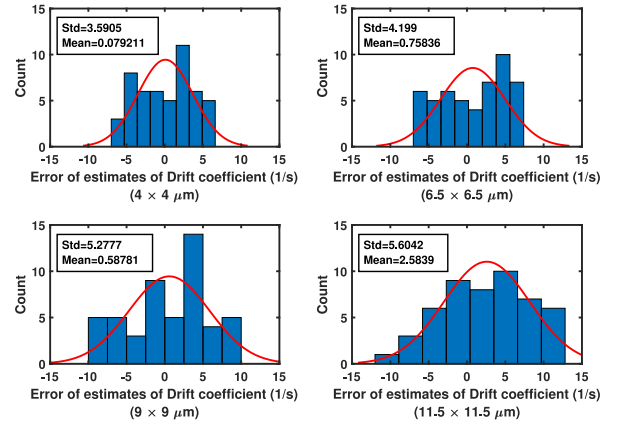

Fig. 8. Analysis of the error of drift coefficient estimates from pixelated images with different pixel sizes. Analysis of the error of drift coefficient estimates from 100×100 pixelated image data sets of single molecule trajectories for the Gaussian measurement noise case simulated using the parameters of Fig. 4, assuming that four photons are detected in each image, where the pixel size ranges from 4×4 to $11.5 \times 11.5 \ \mu m$.
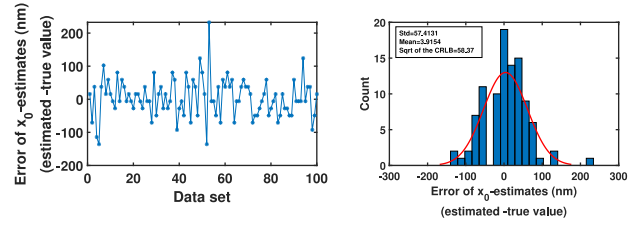

Fig. 9. Analysis of the error of location estimates from pixelated image data sets of a stationary molecule for the Gaussian measurement noise case. Differences between the estimates of the initial x_0- and y_0-location of the molecule and their true values from the simulated images of a stationary molecule using a pixelated detector of pixel size and image size of $6.5 \times 6.5 \ \mu m$ and 60×60 pixels, respectively, assuming that three photons are detected by the detector. The locations of the photons in the image space are simulated with the Gaussian measurement noise (Eq. (4)) and $\sigma = 0.1 \ \mu m$.

given as 58.37 nm for both x- and y-directions, reported in [18], [28]–[30].

Example 5.4: In this example, we investigate the effect of pixel size of the detector on the standard deviation of drift coefficient estimates. For this purpose, 100×100 pixelated images were simulated using the parameters provided in Example 5.1, assuming that four photons are detected in each image, and the pixel size ranges from 4×4 to $11.5 \times 11.5 \ \mu m$ (for each pixel size, we have 50 images). As shown in the Figure 8, the standard deviation of the estimates gets worse as the pixel size of the detector increase.

Here, we only consider a small number of photons, since, in general, the computation of the likelihood function (Eq. (11)) is expensive. It is mostly because of the large number of members of the set $A_K^L(z_1, \ldots, z_K)$, which is equal to $\frac{L!}{z_1! \cdots z_K!}$, when L increases. For example, in case of having a 32×32-pixels detector with $L = 1000$ and $K = 1024, z_1 = 500, z_2 = \cdots = z_{501} = 1, z_{502} = \cdots = z_{1024} = 0$, we have $|A_K^L(z_1, \ldots, z_K)| = \frac{1000!}{500!} = 1000 \times \cdots \times 501$, which is an extremely large number. To arrive at

an estimator that can be practically computed, further research is needed for the cases in which the cardinality of the set $A_K^L(z_1, \ldots, z_K)$ is too large.

VI. FISHER INFORMATION MATRIX

The Fisher information matrix was introduced in [18], [19] for the analysis of microscopy image and data analysis problems, in particular for single molecule microscopy. It has since been used extensively for the evaluation of image analysis algorithms [31], [32], experiment design [33], design of novel point spread functions [34]–[36] etc. The purpose of this section is to develop a framework for the computation of the Fisher information for the experimental setting considered here.

According to a well-known result in information theory, known as the Cramér-Rao inequality, the covariance matrix of any unbiased estimator $\hat{\theta}$ of an unknown parameter vector θ is bounded from below by the inverse of the Fisher information matrix $I(\theta)$, i.e., $\text{Cov}(\hat{\theta}) \geq I^{-1}(\theta)$, assuming that $I(\theta)$ is invertible. [37]. Then, the smallest standard deviation of the estimates that can be obtained, which is independent of the estimation method, only depends on the statistical model of the data, and is given by the positive definite square root of the inverse of the Fisher information matrix, referred to as the Cramér-Rao lower bound (CRLB). Large sample theory tells us that for a number of large repeats the maximum likelihood estimator has a variance that is given by the inverse of the Fisher information matrix [38], [39]. The usefulness of the CRLB also derives from the fact that in many practical circumstances for a maximum likelihood estimator it has been found that the variance of the estimator is well approximated by the inverse of the Fisher information matrix. In the following, for the fundamental and practical data models introduced in Sections III-A and III-B, we calculate the Fisher information matrix.

A. Fisher Information Matrix for Fundamental Data Model

In this section, for the fundamental data model, we first, in Definition 6.1, introduce the notation for the Fisher information matrix of the fundamental data model given the number of detected photons.

Definition 6.1: Let the parameter space Θ describe an open subset of $\mathbb{R}^{1 \times n}$ containing the true parameters. For $L = 1, 2, \cdots$, and a row parameter vector $\theta \in \Theta$, we introduce the $n \times n$ Fisher information matrix of the fundamental data model given $N(t) = L$, as

$$
I_{N(t)=L}^f(\theta) := E_{p_{R_{1:L}|N(t)=L}^\theta} \left\{ \left(\frac{\partial \log p_{R_{1:L}|N(t)}^\theta (r_{1:L}|L)}{\partial \theta} \right)^T \right.
$$
$$
\left. \times \left(\frac{\partial \log p_{R_{1:L}|N(t)}^\theta (r_{1:L}|L)}{\partial \theta} \right) \right\}
$$
$$
= \int_{\mathbb{R}^{2 \times L}} p_{R_{1:L}|N(t)}^\theta (r_{1:L}|L)
$$
$$
\times \left(\frac{\partial \log p_{R_{1:L}|N(t)}^\theta (r_{1:L}|L)}{\partial \theta} \right)^T
$$

$$
\times \left(\frac{\partial \log p_{R_{1:L}|N(t)}^\theta (r_{1:L}|L)}{\partial \theta} \right) dr_{1:L}, \quad (23)
$$

where $E_{p_{R_{1:L}|N(t)=L}^\theta}$ is the expected value with respect to the probability density $p_{R_{1:L}|N(t)=L}^\theta$.

In the above definition, we introduced the Fisher information matrix conditioned on the number of detected photons, which will be further used to derive an expression for the the Fisher information matrix of the practical data model. Note that the relationship between the actual (non-conditional) $I^f(\theta)$ and conditional $I_{N(t)=L}^f(\theta)$ Fisher information matrix for the fundamental data model is given by

$$
I^f(\theta) = \sum_{L=1}^\infty \int_{\mathbb{R}^{2 \times L}} p_{R_{1:L}}^\theta (r_{1:L}) \left(\frac{\partial \log p_{R_{1:L}}^\theta (r_{1:L})}{\partial \theta} \right)^T
$$
$$
\times \left(\frac{\partial \log p_{R_{1:L}}^\theta (r_{1:L})}{\partial \theta} \right) dr_{1:L}
$$
$$
= \sum_{L=1}^\infty p_L I_{N(t)=L}^f(\theta). \quad (24)
$$

In [16], we obtained a recursive formulation for the Fisher information matrix of the fundamental data model, which is computationally expensive for non-Gaussian measurements. In the following theorem, based on the new expression derived for the likelihood function in Theorem 3.1, we provide a closed form expression for the Fisher information matrix.

Theorem 6.1: For a row parameter vector $\theta \in \Theta$, the Fisher information matrix $I_{N(t)=L}^f(\theta)$, $L = 1, 2, \cdots$, of the fundamental data model given $N(t) = L$, can be calculated as (we assume that $p_{R_{1:L}|N(t)}^\theta$ is strictly positive)

$$
I_{N(t)=L}^f(\theta)
$$
$$
= \int_{\mathbb{R}^{2 \times L}} \frac{DF_\theta^T (r_{1:L}) \, DF_\theta (r_{1:L})}{\int_{\mathbb{R}^{3 \times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) \, d\tau_{1:L} \right) dx_{1:L}} dr_{1:L}, \quad (25)
$$

where

$$
DF_\theta (r_{1:L})
$$
$$
:= \frac{\partial}{\partial \theta} \int_{\mathbb{R}^{3 \times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) \, d\tau_{1:L} \right) dx_{1:L}, \quad (26)
$$

in which $F_\theta(.)$ is given by Eq. (7).

Proof: See Appendix IX-D. ∎

The structure of the presentation of the Fisher information is interesting to observe. The Fisher information matrix $I_{N(t)=L}^f(\theta)$ is given as a multidimensional integral of the integrand over the different photon impact locations on the detector. Importantly, the same integrand will also play an important role in the expression of the Fisher information matrix for the practical data model, that we will consider next.

B. Fisher Information Matrix for Practical Data Model

In this section, we use the results obtained in the previous section to calculate the Fisher information matrix for the practical data model. We first, in the following definition, introduce a notation for the Fisher information matrix of the practical data model.

Definition 6.2: Let the parameter space Θ describe an open subset of $\mathbb{R}^{1 \times n}$ containing the true parameters. We introduce the following notation for the Fisher information matrix of the practical data model, for a row parameter vector $\theta \in \Theta$,

$$
I^p(\theta) := E_{Pr^\theta [S_1 = z_1, \dots, S_K = z_K]}
$$

$$
\left\{ \left(\frac{\partial \log Pr^\theta [S_1 = z_1, \dots, S_K = z_K]}{\partial \theta} \right)^T \right.
$$

$$
\left. \times \left(\frac{\partial \log Pr^\theta [S_1 = z_1, \dots, S_K = z_K]}{\partial \theta} \right) \right\}
$$

$$
= \sum_{z_1, \dots, z_K = 0}^{\infty} Pr^\theta [S_1 = z_1, \dots, S_K = z_K]
$$

$$
\times \left(\frac{\partial \log Pr^\theta [S_1 = z_1, \dots, S_K = z_K]}{\partial \theta} \right)^T
$$

$$
\times \left(\frac{\partial \log Pr^\theta [S_1 = z_1, \dots, S_K = z_K]}{\partial \theta} \right).
$$

In the following theorem, we calculate the Fisher information matrix of the practical data model introduced in the above definition.

Theorem 6.2: 1) For a row parameter vector $\theta \in \Theta$, the Fisher information matrix $I^p(\theta)$ of the infinite practical data model can be calculated by Eq. (27), shown at the bottom of this page, where DF_θ and F_θ are given by Eqs. (26) and (7).

2) The Fisher information matrix $I^p(\theta)$ of the practical data model can be calculated by Eq. (28), shown at the bottom of this page.

Proof: See Appendix IX-E. ∎

As can be seen from the results of the above theorem, the key to computing the Fisher information expression is through the computation of the derivatives of the probability density function of the states. In [40], for time-invariant systems, an easy-to-compute recursive formulation has been developed to deal with the derivatives of the probability density function of the states, and therefore, to compute the Fisher information matrix.

VII. EFFECT OF NOISE

So far, we have assumed that all the photons detected by a pixelated detector come from the object of interest. However, in practice, fluorescence microscopy images always are corrupted by a background noise corresponding to the photons emitted from background components. The number of these photons in the $k^{th}, k = 1, \dots, K$, pixel is described by an independently

$$
I^p(\theta) = \sum_{L=1}^{\infty} p_L \sum_{z \in \mathbb{N}_0^K : |z| = L} \left(\frac{\sum_{v_{1:L} \in A_K^L(z)} \sum_{v'_{1:L} \in A_K^L(z)} \int_{C_{v_{1:L}}} \int_{C_{v'_{1:L}}} DF_\theta^T(r_{1:L}) DF_\theta(r'_{1:L}) dr'_{1:L} dr_{1:L}}{\sum_{v_{1:L} \in A_K^L(z)} \int_{C_{v_{1:L}}} \left[\int_{\mathbb{R}^{3 \times L}} \left(\int_{\Delta^L} F_\theta(r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L} \right] dr_{1:L}} \right). \tag{27}
$$

$$
I^p(\theta) = \sum_{(\bar{z}, \bar{z}', L) \in \mathbb{N}_0^3 : L + \bar{z} \neq 0, L + \bar{z}' \neq 0} p_{L+\bar{z}} p_{L+\bar{z}'} \sum_{z \in \mathbb{N}_0^K : |z| = L}
$$

$$
\left(\frac{\sum_{v_{1:L+\bar{z}} \in A_{K+1}^{L+\bar{z}}(z, \bar{z})} \sum_{v'_{1:L+\bar{z}'} \in A_{K+1}^{L+\bar{z}'}(z, \bar{z}')} \int_{C_{v_{1:L+\bar{z}}}} \int_{C_{v'_{1:L+\bar{z}'}}} DF_\theta^T(r_{1:L+\bar{z}}) DF_\theta(r'_{1:L+\bar{z}'}) dr'_{1:L+\bar{z}'} dr_{1:L+\bar{z}}}{\sum_{\bar{z}=0, L+\bar{z} \neq 0}^{\infty} p_{L+\bar{z}} \sum_{v_{1:L+\bar{z}} \in A_{K+1}^{L+\bar{z}}(z, \bar{z})} \int_{C_{v_{1:L+\bar{z}}}} \left[\int_{\mathbb{R}^{3 \times (L+\bar{z})}} \left(\int_{\Delta^{L+\bar{z}}} F_\theta(r_{1:L+\bar{z}}, x_{1:L+\bar{z}}, \tau_{1:L+\bar{z}}) d\tau_{1:L+\bar{z}} \right) dx_{1:L+\bar{z}} \right] dr_{1:L+\bar{z}}} \right). \tag{28}
$$

$$
I^p(\theta) = \int_{\mathbb{R}^K} \frac{\sum_{z_1, \dots, z_K = 0}^{\infty} \sum_{z'_1, \dots, z'_K = 0}^{\infty} p_{\mathcal{I}_{1:K} | S_{1:K}}(i_{1:K} | z_{1:K}) p_{\mathcal{I}_{1:K} | S_{1:K}}(i_{1:K} | z'_{1:K})}{\sum_{z_1, \dots, z_K = 0}^{\infty} p_{\mathcal{I}_{1:K} | S_{1:K}}(i_{1:K} | z_{1:K}) Pr^\theta [S_1 = z_1, \dots, S_K = z_K]}
$$

$$
\times \left(\frac{\partial Pr^\theta [S_1 = z_1, \dots, S_K = z_K]}{\partial \theta} \right)^T \left(\frac{\partial Pr^\theta [S_1 = z'_1, \dots, S_K = z'_K]}{\partial \theta} \right) di_{1:K}. \tag{29}
$$

$$
I^p(\theta) := \int_{\mathbb{R}^K} \sum_{L=1}^{\infty} \sum_{L'=1}^{\infty} p_L p_{L'} \sum_{z \in \mathbb{N}_0^K : |z| = L} \sum_{z' \in \mathbb{N}_0^K : |z'| = L'} p_{\mathcal{I}_{1:K} | S_{1:K}}(i_{1:K} | z) p_{\mathcal{I}_{1:K} | S_{1:K}}(i_{1:K} | z')
$$

$$
\times \left(\frac{\sum_{v_{1:L} \in A_K^L(z)} \sum_{v'_{1:L'} \in A_K^{L'}(z')} \int_{C_{v_{1:L}}} \int_{C_{v'_{1:L'}}} DF_\theta^T(r_{1:L}) DF_\theta(r'_{1:L'}) dr'_{1:L'} dr_{1:L}}{\sum_{L=1}^{\infty} p_L \sum_{z \in \mathbb{N}_0^K : |z| = L} p_{\mathcal{I}_{1:K} | S_{1:K}}(i_{1:K} | z) \sum_{v_{1:L} \in A_K^L(z)} \int_{C_{v_{1:L}}} \left[\int_{\mathbb{R}^{3 \times L}} \left(\int_{\Delta^L} F_\theta(r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L} \right] dr_{1:L}} \right) di_{1:K}. \tag{30}
$$

Poisson distributed random variable B_k with mean $\beta_k \geq 0$. Also, in a pixelated detector, the acquired image contains a readout noise, which can be modeled as an independent Gaussian distributed random variable E_k with mean $\eta_k \geq 0$ and variance $\sigma_k^2 > 0$. The acquired image by a pixelated detector is then can be described by a collection $\{\mathcal{I}_1^\theta, \ldots, \mathcal{I}_K^\theta\}$ of random variables given by

$$\mathcal{I}_k^\theta = S_k^\theta + B_k + E_k, \quad k = 1, \ldots, K, \quad \theta \in \Theta.$$

Note that S_k^θ and B_k are non-negative integers, but E_k is real-valued. Hence, $\mathcal{I}_k^\theta$ is real-valued. In this case, the likelihood function $\mathcal{L}_p$ is given by

$$\mathcal{L}_p(\theta|i_1, \ldots, i_K) = p_{\mathcal{I}_1, \ldots, \mathcal{I}_K}^\theta (i_1, \ldots, i_K), \quad i_1, \ldots, i_K \in \mathbb{R}, \tag{31}$$

where $p_{\mathcal{I}_{1:K}}^\theta$ denotes the joint probability density function of $\mathcal{I}_1^\theta, \ldots, \mathcal{I}_K^\theta$, and can be calculated as

$$p_{\mathcal{I}_{1:K}}^\theta (i_{1:K}) = \sum_{z_1=0}^\infty \cdots \sum_{z_K=0}^\infty p_{\mathcal{I}_{1:K}|S_{1:K}} (i_{1:K}|z_{1:K})$$
$$\times Pr^\theta[S_1 = z_1, \ldots, S_K = z_K],$$

in which the conditional probability density function of $\mathcal{I}_{1:K}$, given $S_{1:K}$, can be calculated as [19]

$$p_{\mathcal{I}_{1:K}|S_{1:K}} (i_{1:K}|z_{1:K})$$
$$= \prod_{k=1}^K p_{\mathcal{I}_k|S_k} (i_k|z_k)$$
$$= \prod_{k=1}^K \frac{1}{\sqrt{2\pi}\sigma_k} \sum_{l=0}^\infty \frac{\beta_k^l e^{-\beta_k}}{l!} e^{-\frac{1}{2}\left(\frac{i_k - l - \eta_k - z_k}{\sigma_k}\right)^2}$$

and $Pr^\theta[S_1 = z_1, \ldots, S_K = z_K]$ is given by Eqs. (18) or (16). It has been shown that $I^p(\theta)$ can be calculated by Eq. (29), shown at the bottom of the previous page (see Appendix F).

For the infinite practical data model, as calculated in Theorem 6.2, the Fisher information matrix can be rewritten as Eq. (30), shown at the bottom of the previous page.

The Fisher information matrix for an EMCCD detector can be obtained in a relatively straightforward fashion by applying the approaches developed for EMCCD detector in [41] combined with the approach introduced here.

VIII. CONCLUSION

The estimation of biophysical parameters from the observed trajectories of molecules in a live cell environment using single molecule fluorescence microscopy is one of the key experiments of modern molecular cell biology and biophysics. The approach we introduced here is one where we use a general stochastic dynamical system model to model the dynamics of the molecule, which includes dynamics governed by stochastic differential equations. The parameters of interest that are to be estimated can be any parameters that impact the underlying dynamical equations. Examples of such parameters are diffusion or drift coefficients, but also the coordinates of the particle at a particular point in time, such as the starting point of the trajectory. In a

fluorescence microscopy experiment the acquired data is given by the photons that are emitted by the fluorescent label and are captured by the imaging detector. While the photons are assumed to be emitted based on a temporal Poisson process, a modern imaging detector has pixels and captures photons during an exposure interval. Such a detector therefore cannot capture the precise time points and impact locations of the detected photons and only records the accumulated photons in each pixel during the exposure time. Moreover, the impact points of the detected photons in the detector are related to the location of the imaged molecule through the so-called point spread function of the optical system that describes the image formation process of the microscope.

In this paper we have set up a stochastic framework within which the maximum likelihood estimator and Fisher information matrices could be derived for this parameter estimation problem. Central to our approach is a careful probabilistic model for the dynamics of the molecule and photon detection process for detection with a pixelated imaging detector. The resulting analytical expressions, derived without approximations from the general modeling assumptions, are complex due to the intricate relationship between the statistics of the photons emission process by the imaged object and the photons statistics in the pixels of the detector.

Using a Monte Carlo approach we proposed a numerical method for computing the maximum likelihood estimator. We showed, with simulated examples, that this estimator does indeed have desirable properties, such as low or no bias, even for extreme low photon count examples. The analysis of high photon count data poses significant numerical challenges as a large number of separate Monte Carlo simulations need to be carried out. To allow for efficient computations, this calls for further investigation into the numerics of the computation of the maximum likelihood estimator. Computation of the Fisher information is complicated through the need to compute a possible very large number of iterative integrals. As with the computation of the maximum likelihood estimator, it is Monte Carlo based methods that are expected to provide a successful approach. Using simulated data we illustrate that a detailed incorporation of the pixel size in the model for a parameter estimation problem is indeed of importance. In our example we show how the standard deviation of the estimate of the diffusion coefficient does depend on the pixel size.

We hope that the results that are presented here will provide important reference points for approximations that might lead to approaches that are computationally more efficient but need evaluation regarding their accuracy. Having a precise formulation available for the maximum likelihood estimator and the Fisher information matrices will hopefully help to provide important and well characterized tools to analyze molecular dynamics.

IX. APPENDICES

A. Analysis of Poisson Time Points

Lemma 9.1: For $t_0 \in \mathbb{R}$, let $\{N(\tau), \tau \geq t_0\}$ be a Poisson process with intensity function $\Lambda(\tau), \tau \geq t_0$. Let $t_0 \leq T_1 < T_2 < \cdots$, be 1D random variables which describe ordered events

of the process N. Then, the conditional probability density function $p_{T_1,\ldots,T_L|N(t)}$ of $T_1,\ldots,T_L$, given $N(t), t > t_0$, can be calculated as

$$p_{T_1,\ldots,T_L|N(t)}(\tau_{1:L}|L) = L! \frac{\prod_{l=1}^L \Lambda(\tau_l)}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L}.$$

Proof: See [42]. ∎

B. Proof of Theorem 3.1

We have (see Lemma 9.1 in Appendix A for the probability density function of Poisson time points)

$$p_{R_{1:L}|N(t)}(r_{1:L}|L)$$

$$= \int_{\mathbb{R}^{3\times L}} p_{R_{1:L}|X(T_1),\ldots,X(T_L),N(t)}(r_{1:L}|x_{1:L},L)$$

$$\times p_{X(T_1),\ldots,X(T_L)|N(t)}(x_{1:L}|L)\,dx_{1:L}$$

$$= \int_{\mathbb{R}^{3\times L}} p_{R_1|X(T_1)}(r_1|x_1)\cdots p_{R_L|X(T_L)}(r_L|x_L)$$

$$\times \left(\int_{\Delta^L} p_{X(T_1),\ldots,X(T_L),T_{1:L}|N(t)}(x_{1:L},\tau_{1:L}|L)d\tau_{1:L}\right)dx_{1:L}$$

$$= \int_{\mathbb{R}^{3\times L}} p_{R_1|X(T_1)}(r_1|x_1)\cdots p_{R_L|X(T_L)}(r_L|x_L)$$

$$\times \left(\int_{\Delta^L} p_{X(T_1),\ldots,X(T_L)|T_{1:L},N(t)}(x_{1:L}|\tau_{1:L},L)\right.$$

$$\left.\times p_{T_{1:L}|N(t)}(\tau_{1:L}|L)\,d\tau_{1:L}\right)dx_{1:L}$$

$$= \int_{\mathbb{R}^{3\times L}} f_{x_1}(r_1)\cdots f_{x_L}(r_L)$$

$$\times \left(\int_{\Delta^L} p_{X(\tau_1),\ldots,X(\tau_L)}(x_{1:L})\frac{L!\prod_{l=1}^L \Lambda(\tau_l)}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L}d\tau_{1:L}\right)$$

$$\times dx_{1:L}$$

$$= \int_{\mathbb{R}^{3\times L}}\left[\int_{\Delta^L} \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L}\right.$$

$$\left.\times \left(\prod_{i=1}^L f_{x_i}(r_i)\Lambda(\tau_i)\right) p_{X(\tau_1),\ldots,X(\tau_L)}(x_{1:L})d\tau_{1:L}\right]dx_{1:L}.$$

C. Proof of Theorem 3.2

1) According to the definitions of $S_1,\ldots,S_K$, we have, for $z_1,\ldots,z_K = 0,1,\cdots,Pr(L=0) = p_0$ and if $L \geq 1$:

$$Pr[S_1 = z_1,\ldots,S_K = z_K]$$

$$= Pr\left[\bigcup_{v_{1:L}\in A_K^L(z_{1:K})}\left\{\bigcap_{l=1}^L (U(X(T_l)) \in C_{v_l})\right\}\Big| N(t) = L\right]$$

$$\times Pr[N(t) = L], \tag{32}$$

where $L = \sum_{k=1}^K z_k$. Since the events $\{\bigcap_{l=1}^L(U(X(T_l)) \in C_{v_l})\}$ are mutually exclusive, we have

$$Pr\left[\bigcup_{v_{1:L}\in A_K^L(z_{1:K})}\left\{\bigcap_{l=1}^L (U(X(T_l)) \in C_{v_l})\right\}\Big| N(t) = L\right]$$

$$= \sum_{v_{1:L}\in A_K^L(z_{1:K})} Pr\left[\bigcap_{l=1}^L (U(X(T_l)) \in C_{v_l})\Big| N(t) = L\right],$$

and therefore, according to Eq. (6),

$$Pr[S_1 = z_1,\ldots,S_K = z_K]$$

$$= \sum_{v_{1:L}\in A_K^L(z_{1:K})} Pr\left[\bigcap_{l=1}^L (U(X(T_l)) \in C_{v_l})\Big| N(t) = L\right]$$

$$\times Pr[N(t) = L]$$

$$= \sum_{v_{1:L}\in A_K^L(z_{1:K})} p_L \int_{C_{v_{1:L}}} p_{R_{1:L}|N(t)}(r_{1:L}|L)\,dr_{1:L}$$

$$= p_L \sum_{v_{1:L}\in A_K^L(z_{1:K})}\int_{C_{v_{1:L}}}\left[\int_{\mathbb{R}^{3\times L}}\left(\int_{\Delta^L} F(r_{1:L},x_{1:L},\tau_{1:L})\right.\right.$$

$$\left.\left.\times d\tau_{1:L}\right)dx_{1:L}\right]dr_{1:L}. \tag{33}$$

2) We use the random variable S_{K+1} to describe the number of photons in the complement pixel $C_{K+1} := \mathbb{R}^2 - \bigcup_{k=1}^K \bar{C}_k$ that result from the detection of the photons emitted from the object of interest. According to the definitions of $S_1,\ldots,S_{K+1}$, we have, for $z_1,\ldots,z_K = 0,1,\cdots,$

$$Pr[L=0]$$

$$= p_0 + \sum_{z=1}^\infty Pr\left[\bigcup_{v_{1:z}\in A_{K+1}^z(z_{1:K},z)}\left\{\bigcap_{l=1}^z (U(X(T_l)) \in C_{v_l})\right\}\right], \tag{34}$$

and

$$Pr[S_1 = z_1,\cdots,S_K = z_K]$$

$$= \sum_{z=0}^\infty Pr[S_1 = z_1,\cdots,S_K = z_K, S_{K+1} = z]$$

$$= \sum_{z=0}^\infty Pr\left[\bigcup_{v_{1:L+z}\in A_{K+1}^{L+z}(z_{1:K},z)}\left\{\bigcap_{l=1}^{L+z} (U(X(T_l)) \in C_{v_l})\right\}\right],$$

$L \geq 1,$

where the inner summation term can be calculated from the similar approach used in part 1.

D. Proof of Theorem 6.1

For a row parameter vector $\theta \in \Theta$, the Fisher information matrix $I_{N(t)=L}^f(\theta), L = 1,2,\cdots$, given $N(t) = L$, can be

calculated as, according to Eq. (23) of Definition 6.1,

$$
I^f_{N(t)=L}(\theta)
$$

$$
= \int_{\mathbb{R}^{2\times L}} \frac{1}{p^{\theta}_{R_{1:L}|N(t)}(r_{1:L}|L)} \left(\frac{\partial p^{\theta}_{R_{1:L}|N(t)}(r_{1:L}|L)}{\partial \theta} \right)^{T}
$$

$$
\times \left(\frac{\partial p^{\theta}_{R_{1:L}|N(t)}(r_{1:L}|L)}{\partial \theta} \right) dr_{1:L}. \tag{35}
$$

By substituting Eq. (6) into Eq. (35), we have

$$
I^f_{N(t)=L}(\theta)
$$

$$
= \int_{\mathbb{R}^{2\times L}} \frac{1}{\int_{\mathbb{R}^{3\times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L}}
$$

$$
\times \left(\frac{\partial}{\partial \theta} \int_{\mathbb{R}^{3\times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L} \right)^{T}
$$

$$
\times \left(\frac{\partial}{\partial \theta} \int_{\mathbb{R}^{3\times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L} \right)
$$

$$
\times dr_{1:L}
$$

$$
= \int_{\mathbb{R}^{2\times L}} \frac{DF_\theta^T (r_{1:L}) DF_\theta (r_{1:L})}{\int_{\mathbb{R}^{3\times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L}} dr_{1:L}, \tag{36}
$$

where DF_θ is given by Eq. (26).

E. Proof of Theorem 6.2

1) For a row parameter vector $\theta \in \Theta$, the Fisher information matrix $I^p(\theta)$ of the infinite practical data model can be calculated

$$
I^p(\theta) = \sum_{z_1=0}^{\infty} \cdots \sum_{z_K=0}^{\infty} \frac{1}{Pr^{\theta}[S_1 = z_1, \ldots, S_K = z_K]} \left(\frac{\partial Pr^{\theta}[S_1 = z_1, \ldots, S_K = z_K]}{\partial \theta} \right)^{T} \left(\frac{\partial Pr^{\theta}[S_1 = z_1, \ldots, S_K = z_K]}{\partial \theta} \right) \tag{37}
$$

$$
I^p(\theta) = \sum_{(z_1,\ldots,z_K)\in \mathbb{N}_0^K \setminus 0} p_L \left(\sum_{v_{1:L}\in A_K^L(z_{1:K})} \sum_{v'_{1:L}\in A_K^L(z_{1:K})} \int_{C_{v_{1:L}}} \int_{C_{v'_{1:L}}} \left\{ \int_{\mathbb{R}^{3\times L}} \int_{\mathbb{R}^{3\times L}} \left[\int_{\Delta^L} \int_{\Delta^L} \right. \right. \right.
$$

$$
\left. \left. \left. \frac{\left(\frac{\partial}{\partial \theta} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) \right)^{T} \left(\frac{\partial}{\partial \theta} F_\theta (r'_{1:L}, x'_{1:L}, \tau'_{1:L}) \right)}{\sum_{v_{1:L}\in A_K^L(z)} \int_{C_{v_{1:L}}} \left[\int_{\mathbb{R}^{3\times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L} \right] dr_{1:L}} d\tau_{1:L} d\tau'_{1:L} \right] dx_{1:L} dx'_{1:L} \right\} dr'_{1:L} dr_{1:L} \right)
$$

$$
= \sum_{L=1}^{\infty} p_L \sum_{z\in \mathbb{N}_0^K : |z|=L} \left(\frac{\sum_{v_{1:L}\in A_K^L(z)} \sum_{v'_{1:L}\in A_K^L(z)} \int_{C_{v_{1:L}}} \int_{C_{v'_{1:L}}} DF_\theta^T (r_{1:L}) DF_\theta (r'_{1:L}) dr'_{1:L} dr_{1:L}}{\sum_{v_{1:L}\in A_K^L(z)} \int_{C_{v_{1:L}}} \left[\int_{\mathbb{R}^{3\times L}} \left(\int_{\Delta^L} F_\theta (r_{1:L}, x_{1:L}, \tau_{1:L}) d\tau_{1:L} \right) dx_{1:L} \right] dr_{1:L}} \right) \tag{38}
$$

$$
I^p(\theta) := E_{Pr^{\theta}[\mathcal{I}_1=i_1,\ldots,\mathcal{I}_K=i_K]} \left\{ \left(\frac{\partial \log Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]}{\partial \theta} \right)^{T} \left(\frac{\partial \log Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]}{\partial \theta} \right) \right\}
$$

$$
= \int_{\mathbb{R}^K} Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K] \left(\frac{\partial \log Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]}{\partial \theta} \right)^{T} \left(\frac{\partial \log Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]}{\partial \theta} \right) di_{1:K}
$$

$$
= \int_{\mathbb{R}^K} \frac{1}{Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]} \left(\frac{\partial Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]}{\partial \theta} \right)^{T} \left(\frac{\partial Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]}{\partial \theta} \right) di_{1:K}
$$

$$
= \int_{\mathbb{R}^K} \frac{1}{Pr^{\theta}[\mathcal{I}_1 = i_1, \ldots, \mathcal{I}_K = i_K]} \left(\sum_{z_1,\ldots,z_K=0}^{\infty} p_{\mathcal{I}_{1:K}|S_{1:K}}(i_{1:K}|z_{1:K}) \frac{\partial Pr^{\theta}[S_1 = z_1, \ldots, S_K = z_K]}{\partial \theta} \right)^{T}
$$

$$
\times \left(\sum_{z'_1,\ldots,z'_K=0}^{\infty} p_{\mathcal{I}_{1:K}|S_{1:K}}(i_{1:K}|z'_{1:K}) \frac{\partial Pr^{\theta}[S_1 = z'_1, \ldots, S_K = z'_K]}{\partial \theta} \right) di_{1:K}
$$

$$
= \int_{\mathbb{R}^K} \frac{\sum_{z_1,\ldots,z_K=0}^{\infty} \sum_{z'_1,\ldots,z'_K=0}^{\infty} p_{\mathcal{I}_{1:K}|S_{1:K}}(i_{1:K}|z_{1:K}) p_{\mathcal{I}_{1:K}|S_{1:K}}(i_{1:K}|z'_{1:K})}{\sum_{z_1,\ldots,z_K=0}^{\infty} p_{\mathcal{I}_{1:K}|S_{1:K}}(i_{1:K}|z_{1:K}) Pr^{\theta}[S_1 = z_1, \ldots, S_K = z_K]}
$$

$$
\times \left(\frac{\partial Pr^{\theta}[S_1 = z_1, \ldots, S_K = z_K]}{\partial \theta} \right)^{T} \left(\frac{\partial Pr^{\theta}[S_1 = z'_1, \ldots, S_K = z'_K]}{\partial \theta} \right) di_{1:K} \tag{39}
$$

by Eq. (37), shown at the bottom of the previous page. By substituting Eq. (11) into Eq. (37), we have Eq. (38), shown at the bottom of the previous page, where DF_θ and F_θ are given by Eqs. (26) and (7).

2) It results from the similar approach used in part 1.

F. Effect of Noise on Fisher Information Matrix

In the following, we show the derivation process of Eq. (39), shown at the bottom of the previous page. By substituting Eq. (31) into the general equation of the Fisher information matrix, we then can obtain the Fisher information expression in terms of $Pr^\theta[S_1 = z_1, \ldots, S_K = z_K]$ by the process shown by Eq. (39).

REFERENCES

[1] W. E. Moerner and L. Kador, "Optical detection and spectroscopy of single molecules in a solid," *Phys. Rev. Lett.*, vol. 62, no. 21, pp. 2535–2538, 1989.

[2] M. J. Saxton and K. Jacobson, "Single-particle tracking: applications to membrane dynamics," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 26, no. 1, pp. 373–399, 1997.

[3] N. G. Walter, C. Y. Huang, A. J. Manzo, and M. A. Sobhy, "Do-it-yourself guide: How to use the modern single-molecule toolkit," *Nat. Methods*, vol. 5, no. 6, pp. 475–489, 2008.

[4] G. Seisenberger, M. U. Ried, T. Endress, H. Büning, M. Hallek, and C. Bräuchle, "Real-time single-molecule imaging of the infection pathway of an adeno-associated virus," *Science*, vol. 294, no. 5548, pp. 1929–1932, 2001.

[5] S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, "High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells," *Biophys. J.*, vol. 95, no. 12, pp. 6025–6043, 2008.

[6] A. Kusumi, T. A. Tsunoyama, K. M. Hirosawa, R. S. Kasai, and T. K. Fujiwara, "Tracking single molecules at work in living cells," *Nat. Chem. Biology*, vol. 10, no. 7, pp. 524–532, 2014.

[7] C. Manzo and M. F. Garcia-Parajo, "A review of progress in single particle tracking: From methods to biophysical insights," *Rep. Prog. Phys.*, vol. 78, no. 12, p. 124601, 2015.

[8] X. Michalet, "Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium," *Phys. Rev. E*, vol. 82, no. 4, p. 041914, 2010.

[9] A. J. Berglund, "Statistics of camera-based single-particle tracking," *Phys. Rev. E*, vol. 82, no. 1, p. 011917, 2010.

[10] X. Michalet and A. J. Berglund, "Optimal diffusion coefficient estimation in single-particle tracking," *Phys. Rev. E*, vol. 85, no. 6, p. 061916, 2012.

[11] P. K. Relich, M. J. Olah, P. J. Cutler, and K. A. Lidke, "Estimation of the diffusion constant from intermittent trajectories with variable position uncertainties," *Phys. Rev. E*, vol. 93, no. 4, p. 042401, 2016.

[12] C. P. Calderon, M. A. Thompson, J. M. Casolari, R. C. Paffenroth, and W. E. Moerner, "Quantifying transient 3D dynamical phenomena of single mRNA particles in live yeast cell measurements," *J. Phys. Chem. B*, vol. 117, no. 49, pp. 15 701–15 713, 2013.

[13] C. P. Calderon and K. Bloom, "Inferring latent states and refining force estimates via hierarchical Dirichlet process modeling in single particle tracking experiments," *PLoS ONE*, vol. 10, no. 9, p. e0137633, 2015.

[14] C. P. Calderon, "Motion blur filtering: A statistical approach for extracting confinement forces and diffusivity from a single blurred trajectory," *Phys. Rev. E*, vol. 93, no. 5, p. 053303, 2016.

[15] T. T. Ashley and S. B. Andersson, "Method for simultaneous localization and parameter estimation in particle tracking experiments," *Phys. Rev. E*, vol. 92, no. 5, p. 052707, 2015.

[16] M. R. Vahid, B. Hanzon, and R. J. Ober, "Fisher information matrix for single molecules with stochastic trajectories," *SIAM J. Imag. Sci.*, vol. 13, no. 1, pp. 234–264, 2020.

[17] Y. Wong, Z. Lin, and R. J. Ober, "Limit of the accuracy of parameter estimation for moving single molecules imaged by fluorescence microscopy," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 895–911, Mar. 2011.

[18] R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophys. J.*, vol. 8, no. 2, pp. 1185–1200, 2004.

[19] S. Ram, E. S. Ward, and R. J. Ober, "A stochastic analysis of performance limits for optical microscopes," *Multidimensional Syst. Signal Process.*, vol. 17, no. 1, pp. 27–57, 2006.

[20] M. Born and E. Wolf, "Principles of optics," *Camb. Univ.*, p. 1999.

[21] Z. Schuss, *Theory and Applications Stochastic Processes: An Analytical Approach*. Springer New York, 2009.

[22] V. Briane, C. Kervrann, and M. Vimond, "Statistical analysis of particle trajectories in living cells," *Phys. Rev. E*, vol. 97, no. 6, p. 062121, 2018.

[23] H. Risken and T. Frank, *The Fokker-Planck equation: methods of solution and applications*. Springer-Verlag Berlin Heidelberg, 1996.

[24] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[25] T. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, 2011.

[26] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Practice*. Springer-Verlag, 2001.

[27] G. Poyiadjis, A. Doucet, and S. S. Singh, "Particle approximations of the score and observed information matrix in state space models with application to parameter estimation," *Biometrika*, vol. 98, no. 1, pp. 65–80, 2011.

[28] A. V. Abraham, S. Ram, J. Chao, E. S. Ward, and R. J. Ober, "Quantitative study of single molecule location estimation techniques," *Opt. Exp.*, vol. 17, no. 26, pp. 23 352–23 373, 2009.

[29] M. R. Vahid, J. Chao, D. Kim, E. S. Ward, and R. J. Ober, "State space approach to single molecule localization in fluorescence microscopy," *Biomed. Opt. Exp.*, vol. 8, no. 3, pp. 1332–1355, 2017.

[30] M. R. Vahid, J. Chao, E. S. Ward, and R. J. Ober, "A state space based approach to localizing single molecules from multi-emitter images," in *Proc. SPIE*, p. 10070: 100700J.

[31] E. A. K. Cohen, D. Kim, and R. J. Ober, "Cramer-Rao lower bound for point based image registration with heteroscedastic error model for application in single molecule microscopy," *IEEE Trans. Med. Imag.*, vol. 34, no. 12, pp. 2632–2644, 2015.

[32] E. A. K. Cohen, A. V. Abraham, S. Ramakrishnan, and R. J. Ober, "Resolution limit of image analysis algorithms," *Nat. Commun.*, vol. 10, no. 1, pp. 1–12, 2019.

[33] A. Tahmasbi *et al.* "Designing the focal plane spacing for multifocal plane microscopy," *Opt. Exp.*, vol. 22, no. 14, pp. 16 706–16 721, 2014.

[34] M. A. Thompson, M. D. Lew, M. Badieirostami, and W. E. Moerner, "Localizing and tracking single nanoscale emitters in three dimensions with high spatiotemporal resolution using a double-helix point spread function," *Nano Lett.*, vol. 10, no. 1, pp. 211–218, 2012.

[35] A. Tahmasbi, E. S. Ward, and R. J. Ober, "Determination of localization accuracy based on experimentally acquired image sets: applications to single molecule microscopy," *Opt. Exp.*, vol. 23, no. 6, pp. 7630–7652, 2015.

[36] P. Petrov, Y. Shechtman, and W. E. Moerner, "Measurement-based estimation of global pupil functions in 3D localization microscopy," *Opt. Exp.*, vol. 25, no. 7, pp. 7945–7959, 2017.

[37] J. Chao, E. S. Ward, and R. J. Ober, "Fisher information theory for parameter estimation in single molecule microscopy: tutorial," *J. Opt. Soc. America A*, vol. 33, no. 7, pp. B36–B57, 2016.

[38] C. R. Rao, *Linear Statistical Inference and Its Appl.*. Wiley, 1965.

[39] S. M. Kay, *Fundamentals Statistical Signal Processing*. Prentice Hall, 1993.

[40] R. L. M. Peeters and B. Hanzon, "Symbolic computation of Fisher information matrices for parametrized state-space systems," *Automatica*, vol. 35, no. 6, pp. 1059–1071, 1999.

[41] J. Chao, E. S. Ward, and R. J. Ober, "Fisher information matrix for branching processes with application to electron-multiplying charge-coupled devices," *Multidimensional Syst. and Signal Process.*, vol. 23, pp. 349–379, 2012.

[42] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. Springer-Verlag New York, 1991.

Milad R. Vahid received the M.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2014 and the Ph.D. degree in biomedical engineering from Texas A&M University, College Station, TX, USA, in 2019. He is currently a Postdoctoral Researcher with the Department of Biomedical Data Science at Stanford University, Palo Alto, USA. His current research interests include biomedical signal and image processing, Biostatistics, and multiomics data analysis.

Bernard Hanzon received the Ph.D. degree on a thesis in the area of mathematical systems theory from Erasmus University Rotterdam, Rotterdam, The Netherlands, in 1986. He is currently a Professor in Mathematics at the School of Mathematical Sciences, University College Cork, Ireland. He studied econometrics with the Erasmus University Rotterdam and Mathematics and with the University of Amsterdam. In 1982, he visited the Department of Electrical Engineering, University of Texas at Austin. From 1984 to 1989, he was the University Lecturer with the Department of Mathematics, Technical University Delft. From 1989 to 2001, he was University Lecturer and later (from 1991 onwards) Senior Lecturer (universitair hoofddocent) with the Department of Econometrics and Operations Research of the Vrije Universtiteit Amsterdam. In the spring semester of 2002, he was Guest Professor with the Institute for Econometrics and Systems Theory, Technical University Vienna. He was an advisor with the Centre for Mathematics and Computer Science in Amsterdam and Guest Researcher with the Mathematical Institute, University of Leiden.

Raimund J. Ober (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Cambridge, U.K., in 1987. From 1987 to 1990, he was a Research Fellow at Girton College and the Engineering Department, University of Cambridge. In 1990, he joined the University of Texas at Dallas, Richardson, where he was a Professor with the Department of Electrical Engineering and the Department of Bioengineering. He was also Adjunct Professor at the University of Texas Southwestern Medical Center, Dallas. In 2014, he joined Texas A&M University, where he was a Professor with the Department of Biomedical Engineering and the Department of Molecular and Cellular Medicine. In 2018, he joined the University of Southampton, where he is currently a Professor with the Centre for Cancer Immunology, Faculty of Medicine. His research interests include the development of microscopy techniques for cellular investigations, in particular at the single molecule level, the study of cellular trafficking pathways using microscopy investigations, and signal/image processing of bioengineering data.