

# FISHER INFORMATION MATRIX FOR SINGLE MOLECULES WITH STOCHASTIC TRAJECTORIES \*

MILAD R. VAHID <sup>¶†‡</sup>, BERNARD HANZON<sup>§</sup>, AND RAIMUND J. OBER <sup>†¶||</sup>

**Abstract.** Tracking of objects in cellular environments has become a vital tool in molecular cell biology. A particularly important example is single molecule tracking which enables the study of the motion of a molecule in cellular environments by locating the molecule over time and provides quantitative information on the behavior of individual molecules in cellular environments, which were not available before through bulk studies. Here, we consider a dynamical system where the motion of an object is modeled by stochastic differential equations (SDEs), and measurements are the detected photons emitted by the moving fluorescently labeled object, which occur at discrete time points, corresponding to the arrival times of a Poisson process, in contrast to equidistant time points which have been commonly used in the modeling of dynamical systems. The measurements are distributed according to the optical diffraction theory, and therefore, they would be modeled by different distributions, e.g., an Airy profile for an in-focus and a Born and Wolf profile for an out-of-focus molecule with respect to the detector. For some special circumstances, Gaussian image models have been proposed. In this paper, we introduce a stochastic framework in which we calculate the maximum likelihood estimates of the biophysical parameters of the molecular interactions, e.g., diffusion and drift coefficients. More importantly, we develop 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. There exists no established method, even for Gaussian measurements, to systematically calculate the CRLB for the general motion model that we consider in this paper. We apply the developed methodology to simulated data of a molecule with linear trajectories and show that the standard deviation of the estimates matches well with the square root of the CRLB. We also show that equally sampled and Poisson distributed time points lead to significantly different Fisher information matrices.

**Key words.** Object tracking, Single molecule microscopy, Stochastic differential equation, Maximum likelihood estimation, Fisher information matrix, Cramér-Rao lower bound.

**AMS subject classifications.** 93B30, 62N02, 92C55

**1. Introduction.** The ability to track objects of interest, e.g., subcellular organelles and molecules, in cellular environments plays an important role in studying biological systems. In particular, single molecule tracking, which enables following subcellular processes at the single molecule level, has become a vital tool in cell biology [29, 28, 27]. Traditionally, microscopy studies were bulk studies and the information from such studies reflected the behavior of ensembles of molecules as opposed to individual ones [23]. Single molecule microscopy techniques have revolutionized the field of microscopy by providing quantitative information on the behavior of individual molecules in cellular environments, which were not available before through

---

\*Submitted to the editors February 2, 2019.

**Funding:** This work was supported in part by the National Institutes of Health (R01 GM085575).

<sup>†</sup>Department of Biomedical Engineering, Texas A&M University, College Station, TX 77843, USA ([milad.rafee@tamu.edu](mailto:milad.rafee@tamu.edu), [raimund.ober@tamu.edu](mailto:raimund.ober@tamu.edu)).

<sup>‡</sup>Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA ([miladrv@stanford.edu](mailto:miladrv@stanford.edu)).

<sup>§</sup>Edgeworth Centre for Financial Mathematics, School of Mathematical Sciences, University College Cork, Ireland ([b.hanzon@ucc.ie](mailto:b.hanzon@ucc.ie)).

<sup>¶</sup>Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, TX 77843, USA ([milad.rafee@tamu.edu](mailto:milad.rafee@tamu.edu), [raimund.ober@tamu.edu](mailto:raimund.ober@tamu.edu)).

<sup>||</sup>Center for Cancer Immunology, Faculty of Medicine, University of Southampton, Southampton, UK ([r.ober@soton.ac.uk](mailto:r.ober@soton.ac.uk)).

bulk studies [21, 24]. In biological studies, single molecule tracking methods have been used to study the intracellular trafficking of fluorescently labeled antibodies, e.g., prostate-specific membrane antigen (PSMA) antibodies [13, 12, 2], by analyzing the velocity and path of the fluorescent molecules.

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 [31, 7]. It has been shown that the motion of a moving object in such environments can be modeled by stochastic differential equations (SDEs) [26]. In particular, in many biological applications, solutions of linear SDEs are good fits to experimental single molecule trajectories [10, 9, 8]. In a basic fluorescence microscope, a fluorescently labeled object of interest is imaged by a detector which detects the photons emitted by the object during the acquisition time. Since the detection process of the emitted photons is inherently a random phenomenon, the acquired measurements are stochastic in nature. These measurements, according to the optical diffraction theory, can be modeled by different distributions. For example, a typical distribution for an in-focus molecule is an Airy profile [11], whereas, classical Born and Wolf profiles [6] are used instead for out-of-focus molecules. In some cases, it is possible and computationally beneficial to approximate these complex profiles with simple Gaussian models [1].

In many dynamical systems, the time points of the measurements are assumed to be equidistant. However, the time points of detection of the photons correspond to the arrival times of a Poisson process [21, 24]. This gives rise to non-uniform sampling of the continuous-time stochastic process that describes the motion of the object. Since the parameters of the motion model of the object are highly time-dependent, this randomized non-uniform sampling causes significant fluctuations in the values of these parameters.

In recent years, many methods have been developed to analyze the trajectories of a molecule in cellular environments. In most of these methods, the model for the motion of the molecule is assumed to be limited to a Brownian motion (pure diffusion) model described only by the diffusion coefficient, and only few of the available methods consider more general motion models. The methods developed to analyze pure diffusion models are mostly based on the mean square displacement approach [22], in which the diffusion coefficient is estimated by a linear regression of the mean square displacement of the Gaussian distributed observed locations of the molecule as a function of the time lag [5, 19, 18]. Mean square displacement-based methods are not the only approaches used to estimate the diffusion coefficient from a set of measurements. For example, Relich et al. [25] have proposed a method for the maximum likelihood estimation of the diffusion coefficient, with an information-based confidence interval, from Gaussian measurements. In all of these methods, the motion of a molecule is assumed as a pure diffusion model, and the measurements are modeled by independent and identically distributed Gaussian random variables [17].

However, in general, the motion of a molecule is not limited to the pure diffusion model, and the diffusion coefficient is only one of the parameters that play a role in the motion of the molecule. Also, the Gaussian assumption for the measurements is problematic in practice due to the fact that the Gaussian model is often not an accurate analytical model. In [3], Ashley and Andersson have proposed a simultaneous localization and parameter estimation algorithm for more complex motion models, such as confined [26] and tethered motions [20], which employs the expectation maximization algorithm in conjunction with sequential Monte Carlo methods [30]. For the general object tracking problem, in [16, 15], a sequential Monte Carlo method has been devel-

oped for the parameter estimation from nonlinear non-Gaussian state-space models. Briane et al. [7] have developed a method for classifying the object trajectories in living cells into three types of diffusion: Brownian motion, subdiffusion (diffusion in a closed domain or in a crowded area) and superdiffusion (diffusion in a specific direction). In [10, 9, 8], the motion of a moving object has been described more generally by a linear SDE, and the parameters of the model has been estimated using a maximum likelihood estimation method. However, they do not consider randomness of the time points at which the measurements occur. Their proposed framework also does not allow for non-Gaussian measurements.

In this paper, we address the above limitations by considering a more general dynamical system with arbitrary distributed measurements, which occur at Poisson distributed time points, that allows for more general motion models for an object of interest. Here, the motion of an object in cellular environments is modeled by stochastic differential equations, and the measurements are the detected photons emitted by the moving fluorescently labeled object. As mentioned earlier, these measurements can be modeled by non-Gaussian distributions. We develop a stochastic framework in which we calculate the maximum likelihood estimates of the biophysical parameters of the molecular interactions, e.g., diffusion and drift coefficients.

According to a well-known result from estimation theory, assuming that the estimator is unbiased, its standard deviation is then at best equal to the square root of the CRLB, which is given by the inverse of the Fisher information matrix [21, 24, 11]. More importantly, in order to evaluate the performance of our proposed estimation method, we develop a general framework to calculate the Fisher information matrix of the unknown parameters of the general motion model. There are some cases in which Gaussian approximations of measurements are very useful due to, for example, the ability of using computationally efficient algorithms in linear systems or the Kalman filter formulae. In particular, for Gaussian measurements, we calculate the Fisher information matrix by taking advantage of its relationship with the Kalman filter formula through a computationally efficient algorithm. To the best of our knowledge, even for Gaussian measurements, there currently exists no systematic methodology to evaluate the standard deviations of the estimates using the CRLB for the general motion model considered here.

To assess the performance of the proposed estimation method, we apply it to simulated data sets comprising linear trajectories of a molecule with Gaussian, Airy and classical model of Born and Wolf measurements. The results show that there is no systematic bias associated with the method. In addition, we show that the means of the distributions of the prediction of the molecule locations are able to follow the true locations of the molecule for the all different types of measurements. In particular, for data sets comprising repeat trajectories of a molecule with Gaussian measurements, it is shown that the standard deviations of the diffusion and drift estimates are close to the square roots of their corresponding CRLBs. We also show that, in case that we have one detected photon, the Fisher information matrices obtained for an Airy and its corresponding approximating Gaussian profile are different from each other, and therefore, the use of the Gaussian approximation can be problematic in some applications. We show that equally sampled time points, which have been commonly used in most dynamical systems, and Poisson distributed time points can lead to significantly different Fisher information matrices. We further show that even the results obtained for different realizations of a Poisson process can vary notably.

This paper is organized as follows. In Section 2, we present the statistical description of the acquired data, and derive a general formula for the likelihood function of

the described data model. Section 3 is devoted to introduce linear stochastic systems and calculate the likelihood function in case that the object is undergoing this type of trajectories. In Section 4, we propose a mathematical framework to calculate the maximum likelihood estimates of the parameters of interest, such as the parameters of the motion model of the molecule. Section 5 is devoted to calculate general expressions for the CRLB and Fisher information matrix relating to the parameter estimation problem.

In this paper, we use the following notation

$$(1.1) \quad \mathcal{C}^l \times \mathbb{R}_{[t]}^l := \{(r_1, \dots, r_l, \tau_1, \dots, \tau_l) \mid r_1, \dots, r_l \in \mathcal{C}, t_0 \leq \tau_1 < \dots < \tau_l \leq t\},$$

where  $\mathcal{C} := \mathbb{R}^2$ ,  $t_0 \in \mathbb{R}$ , and  $l = 1, 2, \dots$ . If there is no bound on  $\tau_l$ , we denote the set in Eq. (1.1) by  $\mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ .

**2. Fundamental data model.** A basic setup of an optical system considered here is shown in Fig. 1, where an object is in the object space and its image is captured by a planar detector in the image space. In the fundamental data model, we assume that the microscopy image data is acquired under ideal conditions. It assumes the use of an image detector that has an unpixelated photon detection area. The detection of a photon is intrinsically random in terms of both the time and the location on the detector at which the photon is detected. In general, the temporal part of the detection of the emitted photons can be modeled as a counting process  $\{N(\tau), \tau \geq t_0\}$ . Here, we assume that  $\{N(\tau), \tau \geq t_0\}$  is a Poisson process referred to as the *photon detection process* that is characterized by the intensity function  $\Lambda(\tau), \tau \geq t_0$ , referred to as the *photon detection rate*. The spatial component of the photon detection process is specified by random variables, referred to as the *photon location variables*, that describe the locations at which photons emitted by the object of interest are detected.

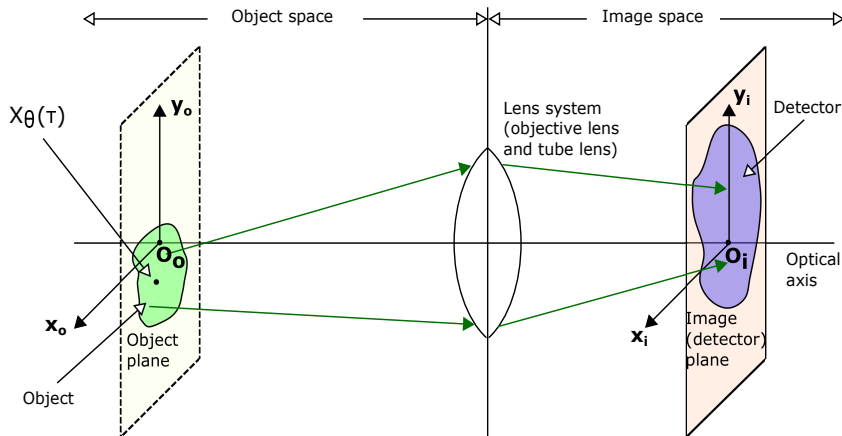


FIG. 1. Schematic of an optical microscope. An object located in the object (focal) plane is imaged by an optical lens system and the image of the object is acquired by the planar detector in the image space. A 3D random variable  $X_\theta(\tau), \tau \geq t_0$ , describes the location of the object in the object plane at time  $\tau$ .

In the following definition, we define a spatio-temporal process referred to as the *image detection process*, which models the acquired data, for two different acquisition methods, one when the time interval over which photons are detected is given and

the other when the total number of detected photons is given. For a fixed acquisition time, due to the stochastic nature of photon emission, the total number of detected photons varies for every image, while in the other case, the number of detected photons remains the same.

**DEFINITION 2.1.** *Let  $\mathcal{C} := \mathbb{R}^2$  denote a non-pixelated detector. Let  $\mathbb{R}^n, n = 1, 2, \dots$ , be the  $n$ -dimensional full parameter space. Let the parameter space  $\Theta$  describe an open subset of  $\mathbb{R}^n$  containing the true parameters. Elements in  $\Theta$  are described by a parameter vector  $\theta \in \Theta$ . Let the one-dimensional (1D) random variables  $T_1, T_2, \dots$ , describe the time points of detection of the photons that impact the detector  $\mathcal{C}$ , which are arrival time points associated with a Poisson process with intensity function  $\Lambda(\tau), \tau \geq t_0, t_0 \in \mathbb{R}$ . Let  $U_1, U_2, \dots$ , be 2D random variables that describe the locations of detection of the photons that impact the detector  $\mathcal{C}$ . For  $l = 1, 2, \dots$ , let  $\mathcal{U}_l := (U_1, \dots, U_l), \mathcal{U}_0 = \emptyset$ , and  $\mathcal{T}_l := (T_1, \dots, T_l), \mathcal{T}_0 = \emptyset$ . Assume that the current location of the detected photon, given the current and previous time points, is independent of the future time points, i.e., for  $r \in \mathcal{C}$  and  $t_0 \leq \tau_1 < \tau_2 < \dots$ ,*

$$p_{U_l | \mathcal{T}_k}(r | \tau_1, \dots, \tau_k) = p_{U_l | \mathcal{T}_l}(r | \tau_1, \dots, \tau_l), \quad \text{for all } k, l = 1, 2, \dots, \quad k \geq l,$$

where, for random vectors  $X$  and  $Y$ , the conditional probability density function of  $X$ , given  $Y$ , is denoted by  $p_{X|Y}$ . In other words, we assume that it may depend on past and current inputs but not future inputs. This assumption is natural in the context of the modeling of the dynamics of biomolecular processes such as the stochastic trajectory of a single molecule or organelle in a cellular context, where future effects do not impact the present.

1. For a fixed acquisition time interval  $[t_0, t]$ , an image detection process

$\mathcal{G}_{[t]} \left( (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \mathcal{C}, \Theta \right)$  for a time interval  $[t_0, t]$  is defined as a spatio-temporal process whose temporal part  $\mathcal{T}_{[t]}$  and spatial part  $\mathcal{U}_{[t]}$  describe the time points and the locations of detection of the photons that impact the detector  $\mathcal{C}$  in the time interval  $[t_0, t]$ , respectively, i.e., for  $\omega \in \Omega$ , where  $\Omega$  is the sample space,

$$U_{S_t(\omega)} = T_{S_t(\omega)} = \emptyset, \quad S_t(\omega) = 0,$$

and

$$\mathcal{T}_{[t]}(\omega) := (T_1(\omega), \dots, T_{S_t(\omega)}(\omega)), \quad \mathcal{U}_{[t]}(\omega) := (U_1(\omega), \dots, U_{S_t(\omega)}(\omega)), \quad S_t(\omega) > 0,$$

where  $t_0 \leq T_1(\omega) < \dots < T_{S_t(\omega)}(\omega) \leq t$ , and  $S_t$  is a discrete 1D random variable that takes its values in the non-negative integers such that  $T_{S_t(\omega)}(\omega) \leq t, T_{S_t(\omega)+1}(\omega) > t, S_t(\omega) > 0$ .

2. Given a fixed number  $L = 1, 2, \dots$ , of photons, an image detection process  $\mathcal{G}_L \left( (\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta \right)$  for a fixed number  $L$  of photons is defined as a spatio-temporal process whose temporal and spatial parts describe the time points and the locations of detection of the  $L$  photons that impact the detector  $\mathcal{C}$ , respectively. Moreover, given  $\mathcal{T}_L = (\tau_1, \dots, \tau_L), t_0 \leq \tau_1 < \tau_2 < \dots < \tau_L$ ,  $\mathcal{G}_{\tau_1, \dots, \tau_L} \left( (\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta \right)$  is referred to as the image detection process at fixed time points  $\tau_1, \dots, \tau_L$ .

In Theorem 2.2, we state expressions for the probability/probability density functions of image detection processes for a fixed time interval and for a fixed number of photons in terms of the conditional distributions of the locations of the detected photons, given the previous locations and the current and previous time points of the

detected photons. We further show that each of these conditional distributions can be expressed in terms of a scaled and shifted version of the image of the object and the distribution of the prediction of the object location, given the previous locations and time points of the detected photons. All the proofs in the paper are placed in the supplementary material. We drop the parameter vector  $\theta \in \Theta$ , when it is clear from the context.

**THEOREM 2.2.** *Let  $\mathcal{G}_{[t]}((\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \mathcal{C}, \Theta)$  and  $\mathcal{G}_L((\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta)$  be image detection processes for a time interval  $[t_0, t]$  and for a fixed number  $L$  of photons, respectively. Let  $\mathcal{D}_{[t]} := (\mathcal{U}_{[t]}, \mathcal{T}_{[t]})$ ,  $\mathcal{D}_k := (\mathcal{U}_k, \mathcal{T}_k)$ ,  $k = 0, 1, \dots$ .*

1. *Then, the probability of  $\mathcal{D}_{[t]} = \emptyset$  and  $N(t) = 0$  is given by*

$$P(\mathcal{D}_{[t]} = \emptyset, N(t) = 0) = e^{-\int_{t_0}^t \Lambda(\tau) d\tau},$$

and the probability density function  $p_{[t]}$  of  $\mathcal{D}_{[t]}$  and  $N(t)$  is given by

$$(2.1) \quad p_{[t]}(d_K, K) = e^{-\int_{t_0}^t \Lambda(\tau) d\tau} \prod_{k=1}^K \Lambda(\tau_k) \left[ \prod_{l=1}^K p_{U_l|T_l, \mathcal{D}_{l-1}}(r_l|\tau_l, d_{l-1}) \right],$$

where  $d_K \in \mathcal{C}^K \times \mathbb{R}_{[t]}^K$ ,  $K = 1, 2, \dots$ , and  $p_{U_l|T_l, \mathcal{D}_{l-1}}$  denotes the conditional probability density function of  $U_l$ , given  $T_l, \mathcal{D}_{l-1}$ , with  $p_{U_1|T_1, \mathcal{D}_0}(r_1|\tau_1, d_0) := p_{U_1|T_1}(r_1|\tau_1)$ .

2. *Moreover, the probability density function  $p_L$  of  $\mathcal{D}_L$  is given by*

$$(2.2) \quad p_L(d_L) = e^{-\int_{t_0}^{\tau_L} \Lambda(\tau) d\tau} \prod_{k=1}^L \Lambda(\tau_k) \left[ \prod_{l=1}^L p_{U_l|T_l, \mathcal{D}_{l-1}}(r_l|\tau_l, d_{l-1}) \right], \quad d_L \in \mathcal{C}^L \times \mathbb{R}_{[\infty]}^L.$$

*Proof.* See Section **SM1** in the supplementary material.  $\square$

Note that, as can be seen in the above theorem, the probability density function of an image detection process for a time interval  $[t_0, t]$  depends on the integral of the photon detection rate  $\Lambda(\tau)$ ,  $\tau \geq t_0$ , over the time interval  $[t_0, t]$ , and the probability density function of an image detection process for a fixed number  $L$  of photons depends on the integral of the photon detection rate over the time interval  $[t_0, \tau_L]$ , where  $\tau_L$  denotes the time point of the  $L^{\text{th}}$  (last) detected photon.

The probability density function of the location at which a photon emitted by the object of interest is detected, is referred to as the *image profile* of the object. So far we have made no assumptions about the specific functional form of the image profile of the object. In many practical cases, the image profile can be described as a scaled and shifted version of the image function. In such cases, an image function describes the image of an object on the detector plane at unit lateral magnification. Also, in general, the trajectory of the object can be described by a random process. In the following definition, we define image detection processes driven by a stochastic trajectory of the object and the image function for a fixed time interval and for a fixed number of photons.

**DEFINITION 2.3.** *Let  $\mathcal{G}_{[t]}((\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \mathcal{C}, \Theta)$  and  $\mathcal{G}_L((\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta)$  be image detection processes for a time interval  $[t_0, t]$  and for a fixed number  $L$  of photons, respectively. Let  $X(\tau)$ ,  $\tau \geq t_0$ , denote a 3D random process that describes the 3D stochastic trajectory of the object. Also, let  $\{f_x\}_{x \in \mathbb{R}^3}$  defined on the detector  $\mathcal{C}$ , be a family of image profiles of an object located at  $x \in \mathbb{R}^3$  in the object space. Assume that the current location of the detected photon, given the current location of the object, is*

independent of the previous locations and time points of the detected photons, i.e., for all  $x \in \mathbb{R}^3$ ,

$$p_{U_l|X(T_l), T_l, \mathcal{D}_{l-1}}(r_l|x, \tau_l, d_{l-1}) = p_{U_l|X(\tau_l)}(r_l|x) := f_x(r_l), \quad r_l \in \mathcal{C},$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$  for  $\mathcal{G}_{[t]}$ ,  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$  for  $\mathcal{G}_L$ ,  $p_{U_l|X(T_l), T_l, \mathcal{D}_{l-1}}$  is the conditional probability density function of  $U_l$ , given  $X(T_l), T_l, \mathcal{D}_{l-1}$ , and  $p_{U_l|X(\tau_l)}$  denotes the conditional probability density function of  $U_l$ , given  $X(\tau_l)$ . This assumption is justified as the process of the image formation, photon emission etc. only depends on the position of the emitting fluorescent object at the particular point in time and not on prior events.

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  $x := (x_0, y_0, z_0) \in \mathbb{R}^3$ ,

$$(2.3) \quad f_x(r) := \frac{1}{|\det(M)|} q_{z_0}(M^{-1}r - (x_0, y_0)^T), \quad r \in \mathcal{C}.$$

In the above equation,  $q_{z_0}$ , which is referred to as the image function, is a function that describes, at unit lateral magnification, the image of the object in the detector plane when the object is located at  $(0, 0, z_0)$  in the object space.

Image detection processes  $\mathcal{G}_{[t]}(X, (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), q, \mathcal{C}, \Theta)$  and  $\mathcal{G}_L(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta)$  driven by the stochastic trajectory  $X$  and image function  $q$  for a time interval  $[t_0, t]$  and for a fixed number  $L$  of photons are defined as the spatio-temporal processes  $\mathcal{G}_{[t]}$  and  $\mathcal{G}_L$ , respectively.

In the classical case of a measurement error, the image function  $q_{z_0}$  is defined as a function of  $(r - M(x_0, y_0)^T)$ , which is the deviation between two locations in the image space. Here, however, in order to be consistent with our previous framework developed for a static object,  $q_{z_0}$  is defined as a function of  $(M^{-1}r - (x_0, y_0)^T)$ , which is the difference between two points in the object space.

We next illustrate specific image functions that describe the image of a point source. According to the optical diffraction theory, when a point source is in-focus with respect to the detector, the intensity distribution of the image of the point source is described by an Airy profile given by [24] (see Fig. 2(a))

$$(2.4) \quad q(x, y) = \frac{J_1^2\left(\frac{2\pi n_a}{\lambda} \sqrt{x^2 + y^2}\right)}{\pi(x^2 + y^2)}, \quad (x, y) \in \mathbb{R}^2,$$

where  $n_a$  denotes the numerical aperture of the objective lens,  $\lambda$  denotes the emission wavelength of the molecule, and  $J_1$  denotes the first order Bessel function of the first kind. The 2D Gaussian profile, on the other hand, which has been widely used to approximate the Airy profile, is given by

$$(2.5) \quad q(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\left(\frac{x^2+y^2}{\sigma^2}\right)}, \quad (x, y) \in \mathbb{R}^2,$$

where  $\sigma > 0$ .

For an out-of-focus point source, the image function can be obtained by the classical Born and Wolf model given by [6]

$$(2.6) \quad q_{z_0}(x, y) = \frac{4\pi n_a^2}{\lambda^2} \left| \int_0^1 J_0\left(\frac{2\pi n_a}{\lambda} \sqrt{x^2 + y^2} \rho\right) e^{\frac{j\pi n_a^2 z_0}{n_a \lambda} \rho^2} \rho d\rho \right|^2, \quad (x, y) \in \mathbb{R}^2,$$



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. When the point source is in-focus with respect to the detector, i.e., it lies in the object plane, then  $z_0 = 0$  and Eqs. (2.4) and (2.6) are equivalent.

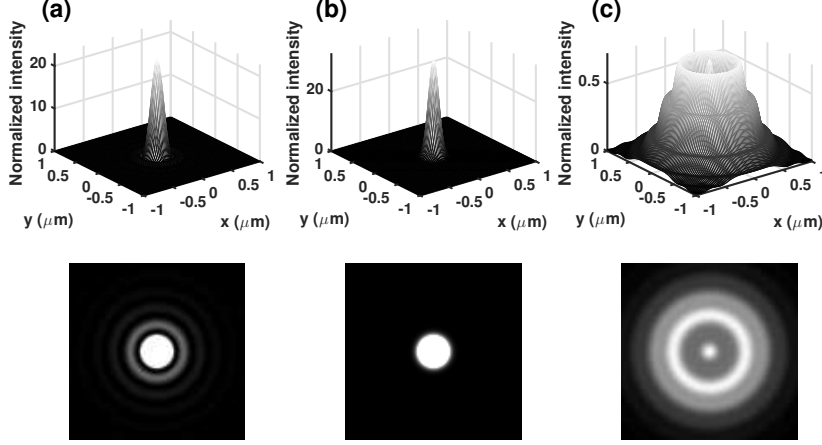


FIG. 2. *Image function examples. (a) Airy and (b) symmetric Gaussian profiles, which describe the images of an in-focus point source, simulated by Eqs. (2.4) and (2.5), respectively, with the parameters given in Section 4.1. (c) Born and Wolf profile simulated by Eq. (2.6) with the out-of-focus level  $z_0 = 1 \mu\text{m}$ , and the parameters given in Section 4.1.*

We calculate  $p_{U_l|T_l, \mathcal{D}_{l-1}}, l = 1, 2, \dots$ , for more general cases. In the following corollary to Theorem 2.2, by describing these conditional probability density functions in terms of the image function, we derive expressions for the probability density functions of the image detection processes driven by the stochastic trajectory  $X$  and image function  $q$  for a time interval  $[t_0, t]$  and for a fixed number  $L$  of photons.

COROLLARY 2.4. *Let  $\mathcal{G}_{[t]}(X, (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), q, \mathcal{C}, \Theta)$  (or  $\mathcal{G}_L(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta)$ ) be an image detection process driven by the stochastic trajectory  $X$  and image function  $q$  for a time interval  $[t_0, t]$  (or for a fixed number  $L$  of photons). Then, the conditional probability density function  $p_{U_l|T_l, \mathcal{D}_{l-1}}, l = 1, 2, \dots$ , in Eq. (2.1) (or in Eq. (2.2)) of Theorem 2.2 is given by, for  $x := (x_0, y_0, z_0) \in \mathbb{R}^3$ ,*

$$\begin{aligned}
 p_{U_l|T_l, \mathcal{D}_{l-1}}(r_l|\tau_l, d_{l-1}) &= \int_{\mathbb{R}^3} f_x(r_l) p_{pr_l}(x|\tau_l, d_{l-1}) dx \\
 (2.7) \qquad \qquad \qquad &= \frac{1}{|\det(M)|} \int_{\mathbb{R}^3} q_{z_0}(M^{-1}r_l - (x_0, y_0)) p_{pr_l}(x|\tau_l, d_{l-1}) dx,
 \end{aligned}$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$  (or  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ ),  $p_{pr_l} := p_{X(T_l)|T_l, \mathcal{D}_{l-1}}$  denotes the distribution of the prediction of the object location,  $p_{pr_1}(x|\tau_1, d_0) := p_{pr_1}(x|\tau_1)$ , and  $f_x, x \in \mathbb{R}^3$ , is the image profile of an object located at  $x$  in the object space.

*Proof.* See Section SM2 in the supplementary material.  $\square$



As can be seen in the above corollary, the expression of the probability density function of the image detection process depends on the distribution  $p_{pr_l}, l = 1, 2, \dots$ , of the prediction of the object location, given the previous locations of the detected photons and the current and previous time points. In the following section, we introduce linear stochastic systems and calculate  $p_{pr_l}, l = 1, 2, \dots$ , for them.

In Theorem 2.2, we expressed the probability density functions of image detection processes in terms of conditional probability densities  $p_{U_i|T_i, \mathcal{D}_{l-1}}, l = 1, 2, \dots$ , of the locations of the detected photons, given the previous locations and the current and previous time points of the detected photons. In particular, for an object with a deterministic trajectory or a static object, the conditional probability densities  $p_{U_i|T_i, \mathcal{D}_{l-1}}, l = 1, 2, \dots$ , are given as follows. For an object with deterministic trajectory  $X(\tau) \in \mathbb{R}^3, \tau \geq t_0$ , we have

$$(2.8) \quad p_{U_i|T_i, \mathcal{D}_{l-1}}(r_l|\tau_l, d_{l-1}) = p_{U_i|T_i}(r_l|\tau_l) := f_{X(\tau_l)}(r_l).$$

Also, for a static object with position  $X_0 \in \mathbb{R}^3$ , we have

$$(2.9) \quad p_{U_i|T_i, \mathcal{D}_{l-1}}(r_l|\tau_l, d_{l-1}) = p_{U_i}(r_l) := f_{X_0}(r_l).$$

**3. 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 [31, 7]. 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

$$(3.1) \quad \tilde{X}(\tau_{l+1}) = \tilde{\phi}(\tau_l, \tau_{l+1})\tilde{X}(\tau_l) + \tilde{W}(\tau_l, \tau_{l+1}), \quad \tau_0 := t_0 \leq \tau_1 < \dots < \tau_{l+1} < \dots,$$

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  $\{\tilde{W}(\tau_l, \tau_{l+1}), l = 1, 2, \dots\}$  is a sequence of  $k$ -dimensional random variables with probability density functions  $p_{\tilde{W}(\tau_l, \tau_{l+1})}$ . 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 general system of discrete evolution equations described by Eq. (3.1) can arise, for example, from stochastic differential equations [26]. In particular, in many biological applications, solutions of linear stochastic differential equations are good fits to experimental single-molecule trajectories [26]. 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 [8]

$$(3.2) \quad dX(\tau) = (V + F(\tau)X(\tau))d\tau + G(\tau)dB(\tau), \quad \tau \geq t_0,$$

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 a random process [4].

Here, we assume that  $\{B(\tau) \in \mathbb{R}^r, \tau \geq t_0\}$  is an  $r$ -vector Brownian motion 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 [10, 9, 8]. Then, the solution of Eq. (3.2) at discrete time points  $\tau_0 := t_0 \leq \tau_1 < \dots < \tau_{l+1} < \dots$  is given by [14]

$$(3.3) \quad X(\tau_{l+1}) = \phi(\tau_l, \tau_{l+1})X(\tau_l) + a(\tau_l, \tau_{l+1}) + W_g(\tau_l, \tau_{l+1}),$$

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

$$\begin{aligned} \frac{d\phi(t, \tau)}{dt} &= F(t)\phi(t, \tau), \quad \phi(\tau, \tau) = I_{3 \times 3}, \quad \text{for all } t, \tau \geq t_0, \\ \phi(t, \tau)\phi(\tau, \psi) &= \phi(t, \psi), \quad \text{for all } t, \tau, \psi \geq t_0, \end{aligned}$$

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

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

Also, in this case,  $\left\{ W_g(\tau_l, \tau_{l+1}) := \int_{\tau_l}^{\tau_{l+1}} \phi(\tau, \tau_{l+1})G(\tau)dB(\tau), l = 1, 2, \dots \right\}$  is a zero mean white Gaussian sequence with covariance  $Q_g(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3 \times 3}$  given by

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

By letting  $X(\tau) = H\tilde{X}(\tau) = I_{3 \times 3}\tilde{X}(\tau) = \tilde{X}(\tau), \tau \geq t_0$ , and  $\phi(\tau_l, \tau_{l+1}) = \tilde{\phi}(\tau_l, \tau_{l+1})$ , we obtain expressions of the form of Eq. (3.1), where we assume that

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

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

As an another example, for pure diffusion motion, when  $V$  and  $F(\tau), \tau \geq 0$ , in Eq. (3.2) are equal to zero, the discrete motion model is given by

$$(3.4) \quad X(\tau_{l+1}) = X(\tau_l) + W_g(\tau_l, \tau_{l+1}), \quad \tau_0 := t_0 \leq \tau_1 < \dots < \tau_{l+1} < \dots.$$

Setting  $\tilde{X}(\tau) := X(\tau), \tau \geq t_0$ , with  $H$  the identity matrix,  $\tilde{\phi}(\tau_l, \tau_{l+1}) = \phi(\tau_l, \tau_{l+1}) = I_{3 \times 3}$ , and  $\tilde{W}(\tau_l, \tau_{l+1}) = W_g(\tau_l, \tau_{l+1})$ , we again obtain expressions of the form of Eq. (3.1).

The above discussion motivates us to model the motion of the object, in the following definition, by Eq. (3.1) with, in general, an arbitrary distributed process noise  $\tilde{W}$ . In particular, we also consider the special case of Gaussian distributed process noise  $\tilde{W}_g$ , separately.

**DEFINITION 3.1.** Let  $\mathcal{G}_{[t]}(X, (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), q, \mathcal{C}, \Theta)$  and  $\mathcal{G}_L(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta)$  be image detection processes driven by a stochastic trajectory  $\tilde{X}$  and image function  $q$  for a fixed time interval  $[t_0, t]$  and for a fixed number  $L$  of photons. Let  $p_{X(t_0)}$  be the probability density function of the initial location  $X(t_0)$  of the object. We assume that

a. the motion of the object is modeled through a general state space system with state  $\tilde{X}(\tau) \in \mathbb{R}^k, \tau \geq t_0$ , as

$$(3.5) \quad \tilde{X}(\tau_{l+1}) = \tilde{\phi}(\tau_l, \tau_{l+1})\tilde{X}(\tau_l) + \tilde{W}(\tau_l, \tau_{l+1}), \quad \tau_0 := t_0 \leq \tau_1 < \dots < \tau_{l+1} < \dots,$$

where we assume that there exists a matrix  $H \in \mathbb{R}^{3 \times k}$  such that  $X(\tau) = H\tilde{X}(\tau), \tau \geq t_0$ ,  $\tilde{\phi}(\tau_l, \tau_{l+1}) \in \tilde{\Phi}$ , where  $\tilde{\Phi} = \left\{ \tilde{\phi}(\tau, \psi) \right\}_{\psi > \tau \geq t_0}$  is a family of  $k \times k$  invertible real-valued state-transition matrices, and  $\left\{ \tilde{W}(\tau_l, \tau_{l+1}), l = 0, 1, 2, \dots \right\}$  is a process noise sequence of independent  $k$ -dimensional random variables with probability density functions  $p_{\tilde{W}(\tau_l, \tau_{l+1})}$ .

b. We assume that

$$(3.6) \quad U_l = Z(X(\tau_l)), \quad l = 1, 2, \dots,$$

where  $\{Z(X(\tau_l)), l = 1, 2, \dots\}$  is a measurement sequence of independent 2D random variables with probability density functions  $p_{Z(X(\tau_l))} = f_{X(\tau_l)}$ , where  $Z$  is a random function that maps the object space into the image space,  $f_{X(\tau_l)}$  is the image profile of an object located at  $X(\tau_l)$  defined in Definition 2.3 and

c. We assume that the sequences  $\{\tilde{W}(\tau_l, \tau_{l+1}), l = 0, 1, \dots\}$ ,  $\{Z(X(\tau_l)), l = 1, 2, \dots\}$ , and  $\tilde{X}(t_0)$  are independent of one another.

The image detection process  $\mathcal{G}_{[t]}(X, (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), q, \mathcal{C}, \Theta)$  (or  $\mathcal{G}_L(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta)$ ) with the additional properties (a)-(c) is called an image detection process with expanded state space  $\tilde{X}$  for a time interval  $[t_0, t]$  (or for a fixed number  $L$  of photons), and is denoted by  $\mathcal{G}_{[t]}(\tilde{X}, H, \tilde{W}, Z), (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \tilde{\Phi}, \mathcal{C}, \Theta)$  (or  $\mathcal{G}_L(\tilde{X}, H, \tilde{W}, Z), (\mathcal{U}_L, \mathcal{T}_L), \tilde{\Phi}, \mathcal{C}, \Theta)$ ).

We further assume that

$\alpha$ .  $\{\tilde{W}_g(\tau_l, \tau_{l+1}) := \tilde{W}(\tau_l, \tau_{l+1}), l = 0, 1, \dots\}$  is a white Gaussian sequence with mean  $\tilde{a}(\tau_l, \tau_{l+1}) \in \mathbb{R}^k$  and covariance matrix  $\tilde{Q}_g(\tau_l, \tau_{l+1}) \in \mathbb{R}^{k \times k}$ ,  $\tilde{Q}_g(\tau_l, \tau_{l+1}) > 0$ ,  $\beta$ .

$$(3.7) \quad Z(X(\tau_l)) = M'X(\tau_l) + Z_{g,l} \quad l = 1, 2, \dots,$$

where  $M' := [M \quad 0_{2 \times 1}] \in \mathbb{R}^{2 \times 3}$ , in which  $M \in \mathbb{R}^{2 \times 2}$  is an invertible magnification matrix used in the definition of the image function (Eq. (2.3)), where  $0_{2 \times 1}$  is the  $2 \times 1$  zero matrix, and  $\{Z_{g,l}, l = 1, 2, \dots\}$  is a measurement noise sequence of independent 2D Gaussian random variables with mean zero and the same covariance matrix  $\Sigma_g \in \mathbb{R}^{2 \times 2}$ ,  $\Sigma_g > 0$ .

$\gamma$ . We assume that the initial state  $\tilde{X}(t_0)$  is Gaussian distributed with mean  $\tilde{x}_0 \in \mathbb{R}^k$  and covariance matrix  $\tilde{P}_0 \in \mathbb{R}^{k \times k}$ ,  $\tilde{P}_0 > 0$ .

If, in addition, an image detection process with expanded state space has the properties ( $\alpha$ )-( $\gamma$ ), it is called an image detection process with expanded state space  $\tilde{X}$  and Gaussian process and measurement noise models, and is denoted by  $\mathcal{G}_{[t]}^g(\tilde{X}, H, \tilde{W}_g, Z_g), (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \tilde{\Phi}, M', \mathcal{C}, \Theta)$  (or  $\mathcal{G}_L^g(\tilde{X}, H, \tilde{W}_g, Z_g), (\mathcal{U}_L, \mathcal{T}_L), \tilde{\Phi}, M', \mathcal{C}, \Theta)$ ) for a time interval  $[t_0, t]$  (or for a fixed number  $L$  of photons).

In Corollary 2.4, we calculated the probability density function of the image detection process in terms of the image function  $q$  and the distribution  $p_{pr_l}, l = 1, 2, \dots$ , of the prediction of the object location, given the previous locations of the detected photons and the current and previous time points. In the following theorem, for a linear stochastic system and Gaussian process and measurement noise, we calculate these distributions using the Kalman filter formulae. Also, for a more general Markov motion model described by a first order system with arbitrary distributed process and measurement noise, we calculate these distributions recursively.

**THEOREM 3.2.** Let  $\mathcal{G}_{[t]}(\tilde{X}, H, \tilde{W}, Z), (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \tilde{\Phi}, \mathcal{C}, \Theta)$  (or  $\mathcal{G}_L(\tilde{X}, H, \tilde{W}, Z), (\mathcal{U}_L, \mathcal{T}_L), \tilde{\Phi}, \mathcal{C}, \Theta)$ ) be an image detection process with expanded state space  $\tilde{X}$  for

a time interval  $[t_0, t]$  (or for a fixed number  $L$  of photons). Let  $\mathcal{D}_k := (\mathcal{U}_k, \mathcal{T}_k)$ ,  $k = 0, 1, \dots$ , and

$$p_{pr_l}(x|\tau_l, d_{l-1}) := p_{X(T_l)|T_l, \mathcal{D}_{l-1}}(x|\tau_l, d_{l-1}), \quad x \in \mathbb{R}^3,$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$  (or  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ ), be the probability density function of the prediction of the object location, and  $p_{pr_1}(x|\tau_1, d_0) := p_{pr_1}(x|\tau_1)$ .

1. Assume that there exist non-singular matrix  $H_1 \in \mathbb{R}^{3 \times 3}$  and matrix  $H_2 \in \mathbb{R}^{3 \times (k-3)}$  such that  $H = [H_1 \ H_2]$ . Let

$$S := \begin{bmatrix} H_1 & H_2 \\ 0_{(k-3) \times 3} & I_{(k-3) \times (k-3)} \end{bmatrix} \in \mathbb{R}^{k \times k}.$$

Then, for  $x := (x_1, x_2, x_3)^T \in \mathbb{R}^3$  and  $\bar{x} := (x_1, x_2, x_3, \bar{x}_4, \dots, \bar{x}_k)^T \in \mathbb{R}^k$ ,

$$p_{pr_l}(x|\tau_l, d_{l-1}) = \int_{\mathbb{R}^{k-3}} \tilde{p}_{pr_l}(S^{-1}\bar{x}|\tau_l, d_{l-1}) |H_1|^{-1} d\bar{x}_4 \cdots d\bar{x}_k,$$

where  $\tilde{p}_{pr_l} := p_{\tilde{X}(T_l)|T_l, \mathcal{D}_{l-1}}$ ,  $l = 0, 1, 2, \dots$ , and  $S^{-1}$  is given by

$$S^{-1} = \begin{bmatrix} H_1^{-1} & -H_1^{-1}H_2 \\ 0_{(k-3) \times 3} & I_{(k-3) \times (k-3)} \end{bmatrix}.$$

If  $H = [I_{3 \times 3} \ 0_{3 \times (k-3)}]$ , then,

$$p_{pr_l}(x|\tau_l, d_{l-1}) = \int_{\mathbb{R}^{k-3}} \tilde{p}_{pr_l}(\bar{x}|\tau_l, d_{l-1}) d\bar{x}_4 \cdots d\bar{x}_k.$$

2. The probability density function  $\tilde{p}_{pr_l}$ ,  $l = 0, 1, 2, \dots$ , can be calculated through the following recursive formula, for  $\bar{x} \in \mathbb{R}^k$ ,

$$(3.8) \quad \tilde{p}_{pr_{l+1}}(\bar{x}|\tau_{l+1}, d_l) = \frac{1}{|\det(\phi(\tau_l, \tau_{l+1}))|} \int_{\mathbb{R}^k} \tilde{p}_{f_{i_l}}(\tilde{\phi}^{-1}(\tau_l, \tau_{l+1})\bar{x}_o|d_l) p_{\tilde{W}(\tau_l, \tau_{l+1})}(\bar{x} - \bar{x}_o) d\bar{x}_o,$$

where  $d_0 = \emptyset$ , and the distribution  $\tilde{p}_{f_{i_l}}(\bar{x}|d_l) := p_{\tilde{X}(T_l)|\mathcal{D}_l}(\bar{x}|d_l)$  of the filtered object location is given by

$$(3.9) \quad \tilde{p}_{f_{i_l}}(\bar{x}|d_l) = \frac{p_{Z(H\bar{x})}(r_l) \tilde{p}_{pr_l}(\bar{x}|\tau_l, d_{l-1})}{\int_{\mathbb{R}^k} p_{Z(H\bar{x}_o)}(r_l) \tilde{p}_{pr_l}(\bar{x}_o|\tau_l, d_{l-1}) d\bar{x}_o}.$$

3.1. Let  $\mathcal{G}_{[t]}^g \left( (\tilde{X}, H, \tilde{W}_g, Z_g), (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$  (or  $\mathcal{G}_L^g \left( (\tilde{X}, H, \tilde{W}_g, Z_g), (\mathcal{U}_L, \mathcal{T}_L), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$ ) be an image detection process with expanded state space  $\tilde{X}$  and Gaussian process and measurement noise models for a time interval  $[t_0, t]$  (or for a fixed number  $L$  of photons). Let  $C := M'H$ . Then, for  $l = 0, 1, \dots$ , and  $\bar{x} \in \mathbb{R}^k$ ,

$$(3.10) \quad \tilde{p}_{pr_{l+1}}(\bar{x}|d_l, \tau_{l+1}) = \frac{1}{(2\pi)^{k/2} [\det(P_{l+1}^l)]^{1/2}} \exp \left( -\frac{1}{2} (\bar{x} - \hat{x}_{l+1}^l)^T (P_{l+1}^l)^{-1} (\bar{x} - \hat{x}_{l+1}^l) \right),$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$  (or  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ ),  $\hat{x}_1^0 = \tilde{\phi}(\tau_0, \tau_1)\bar{x}_0$ ,  $P_1^0 = \tilde{\phi}(\tau_0, \tau_1)\tilde{P}_0\tilde{\phi}^T(\tau_0, \tau_1) + \tilde{Q}_g(\tau_0, \tau_1)$ , and for  $l = 1, 2, \dots$ ,

$$(3.11) \quad \begin{aligned} \hat{x}_{l+1}^l &= \tilde{\phi}(\tau_l, \tau_{l+1})\hat{x}_l^l + \tilde{a}(\tau_l, \tau_{l+1}), \\ P_{l+1}^l &= \tilde{\phi}(\tau_l, \tau_{l+1})P_l^l\tilde{\phi}^T(\tau_l, \tau_{l+1}) + \tilde{Q}_g(\tau_l, \tau_{l+1}), \end{aligned}$$

with

$$(3.12) \quad \begin{aligned} K_l &= P_l^{l-1}C^T \left( CP_l^{l-1}C^T + \Sigma_g \right)^{-1}, \\ \hat{x}_l^l &= \hat{x}_l^{l-1} + K_l(r_l - C\hat{x}_l^{l-1}), \\ P_l^l &= P_l^{l-1} - K_lCP_l^{l-1}. \end{aligned}$$

3.2. Moreover, the conditional probability density function  $p_{U_l|T_l, \mathcal{D}_{l-1}}$  is given by

$$(3.13) \quad p_{U_l|T_l, \mathcal{D}_{l-1}}(r_l|\tau_l, d_{l-1}) = \frac{1}{2\pi [\det(R_l)]^{1/2}} \exp\left(-\frac{1}{2}(r_l - \hat{r}_l)^T R_l^{-1}(r_l - \hat{r}_l)\right),$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$  (or  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ ),  $R_l := CP_l^{l-1}C^T + \Sigma_g$  and  $\hat{r}_l := C\hat{x}_l^{l-1}$ .

*Proof.* See Section SM3 in the supplementary material.  $\square$

**4. Maximum likelihood estimation.** The main purpose of the presented materials in the previous section is to provide a mathematical framework to estimate the parameters of interest, such as the parameters of the model that describes the motion of a moving object with stochastic trajectories, from the acquired data. In this paper, we use the maximum likelihood estimation approach as follows. For a general parameter estimation problem, denoting the acquired data by  $\bar{d} \in \mathbb{R}^m$ ,  $m = 1, 2, \dots$ , the maximum likelihood estimate  $\hat{\theta}_{mle}$  of  $\theta \in \Theta$ , if it exists, is given by

$$\hat{\theta}_{mle} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left( -\log \mathcal{L}(\theta|\bar{d}) \right),$$

where  $\mathcal{L}$  denotes the likelihood function. In our specific problem, the acquired data for the fixed time interval  $[t_0, t]$  acquisition case is denoted by  $\bar{d}_K \in \mathcal{C}^K \times \mathbb{R}_{[t]}^K$ ,  $K = 0, 1, \dots$ . Then, the likelihood function  $\mathcal{L}_{[t]}$  of  $\mathcal{G}_{[t]}((\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \mathcal{C}, \Theta)$  is given by, according to Theorem 2.2 (see also [32, 33]), for  $\theta \in \Theta$ ,

$$(4.1) \quad \mathcal{L}_{[t]}(\theta|\bar{d}_K) = \begin{cases} e^{-\int_{t_0}^t \Lambda_\theta(\tau) d\tau}, & K = 0, \\ e^{-\int_{t_0}^t \Lambda_\theta(\tau) d\tau} \prod_{k=1}^K \Lambda_\theta(\bar{\tau}_k) \left[ \prod_{l=1}^K p_{U_l|T_l, \mathcal{D}_{l-1}}^\theta(\bar{r}_l|\bar{\tau}_l, \bar{d}_{l-1}) \right], & K = 1, 2, \dots, \end{cases}$$

and the likelihood function  $\mathcal{L}_L$  of  $\mathcal{G}_L((\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta)$  is given by

$$(4.2) \quad \mathcal{L}_L(\theta|\bar{d}_L) = p_L^\theta(\bar{d}_L) = e^{-\int_{t_0}^{\bar{\tau}_L} \Lambda_\theta(\tau) d\tau} \prod_{k=1}^L \Lambda_\theta(\bar{\tau}_k) \left[ \prod_{l=1}^L p_{U_l|T_l, \mathcal{D}_{l-1}}^\theta(\bar{r}_l|\bar{\tau}_l, \bar{d}_{l-1}) \right],$$

where  $\bar{d}_L \in \mathcal{C}^L \times \mathbb{R}_{[\infty]}^L$ ,  $L = 1, 2, \dots$ .

In supplementary Section SM4, we provide an example to illustrate our results for the specific case that the motion model is described by a linear stochastic differential equation.

In the following, we present and discuss the results of the proposed maximum likelihood estimation method when applied to simulated data sets of trajectories of a single molecule.

**4.1. Simulated parameters.** To analyze the performance of the proposed maximum likelihood estimation method, we simulated different data sets using parameters commonly used in single molecule experiments. Unless otherwise stated, the images of in-focus and out-of-focus molecules were generated with Airy and Born and Wolf profiles (Eqs. (2.4) and (2.6)), respectively, where  $n_a = 1.4$ ,  $\lambda = 520$  nm,  $n_o = 1.515$ , and  $z_0 = 1$   $\mu\text{m}$ . For the Gaussian measurement case, the image of a molecule was generated with a zero-mean Gaussian measurement noise with the probability density function given by Eq. (2.5), where  $\sigma = 70$  nm, which is related to the corresponding Airy profile.

Furthermore, a measurement (magnification) matrix  $M = 100I_{2 \times 2}$  was assumed to map the object space to the image space.

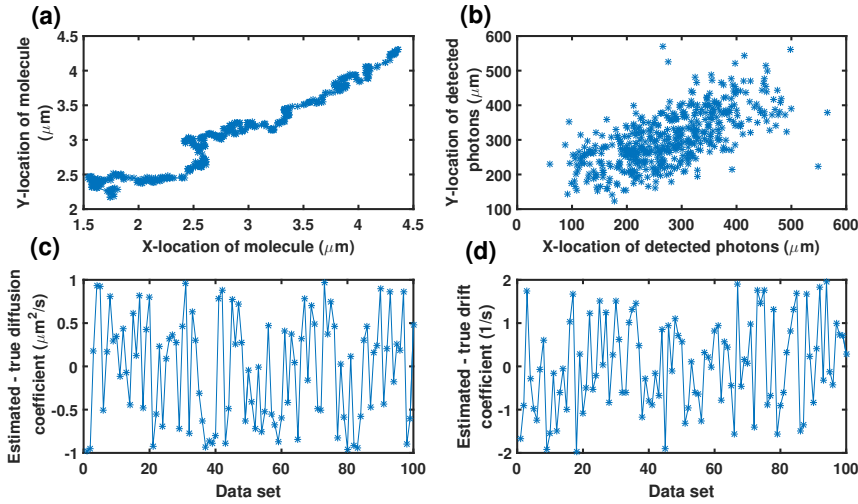


FIG. 3. Analysis of the error of diffusion coefficient and first order drift coefficient estimates produced by the maximum likelihood estimation method for the Born and Wolf measurement model. (a) A trajectory of an out-of-focus molecule, with the out-of-focus level  $z_0 = 1$   $\mu\text{m}$ , in the object space simulated using Eq. (SM4.1) where the time points are drawn from a Poisson process with mean 500 in the time interval  $[0, 100]$  ms with the first order drift coefficient  $F = -10I_{2 \times 2}/\text{s}$  and the diffusion coefficient  $D = 1$   $\mu\text{m}^2/\text{s}$ . We assume the zero order drift is equal to 0. Also, we assume that the initial location of the molecule is Gaussian distributed with mean  $x_0 = (4.4, 4.4)^T$   $\mu\text{m}$  and covariance  $P_0 = 10I_{2 \times 2}$   $\text{nm}^2$ . (b) Detected locations of the photons emitted from the molecule trajectory of part (a) in the image space which are simulated using Eq. (3.6) with the Born and Wolf profile (Eq. (2.6)) and the parameters given in Section 4.1. (c) Differences between the diffusion coefficient estimates and the true diffusion coefficient value for 100 data sets, each containing a trajectory of a molecule simulated using Eqs. (SM4.1) and (3.6) with the Born and Wolf profile, and the parameters given in parts (a) and (b). (d) Differences between the first order drift coefficient estimates and its true value for the data sets of part (c).

**4.2. Estimation results.** Using simulated data sets, we first examine the performance of the maximum likelihood estimation method used to estimate the parameters of the linear motion model of a moving molecule in terms of the bias of the method. The bias is assessed by the average of the deviations of the estimates from the true value. For this purpose, we simulated 100 data sets, each containing a trajectory of an out-of-focus molecule, with the out-of-focus level  $z_0 = 1$   $\mu\text{m}$ , simulated using Eqs. (SM4.1) and (3.6), with the Born and Wolf profile (Eq. (2.6)) and the

parameters given in Section 4.1, with a mean photon count of 500 photons in the time interval  $[0, 100]$  ms, where the first order drift coefficient  $F = -10I_{2 \times 2}/s$  and the diffusion coefficient  $D = 1 \mu\text{m}^2/s$ . We assume the zero order drift is equal to 0. In Figs. 3(a) and 3(b), an example of a molecule trajectory in the object space and its image in the image space are shown. For these data sets, we calculated the maximum likelihood estimates of the diffusion and drift coefficients, separately. For this purpose, we needed to obtain the distributions of the prediction in the likelihood function expressions (Eqs. (4.1) and (4.2)) through Eqs. (3.8) and (3.9), which in general is a computationally expensive problem. We approximated the distributions of the prediction using a sequential Monte Carlo algorithm proposed in [30]. The overall approach is explained in supplementary Section SM6 in detail. In Figs. 3(c) and 3(d), the differences between the maximum likelihood estimates of the diffusion and the first order drift coefficients and the true values are plotted. We also estimated the  $z_0$ -location of the molecule, i.e., the out-of-focus level, and show the errors of estimation in Fig. 4. As can be seen, the deviations of the estimates from the ground truth are, overall, centered around 0 nm, which suggests that there is no systematic bias associated with our proposed method (the average of the diffusion coefficient deviations and the first order drift coefficient deviations are  $-0.0319 \mu\text{m}^2/s$  and  $0.0307/s$ , respectively).

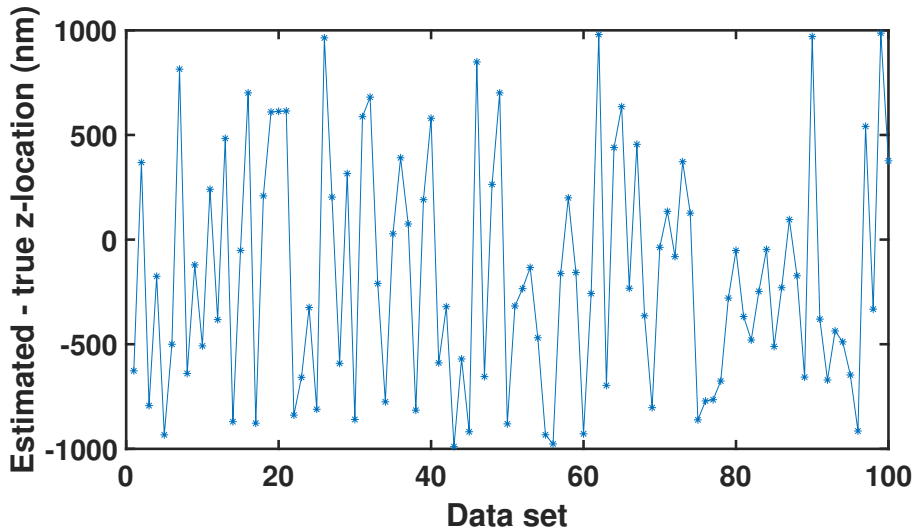


FIG. 4. Analysis of the error of out-of-focus  $z_0$ -location estimates produced by the maximum likelihood estimation method for the Born and Wolf measurement model. Differences between the  $z_0$ -location estimates and its true value,  $z_0 = 1 \mu\text{m}$ , for the data sets of Fig. 3.

We further investigate the distribution  $p_{pr_l}, l = 1, 2, \dots$ , of the prediction of the molecule location, given previous observations, for the molecule trajectory shown in Figs. 3(a) and 3(b). The means of the distributions of the prediction of the molecule  $x$ - and  $y$ -locations and the true  $x$ - and  $y$ -locations are shown in Fig. 5(a) and 5(b). We also show the measurements transformed from the image space to the object space, which are obtained as follows. The location  $X_o := (x_o, y_o, z_o)^T \in \mathbb{R}^3$  in the object space is transformed into the location  $X_i := (x_i, y_i)^T \in \mathbb{R}^2$  in the image space through



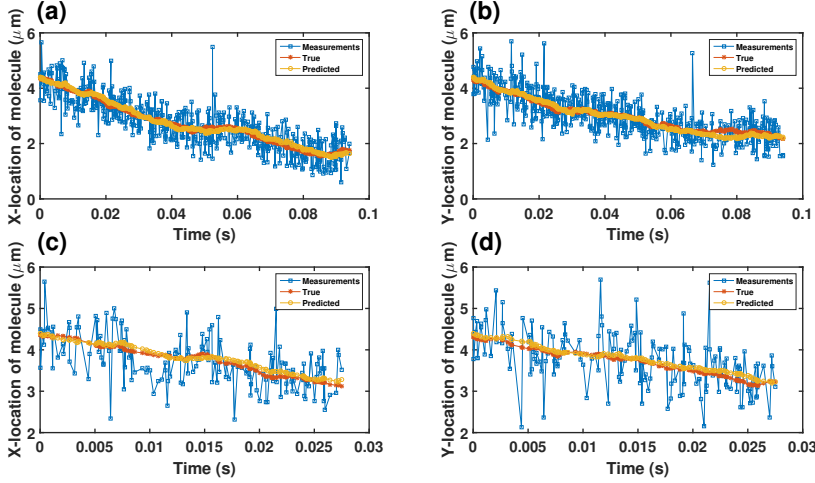


FIG. 5. Predicted locations of the molecule for the Born and Wolf measurement model. (a) and (b) Means of the distributions of the prediction of the molecule  $x$ - and  $y$ -locations and the true  $x$ - and  $y$ -locations of the molecule for the same data set as in Figs. 3(a) and 3(b). The measurements transformed from the image space to the object space are also shown. (c) and (d) Means of the distributions of the prediction of the molecule  $x$ - and  $y$ -locations and the true  $x$ - and  $y$ -locations of the molecule over the time interval  $[0, 27.5]$  ms.

a linear map as, for  $M' \in \mathbb{R}^{2 \times 3}$ ,

$$(4.3) \quad X_i = M' X_o.$$

In Fig. 3, where we have a trajectory of an out-of-focus molecule, with the out-of-focus plane  $z_o = 1 \mu\text{m}$ , it is assumed that the magnification matrix (measurement mapping matrix)  $M' = 100I_{2 \times 2}$ . Then, the  $x$ - and  $y$ -locations of the measurements mapped to the object space are obtained as

$$(4.4) \quad x_o = x_i/100, \quad y_o = y_i/100.$$

For a better visual comparison, the means of the distributions of the prediction of the molecule locations and the true locations for  $x$ - and  $y$ -coordinates are also shown over a shorter time interval in Figs. 5(c) and 5(d). As can be seen, the predicted locations are able to track the true locations of the molecule for both  $x$ - and  $y$ -coordinates. We also show the differences between the means of the distributions of the prediction of the molecule locations and the true locations of the molecule in Fig. SM4 (see Section SM14.6 in the supplementary material). We also applied the proposed method to trajectory data of an in-focus molecule simulated using an Airy profile, with the same standard deviation as the Born and Wolf data, and obtained similar results (see Figs. SM2, SM3 and SM5 in supplementary Section SM7).

As mentioned, in some applications, it is useful to approximate the point spread function of an optical system with a Gaussian profile. We analyzed the error of the estimates for simulated data sets with Gaussian measurement noise, with the same standard deviation as the Born and Wolf data, and obtained similar results (see Figs. 6, 7, 8 and SM6). This time we estimated all the parameters of the trajectory together, i.e., we assumed that the parameter vector  $\theta := (V, F, D, x_0, y_0)$ , where  $V \in \mathbb{R}^2$  and

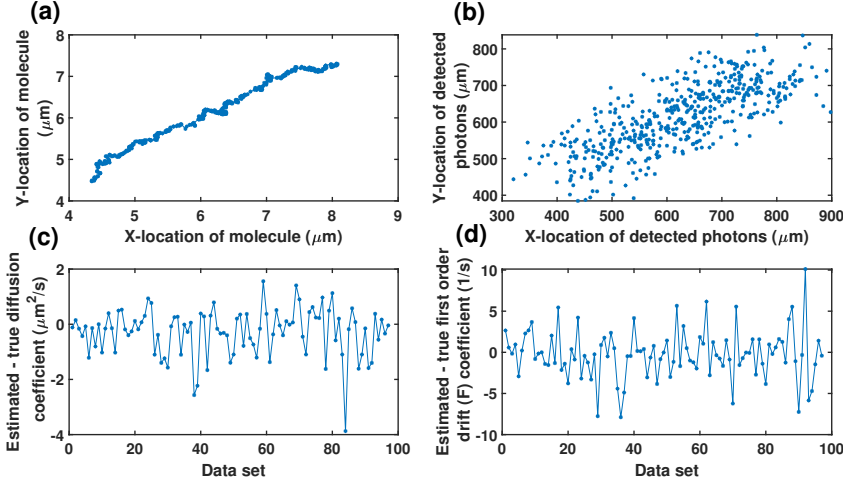


FIG. 6. Analysis of the error of diffusion coefficient and first order drift coefficient estimates produced by the maximum likelihood estimation method for the Gaussian measurement noise case. (a) A 2D trajectory of an in-focus molecule in the object space simulated using Eq. (SM4.1) where the time points are drawn from a Poisson process with mean 500 in the time interval  $[0, 100]$  ms with the first order drift coefficient  $F = -10I_{2 \times 2}/\text{s}$  and the diffusion coefficient  $D = 1 \mu\text{m}^2/\text{s}$ . Also, we assume that the zero order drift coefficient  $V = (100, 100)^T \mu/\text{s}$ , the initial location of the molecule is Gaussian distributed with mean  $x_0 = (4.4, 4.4)^T \mu\text{m}$  and covariance  $P_0 = 10I_{2 \times 2} \text{nm}^2$ . (b) Detected locations of the photons emitted from the molecule trajectory of part (a) in the image space which are simulated using Eq. (3.7) with the Gaussian measurement noise (Eq. (2.5)) and  $\sigma = 0.51 \mu\text{m}$ . (c) Differences between the diffusion coefficient estimates and the true diffusion coefficient value for 100 data sets, each containing a trajectory of a molecule simulated using Eqs. (SM4.1) and (3.7) with the Gaussian profile, and the parameters given in parts (a) and (b). (d) Differences between the first order drift coefficient estimates and its true value for the data sets of part (c).

$F \in \mathbb{R}$  denote the zero order and first order drift, respectively,  $D \in \mathbb{R}$  is the diffusion coefficient and  $(x_0, y_0) \in \mathbb{R}^2$  is the initial location of the molecule. We also consider the more general case where  $F = \begin{bmatrix} F_x & 0 \\ 0 & F_y \end{bmatrix}$ ,  $F_x, F_y \in \mathbb{R}$  (Fig. SM1). In order to calculate the predicted locations of the molecule for Gaussian measurements, we took advantage of the relationship between the likelihood function and Kalman filter formulae (see Theorem 3.2). It improved the computational efficiency significantly.

**5. Fisher information matrix and CRLB.** In any estimation problem, the performance of the estimator can be evaluated by calculating their standard deviations from the true parameter values. According to the Cramér-Rao inequality, the covariance matrix of any unbiased estimator  $\hat{\theta}$  of an unknown vector parameter  $\theta$  is bounded from below by the inverse of the Fisher information matrix  $I(\theta)$ , i.e.,  $\text{Cov}(\hat{\theta}) \geq I^{-1}(\theta)$ . Therefore, a benchmark on the standard deviation of estimates can be obtained by the square root of the inverse of the Fisher information matrix. Note that the Fisher information matrix only depends on the statistical nature of the acquired data and is independent of the applied estimation technique. Since this concept is very important when we have fixed time points, as we defined image detection processes and their probability density functions at fixed time points in Section 2, here, we first introduce a notation for the Fisher information matrix of these processes in Definition 5.1, and

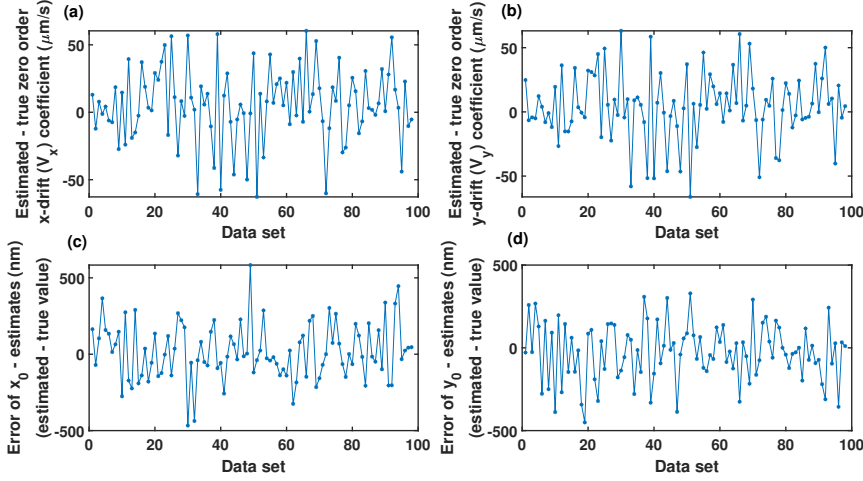


FIG. 7. Analysis of the error of the zero order drift coefficient and initial location estimates produced by the maximum likelihood estimation method for the Gaussian measurement noise case. (a) and (b) Differences between the zero order drift coefficient estimates and its true value, in both  $x$ - and  $y$ -directions, for the data sets of Fig. 6. (c) and (d) Differences between the initial location estimates and its true value, in both  $x$ - and  $y$ -directions, for the data sets of Fig. 6.

use it to calculate the Fisher information matrix of image detection processes for the fixed time interval and for the fixed number of photons in Theorem 5.2.

DEFINITION 5.1. For  $t_0 \leq \tau_1 < \dots < \tau_K$ , let  $\mathcal{G}_{\tau_1, \dots, \tau_K} \left( (\mathcal{U}_K, \mathcal{T}_K), \mathcal{C}, \Theta \right)$  be an image detection process at fixed time points  $\tau_1, \dots, \tau_K$ . We introduce the following notation for the Fisher information matrix of  $\mathcal{G}_{\tau_1, \dots, \tau_K} \left( (\mathcal{U}_K, \mathcal{T}_K), \mathcal{C}, \Theta \right)$  as, for a row parameter vector  $\theta \in \Theta$ ,

$$\begin{aligned} I_{\tau_1, \dots, \tau_K}(\theta) &:= E_{\mathcal{U}_K | \mathcal{T}_K = \tau_{1:K}} \left\{ \left( \frac{\partial \log p_{\mathcal{U}_K | \mathcal{T}_K}^\theta(r_{1:K} | \tau_{1:K})}{\partial \theta} \right)^T \left( \frac{\partial \log p_{\mathcal{U}_K | \mathcal{T}_K}^\theta(r_{1:K} | \tau_{1:K})}{\partial \theta} \right) \right\} \\ &= \int_{\mathcal{C}} \dots \int_{\mathcal{C}} p_{\mathcal{U}_K | \mathcal{T}_K}^\theta(r_{1:K} | \tau_{1:K}) \left( \frac{\partial \log p_{\mathcal{U}_K | \mathcal{T}_K}^\theta(r_{1:K} | \tau_{1:K})}{\partial \theta} \right)^T \\ &\quad \times \left( \frac{\partial \log p_{\mathcal{U}_K | \mathcal{T}_K}^\theta(r_{1:K} | \tau_{1:K})}{\partial \theta} \right) dr_1 \dots dr_K, \end{aligned}$$

for  $t_0 \leq \tau_1 < \dots < \tau_K$ , and  $I_{\tau_1, \dots, \tau_K}(\theta) = 0$ , otherwise, where  $r_{1:K} := (r_1, \dots, r_K)$ ,  $r_1, \dots, r_k \in \mathcal{C}$ ,  $\tau_{1:K} := (\tau_1, \dots, \tau_K)$ ,  $K = 1, 2, \dots$ , and  $E_{\mathcal{U}_K | \mathcal{T}_K = \tau_{1:K}}$  denotes the expected value with respect to the conditional probability density function  $p_{\mathcal{U}_K | \mathcal{T}_K}^\theta$  of  $\mathcal{U}_K$ , given  $\mathcal{T}_K = \tau_{1:K}$ .

THEOREM 5.2. Let  $\mathcal{G}_{[t]} \left( (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \mathcal{C}, \Theta \right)$  and  $\mathcal{G}_L \left( (\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta \right)$  be image detection processes for a time interval  $[t_0, t]$  and for a fixed number  $L$  of photons, respectively. Let  $\mathcal{D}_{[t]} := (\mathcal{U}_{[t]}, \mathcal{T}_{[t]})$ ,  $\mathcal{D}_k := (\mathcal{U}_k, \mathcal{T}_k)$ ,  $k = 0, 1, \dots$ . Assume that the conditional probability density functions  $p_{\mathcal{U}_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^\theta$ ,  $l = 1, 2, \dots$ , of  $\mathcal{U}_l$ , given  $\mathcal{T}_l$  and  $\mathcal{D}_{l-1}$ , satisfy the following regularity conditions, for  $\theta = (\theta_1, \dots, \theta_n) \in \Theta$ ,

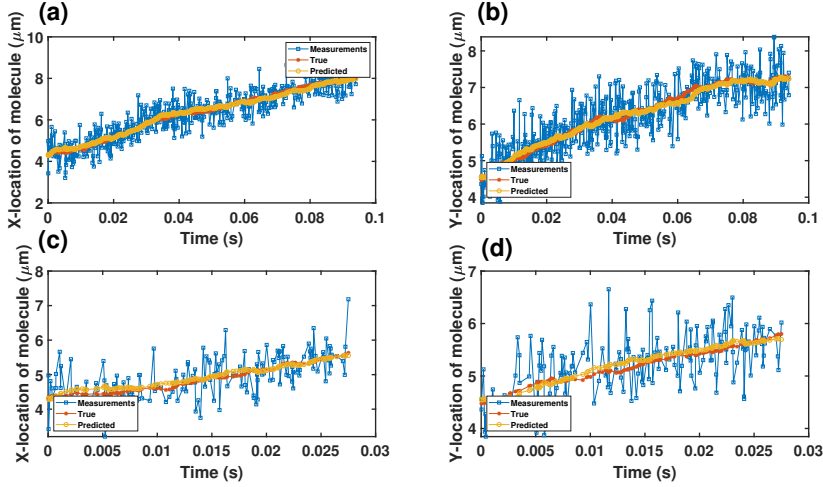


FIG. 8. Predicted locations of the molecule for the Gaussian measurement noise case. (a) and (b) Means of the distributions of the prediction of the molecule  $x$ - and  $y$ -locations and the true  $x$ - and  $y$ -locations of the molecule for the same data set as in Figs. 6(a) and 6(b). The measurements transformed from the image space to the object space are also shown. (c) and (d) Means of the distributions of the prediction of the molecule  $x$ - and  $y$ -locations and the true  $x$ - and  $y$ -locations of the molecule over the time interval  $[0, 27.5]$  ms.

$$(a) \frac{\partial p_{U_i | \mathcal{T}_i, \mathcal{D}_{i-1}}^\theta(r_i | \tau_i, d_{i-1})}{\partial \theta_i} \text{ exists for } i = 1, \dots, n,$$

$$(b) \int_{\mathcal{C}} \left| \frac{\partial p_{U_i | \mathcal{T}_i, \mathcal{D}_{i-1}}^\theta(r | \tau_i, d_{i-1})}{\partial \theta_i} \right| dr < \infty \text{ for } i = 1, \dots, n,$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$  for  $\mathcal{G}_{[t]}$ ,  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$  for  $\mathcal{G}_L$ , and  $p^\theta(r_1 | \tau_1, d_0) := p^\theta(r_1 | \tau_1)$ .

1.1. Then, the Fisher information matrix  $I_{[t]}^\theta$  of  $\mathcal{G}_{[t]}$  is given by

$$(5.1) \quad I_{[t]}^\theta(\theta) = \frac{1}{P_\theta(N(t) = 0)} \left( \frac{\partial P_\theta(N(t) = 0)}{\partial \theta} \right)^T \left( \frac{\partial P_\theta(N(t) = 0)}{\partial \theta} \right) \\ + \sum_{K=1}^{\infty} \int_{t_0}^t \int_{t_0}^{\tau_K} \dots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \left[ \int_{\mathcal{C}} \dots \int_{\mathcal{C}} \frac{1}{p_{[t]}^\theta(d_K, K)} \left( \frac{\partial p_{[t]}^\theta(d_K, K)}{\partial \theta} \right)^T \left( \frac{\partial p_{[t]}^\theta(d_K, K)}{\partial \theta} \right) \right. \\ \left. \times dr_1 \dots dr_K \right] d\tau_1 d\tau_2 \dots d\tau_{K-1} d\tau_K,$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$ , and  $p_{[t]}^\theta$  denotes the probability density function of  $\mathcal{D}_{[t]}$  and  $N(t)$ .

1.2. Assume that the photon detection rate  $\Lambda$  is independent of  $\theta$ . Then,  $I_{[t]}^\theta$  can

be calculated as

$$(5.2) \quad I_{[t]}(\theta) = e^{-\int_{t_0}^t \Lambda(\tau) d\tau} \sum_{K=1}^{\infty} \left\{ \int_{t_0}^t \int_{t_0}^{\tau_K} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} I_{\tau_1, \dots, \tau_K}(\theta) \prod_{k=1}^K \Lambda(\tau_k) \right. \\ \left. \times d\tau_1 d\tau_2 \cdots d\tau_{K-1} d\tau_K \right\},$$

where the Fisher information matrix  $I_{\tau_1, \dots, \tau_K}$  of the image detection process at fixed time points  $\tau_1, \dots, \tau_K$   $\mathcal{G}_{\tau_1, \dots, \tau_K} \left( (\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta \right)$  is given by

$$(5.3) \quad I_{\tau_1, \dots, \tau_K}(\theta) = \begin{cases} \sum_{l=1}^K I_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}(\theta), & t_0 \leq \tau_1 < \cdots < \tau_K \leq t, \\ 0, & \text{otherwise,} \end{cases}$$

in which the Fisher information matrix  $I_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}$  calculated with respect to the conditional probability density function  $p_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^\theta$  at fixed time points  $\mathcal{T}_l = \tau_{1:l}$  is given by

$$(5.4) \quad I_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}(\theta) = E_{U_l | \tau_{1:l}} \left\{ \left( \frac{\partial \log p_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^\theta(r_l | \tau_l, d_{l-1})}{\partial \theta} \right)^T \left( \frac{\partial \log p_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^\theta(r_l | \tau_l, d_{l-1})}{\partial \theta} \right) \right\} \\ = \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_{U_{l-1} | \mathcal{T}_{l-1}}^\theta(r_{1:l-1} | \tau_{1:l-1}) \left[ \int_{\mathcal{C}} \frac{1}{p_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^\theta(r_l | \tau_l, d_{l-1})} \right. \\ \left. \times \left( \frac{\partial p_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^\theta(r_l | \tau_l, d_{l-1})}{\partial \theta} \right)^T \left( \frac{\partial p_{U_l | \mathcal{T}_l, \mathcal{D}_{l-1}}^\theta(r_l | \tau_l, d_{l-1})}{\partial \theta} \right) dr_l \right] dr_{l-1} \cdots dr_1,$$

with  $r_{1:l} := (r_1, \dots, r_l)$ ,  $\tau_{1:l} := (\tau_1, \dots, \tau_l)$ , and  $I_{U_1 | \mathcal{T}_1}^{\tau_1}$  given by

$$(5.5) \quad I_{U_1 | \mathcal{T}_1}^{\tau_1}(\theta) = \int_{\mathcal{C}} \frac{1}{p_{U_1 | \mathcal{T}_1}^\theta(r | \tau_1)} \left( \frac{\partial p_{U_1 | \mathcal{T}_1}^\theta(r | \tau_1)}{\partial \theta} \right)^T \left( \frac{\partial p_{U_1 | \mathcal{T}_1}^\theta(r | \tau_1)}{\partial \theta} \right) dr.$$

2.1. The Fisher information matrix  $I_L$  of  $\mathcal{G}_L$  is given by

$$I_L(\theta) = \int_{t_0}^{\infty} \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \left[ \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} \frac{1}{p_L^\theta(d_L)} \left( \frac{\partial p_L^\theta(d_L)}{\partial \theta} \right)^T \left( \frac{\partial p_L^\theta(d_L)}{\partial \theta} \right) dr_1 \cdots dr_L \right] \\ \times d\tau_1 d\tau_2 \cdots d\tau_{L-1} d\tau_L,$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ , and  $p_L^\theta$  denotes the probability density function of  $\mathcal{D}_L$ .

2.2. Assume that the photon detection rate  $\Lambda$  is independent of  $\theta$ . Then,  $I_L$  can be obtained as

$$(5.6) \quad I_L(\theta) = \int_{t_0}^{\infty} \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} I_{\tau_1, \dots, \tau_L}(\theta) e^{-\int_{t_0}^{\tau_L} \Lambda(\tau) d\tau} \prod_{k=1}^L \Lambda(\tau_k) \\ \times d\tau_1 d\tau_2 \cdots d\tau_{L-1} d\tau_L.$$

*Remark 5.3.* Note that for  $K = 1$ , the time integral of Eq. (5.2) is calculated over the interval  $[t_0, t]$ , i.e.,  $\int_{t_0}^t I_{\tau_1}(\theta) \Lambda(\tau_1) d\tau_1$ .

*Proof.* See Section SM8 in the supplementary material.  $\square$

We next derive expressions for the Fisher information matrices of the image detection processes driven by the stochastic trajectory  $X$  and image function  $q$  for a time interval  $[t_0, t]$  and for a fixed number  $L$  of photons in the following corollary to Theorem 5.2.

**COROLLARY 5.4.** *Let  $\mathcal{G}_{[t]}(X, (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), q, \mathcal{C}, \Theta)$  (or  $\mathcal{G}_L(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta)$ ) be an image detection process driven by the stochastic trajectory  $X$  and image function  $q$  for a time interval  $[t_0, t]$  (or for a fixed number  $L$  of photons). Let, for a row parameter vector  $\theta = (\theta_1, \dots, \theta_n) \in \Theta$ , the  $n$ -dimensional vector  $F_l^\theta$  be given by*

$$(5.7) \quad F_l^\theta(x, d_l) := \underbrace{\left[ (df_x^\theta(r_l))^T \quad (dp_{pr_l}^\theta(x|\tau_l, d_{l-1}))^T \right]}_{\text{Block row vector}} \begin{bmatrix} p_{pr_l}^\theta(x|\tau_l, d_{l-1}) \\ f_x^\theta(r_l) \end{bmatrix}, \quad x \in \mathbb{R}^3,$$

where  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$  (or  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ ),  $r_{1:l} := (r_1, \dots, r_l)$ ,  $\tau_{1:l} := (\tau_1, \dots, \tau_l)$ ,  $p_{pr_l}^\theta := p_{X(T_l)|T_l, \mathcal{D}_{l-1}}^\theta$ ,  $p_{pr_1}^\theta(x|\tau_1, d_0) := p_{pr_1}^\theta(x|\tau_1)$ , denotes the distribution of the prediction of the object location, and  $dp_{pr_l}^\theta := \frac{\partial p_{pr_l}^\theta}{\partial \theta}$ ,  $df_x^\theta := \frac{\partial f_x^\theta}{\partial \theta}$ . Assume that the photon detection rate  $\Lambda$  is independent of  $\theta$ . Then,  $I_{\tau_1, \dots, \tau_K}$  in Eq. (5.2) (or Eq. (5.6)) of Theorem 5.2 is given by

$$I_{\tau_1, \dots, \tau_K}(\theta) = \begin{cases} \sum_{l=1}^K I_{U_l|T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}(\theta), & t_0 \leq \tau_1 < \dots < \tau_K \leq t, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$(5.8) \quad I_{U_l|T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}(\theta) = \int_{\mathcal{C}} \dots \int_{\mathcal{C}} p_{U_{l-1}|\tau_{l-1}}^\theta(r_{1:l-1}|\tau_{1:l-1}) \times \left[ \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left( \int_{\mathcal{C}} \frac{F_l^\theta(x_1, d_l) [F_l^\theta(x_2, d_l)]^T}{p_{U_l|T_l, \mathcal{D}_{l-1}}^\theta(r_l|\tau_l, d_{l-1})} dr_l \right) dx_1 dx_2 \right] dr_{l-1} \dots dr_1,$$

and

$$(5.9) \quad p_{U_{l-1}|\tau_{l-1}}^\theta(r_{1:l-1}|\tau_{1:l-1}) = \prod_{i=1}^{l-1} \int_{\mathbb{R}^3} f_{x_o}^\theta(r_i) p_{pr_i}^\theta(x_o|\tau_i, d_{i-1}) dx_o,$$

with  $I_{U_1|T_1}^{\tau_1}$  given by

$$(5.10) \quad I_{U_1|T_1}^{\tau_1}(\theta) = \int_{\mathcal{C}} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{1}{p_{U_1|T_1}^\theta(r|\tau_1)} \left[ (df_{x_1}^\theta(r))^T \quad (dp_{pr_1}^\theta(x_1|\tau_1))^T \right] \begin{bmatrix} p_{pr_1}^\theta(x_1|\tau_1) \\ f_{x_1}^\theta(r) \end{bmatrix} \times \begin{bmatrix} p_{pr_1}^\theta(x_2|\tau_1) \\ f_{x_2}^\theta(r) \end{bmatrix}^T \begin{bmatrix} df_{x_2}^\theta(r) \\ dp_{pr_1}^\theta(x_2|\tau_1) \end{bmatrix} dx_1 dx_2 dr.$$

**Remark 5.5.** Note that if the image function  $q$  is independent of the parameter vector  $\theta$ , then,

$$F_l^\theta(x, d_l) = f_x(r_l) \left( dp_{pr_l}^\theta(x|\tau_l, d_{l-1}) \right)^T, \quad x \in \mathbb{R}^3,$$

and the expression for  $I_{U_l|T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}$  can be simplified as

$$\begin{aligned}
& I_{U_l|T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}(\theta) \\
&= \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_{U_{l-1}|\mathcal{T}_{l-1}}^\theta(r_{1:l-1}|\tau_{1:l-1}) \left[ \int_{\mathcal{C}} \frac{1}{p_{U_l|T_l, \mathcal{D}_{l-1}}^\theta(r_l|\tau_l, d_{l-1})} \right. \\
&\quad \times \left( \frac{\partial}{\partial \theta} \int_{\mathbb{R}^3} f_{x_o}(r_l) p_{p_{r_l}}^\theta(x_o|\tau_l, d_{l-1}) dx_o \right)^T \left( \frac{\partial}{\partial \theta} \int_{\mathbb{R}^3} f_{x_o}(r_l) p_{p_{r_l}}^\theta(x_o|\tau_l, d_{l-1}) dx_o \right) dr_l \Big] \\
&\quad \times dr_{l-1} \cdots dr_1 \\
&= \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_{U_{l-1}|\mathcal{T}_{l-1}}^\theta(r_{1:l-1}|\tau_{1:l-1}) \left\{ \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left[ \int_{\mathcal{C}} \frac{1}{p_{U_l|T_l, \mathcal{D}_{l-1}}^\theta(r_l|\tau_l, d_{l-1})} \right. \right. \\
(5.11) \quad &\quad \times f_{x_1}(r_l) f_{x_2}(r_l) \left( \frac{\partial p_{p_{r_l}}^\theta(x_1|\tau_l, d_{l-1})}{\partial \theta} \right)^T \left( \frac{\partial p_{p_{r_l}}^\theta(x_2|\tau_l, d_{l-1})}{\partial \theta} \right) dr_l \Big] dx_1 dx_2 \Big\} dr_{l-1} \cdots dr_1.
\end{aligned}$$

*Proof.* See Section SM9 in the supplementary material.  $\square$

As mentioned in Section 2, for special cases of an object with a deterministic trajectory and a static object, the probability density function of the image detection process  $\mathcal{G}_{\tau_1, \dots, \tau_K}$  at fixed time points  $t_0 \leq \tau_1 < \cdots < \tau_K$  is simplified as given by Eqs. (2.8) and (2.9), respectively. We next in Corollary 5.6 to Theorem 5.2 calculate the Fisher information matrix for these special cases, and show that the obtained results are consistent with the results presented in [21, 36, 34, 35].

**COROLLARY 5.6.** *For  $t_0 \leq \tau_1 < \cdots < \tau_K$ , let  $\mathcal{G}_{\tau_1, \dots, \tau_K}((\mathcal{U}_K, \mathcal{T}_K), \mathcal{C}, \Theta)$  be an image detection process at fixed time points  $\tau_1, \dots, \tau_K$ . Assume that  $p_{U_l|T_l, \mathcal{D}_{l-1}}(r_l|\tau_l, d_{l-1}) = p_{U_l|T_l}(r_l|\tau_l)$ ,  $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$ ,  $l = 1, 2, \dots$ .*

1. *Then, the Fisher information matrix  $I_{\tau_1, \dots, \tau_K}$  of  $\mathcal{G}_{\tau_1, \dots, \tau_K}((\mathcal{U}_K, \mathcal{T}_K), \mathcal{C}, \Theta)$  is given by*

$$I_{\tau_1, \dots, \tau_K}(\theta) = \begin{cases} \sum_{l=1}^K I_{U_l|T_l}^\tau(\theta), & t_0 \leq \tau_1 < \cdots < \tau_K, \\ 0, & \text{otherwise,} \end{cases}$$

where for  $l = 1, \dots, K$ ,

$$I_{U_l|T_l}^\tau(\theta) = \int_{\mathbb{R}^2} \frac{1}{p_{U_l|T_l}^\theta(r|\tau)} \left( \frac{\partial p_{U_l|T_l}^\theta(r|\tau)}{\partial \theta} \right)^T \left( \frac{\partial p_{U_l|T_l}^\theta(r|\tau)}{\partial \theta} \right) dr.$$

2.1. *For an object with deterministic trajectory  $X_\tau(\theta) := (x_\tau(\theta), y_\tau(\theta)) \in \mathbb{R}^2$ ,  $\tau \geq t_0$ , assume that there exists an image function  $q: \mathbb{R}^2 \mapsto \mathbb{R}$ , which describes the image of an object on the detector plane at unit lateral magnification and it is assumed to be independent of the parameter vector  $\theta = (\theta_1, \dots, \theta_n) \in \Theta$ , such that*

$$p_{U_l|T_l}^\theta(r|\tau) := \frac{1}{M^2} q\left(\frac{x}{M} - x_\tau(\theta), \frac{y}{M} - y_\tau(\theta)\right),$$

where  $r = (x, y) \in \mathbb{R}^2$ ,  $t_0 \leq \tau \leq t$ , and  $M > 1$  is a magnification factor. Let  $D_1q$  and  $D_2q$  be the partial derivatives of  $q$  with respect to the  $x$ - and  $y$ -coordinates, respectively.



Also, let  $D_j x_\tau$  and  $D_j y_\tau, j = 1, \dots, n$ , denote the partial derivatives of  $x_\tau$  and  $y_\tau$  with respect to the  $j^{\text{th}}$  parameter coordinate, respectively. Then, for  $t_0 \leq \tau_1 < \dots < \tau_K$ ,

$$I_{\tau_1, \dots, \tau_K}(\theta) = \sum_{l=1}^K I_{\tau_l}(\theta),$$

where

$$I_{\tau_l}(\theta) = V_\theta^T(\tau_l) \left( \int_{\mathbb{R}^2} \frac{1}{q(u, v)} \begin{bmatrix} (D_1 q)(u, v) \\ (D_2 q)(u, v) \end{bmatrix} \begin{bmatrix} (D_1 q)(u, v) \\ (D_2 q)(u, v) \end{bmatrix}^T dudv \right) V_\theta(\tau_l),$$

and

$$V_\theta(\tau_l) := \begin{bmatrix} (D_1 x_{\tau_l})(\theta) & \dots & (D_n x_{\tau_l})(\theta) \\ (D_1 y_{\tau_l})(\theta) & \dots & (D_n y_{\tau_l})(\theta) \end{bmatrix} \in \mathbb{R}^{2 \times n}.$$

2.2. For a static object with position  $X_0(\theta) = (x_0(\theta), y_0(\theta)) \in \mathbb{R}^2$ , we have, for  $t_0 \leq \tau_1 < \dots < \tau_K$ ,

$$I_{\tau_1, \dots, \tau_K}(\theta) = I(\theta) = K \tilde{I}(\theta),$$

where

$$\tilde{I}(\theta) = V_\theta^T \left( \int_{\mathbb{R}^2} \frac{1}{q(u, v)} \begin{bmatrix} (D_1 q)(u, v) \\ (D_2 q)(u, v) \end{bmatrix} \begin{bmatrix} (D_1 q)(u, v) \\ (D_2 q)(u, v) \end{bmatrix}^T dudv \right) V_\theta,$$

and for  $\theta = (\theta_1, \dots, \theta_n) \in \theta$ ,

$$V_\theta := \begin{bmatrix} (D_1 x_0)(\theta) & \dots & (D_n x_0)(\theta) \\ (D_1 y_0)(\theta) & \dots & (D_n y_0)(\theta) \end{bmatrix} \in \mathbb{R}^{2 \times n}.$$

*Proof.* See Section SM10 in the supplementary material.  $\square$

The material presented in Theorem 5.2 and Corollary 5.4 provides a mathematical framework to calculate the Fisher information matrix of image detection processes for a fixed time interval and for a fixed number of photons for a moving object with a general stochastic motion model. As mentioned before, in many biological applications, the motion of a small object in subcellular environments can be modeled by a linear stochastic differential equation. The solution of this linear stochastic differential equation can be modeled by a first order system driven by Gaussian noise. In Corollary 5.7 to Theorem 5.2, we obtain recursive expressions for the Fisher information matrices for both image detection processes for a fixed time interval and fixed number of photons, in case that the dynamical system is described by a first order system with Gaussian process and measurement noise.

**COROLLARY 5.7.** Let  $\mathcal{G}_{[t]}^g \left( \left( \tilde{X}, H, \tilde{W}_g, Z_g \right), (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$  (or  $\mathcal{G}_L^g \left( \left( \tilde{X}, H, \tilde{W}_g, Z_g \right), (\mathcal{U}_L, \mathcal{T}_L), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$ ) be an image detection process with expanded state space  $\tilde{X}$  and Gaussian process and measurement noise models for a time interval  $[t_0, t]$  (or for a fixed number  $L$  of photons). Let  $C := M'H$ . Assume that the photon detection rate  $\Lambda$ ,  $C$  and  $Z_g$  are independent of  $\theta$ . Let

$$(5.12) \quad \begin{aligned} S_{\theta, l}^{(j i)} - A_{\theta, l}^{(j)} S_{\theta, l-1}^{(j i)} \left( A_{\theta, l}^{(i)} \right)^T &= B_{\theta, l}^{(j)} R_{\theta, l-1} \left( B_{\theta, l}^{(j)} \right)^T, \quad l = 2, 3, \dots, \\ S_{\theta, 1}^{(j i)} &= \begin{bmatrix} \tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta, 0} + \tilde{a}_\theta(\tau_0, \tau_1) \\ \frac{\partial(\tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta, 0} + \tilde{a}_\theta(\tau_0, \tau_1))}{\partial \theta_j} \end{bmatrix} \begin{bmatrix} \left( \tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta, 0} + \tilde{a}_\theta(\tau_0, \tau_1) \right) \\ \left( \frac{\partial(\tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta, 0} + \tilde{a}_\theta(\tau_0, \tau_1))}{\partial \theta_i} \right) \end{bmatrix}^T, \end{aligned}$$

where

$$A_{\theta,l}^{(i)} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_{l-1}, \tau_l) & 0_{k \times k} \\ \frac{\partial \tilde{\phi}_\theta(\tau_{l-1}, \tau_l)}{\partial \theta_i} & \tilde{\phi}_\theta(\tau_{l-1}, \tau_l) (I_{k \times k} - K_{\theta,l-1} C) \end{bmatrix}, \quad B_{\theta,l}^{(i)} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_{l-1}, \tau_l) K_{\theta,l-1} \\ \frac{\partial (\tilde{\phi}_\theta(\tau_{l-1}, \tau_l) K_{\theta,l-1})}{\partial \theta_i} \end{bmatrix},$$

and  $R_{\theta,l} := CP_{\theta,l}^{l-1} C^T + \Sigma_g$ ,  $K_{\theta,l} := P_{\theta,l}^{l-1} C^T (CP_{\theta,l}^{l-1} C^T + \Sigma_g)^{-1}$ ,  $l = 1, 2, \dots$ , where  $P_{\theta,l}^{l-1}$  is obtained through Eqs. (3.11) and (3.12).

Then, the Fisher information matrix  $I_{\tau_1, \dots, \tau_K}$  in Eq. (5.2) (or Eq. (5.6)) of Theorem 5.2 can be calculated as

$$(5.13) \quad I_{\tau_1, \dots, \tau_K}(\theta) = \begin{cases} \sum_{l=1}^K I_{U_l | T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}(\theta), & t_0 \leq \tau_1 < \dots < \tau_K \leq t, \\ 0, & \text{otherwise,} \end{cases}$$

where, for  $\theta = (\theta_1, \dots, \theta_n) \in \Theta$  and  $l = 1, \dots, K$ , the  $i, j^{\text{th}}$ ,  $i, j = 1, \dots, n$ , entry  $\left[ I_{U_l | T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l} \right]_{i,j}$  of  $I_{U_l | T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}$  can be calculated as

$$(5.14) \quad \left[ I_{U_l | T_l, \mathcal{D}_{l-1}}^{\tau_1, \dots, \tau_l}(\theta) \right]_{i,j} = \frac{1}{2} \text{trace} \left[ R_{\theta,l}^{-1} \frac{\partial R_{\theta,l}}{\partial \theta_i} R_{\theta,l}^{-1} \frac{\partial R_{\theta,l}}{\partial \theta_j} \right] + \text{trace} \left\{ R_{\theta,l}^{-1} \tilde{C} S_{\theta,l}^{(j,i)} \tilde{C}^T \right\},$$

with  $\tilde{C} := [0_{2 \times k} \quad C]$ .

*Proof.* See Section SM11 in the supplementary material.  $\square$

In Section SM12, we provide an example to illustrate our results for calculating the Fisher information matrix for the specific case of a linear trajectory described in the example provided in Section SM4 of the supplementary material.

**5.1. CRLB and standard deviation of estimates for different photon counts.** We next evaluate the performance of our proposed maximum likelihood estimation method in terms of the standard deviation of the estimates. For this purpose, we simulated data sets of the detected photons emitted from a molecule, referred to as the images of a molecule, with a stochastic trajectory which differ by the mean photon count, i.e., the mean number of detected photons during the exposure time interval, assumed for each trajectory. This mean photon count ranges from 250 to 1250. For each mean photon count, the data set consists of 100 repeat images simulated using the Gaussian profile (Eq. (3.7)) with the parameters given in Section 4.1. For these data sets, we calculated the maximum likelihood estimates of the diffusion and first order drift coefficients, separately. Also, for the given data set and time points, we obtained the square roots of the CRLBs for the diffusion and first order drift coefficient by calculating the square roots of the inverse of their corresponding Fisher information matrices at the fixed time points. It can be seen in the first row of Fig. 9 that as the mean photon count increases and thereby the amount of data that is available for the estimation increases the CRLB for the estimates improves, consistent with the expectation that an increasing amount of data leads to improved estimation results. The standard deviations of the estimates show the analogous behavior while exhibiting the expected fluctuations due to the stochastic nature of the sample standard deviations. Also, the percentage differences between the standard deviations and the square roots of the CRLBs are shown in the second row of Fig. 9. The percentage difference is the difference between the standard deviation of the estimates and the square root of the corresponding CRLB, expressed as a percentage of the square root of the corresponding CRLB. As can be seen, these percentage differences are at most around 10%.

Note that in theory, the square root of the CRLB provides a lower bound on the standard deviation of an unbiased estimator. However, in simulation results, we deal with individual stochastic trials rather than the probabilistic expressions that are used in the statements of the CRLB. This means that even if the (probabilistic) standard deviation of an estimator attains the CRLB, the sample standard deviation obtained in a stochastic simulation will deviate from the probabilistic expression and could be expected to be both above and below the CRLB.

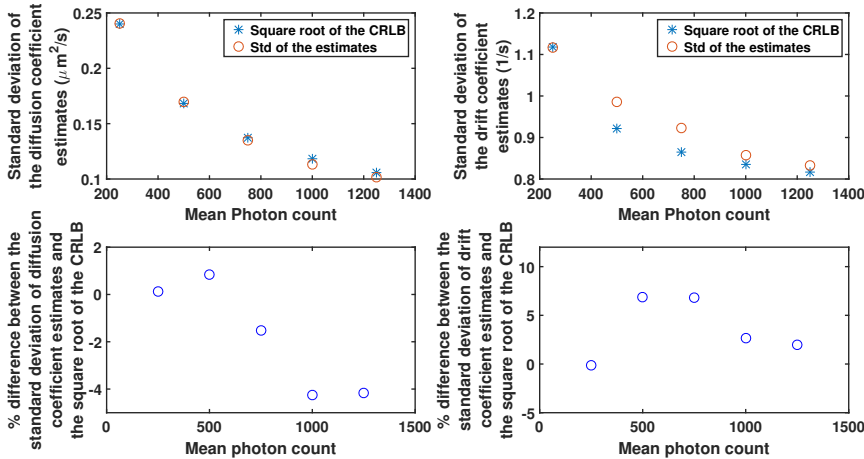


FIG. 9. Analysis of the standard deviation of diffusion coefficient and first order drift coefficient estimates produced by the maximum likelihood estimation method for the Gaussian measurement noise case. Shown in the first row are the standard deviations of the diffusion coefficient and first order drift coefficient estimates versus the square roots of their corresponding CRLBs for simulated data sets. The simulated data sets are the detected photons emitted from a molecule, referred to as the images of a molecule, with a stochastic trajectory which differ by the mean photon count assumed for each trajectory. For each mean photon count, the data set consists of 100 repeat images. For a given data set, the time points of the detected photons are drawn from a Poisson process and are the same for the all trajectories. All trajectories are simulated in the object space using Eq. (SM4.1) with the first order drift coefficient  $F = -10I_{2 \times 2}/s$  and the diffusion coefficient  $D = 1 \mu\text{m}^2/s$ . We assume the zero order drift is equal to 0. Also, we assume that the initial location of the molecule is Gaussian distributed with mean  $x_0 = (5, 5)^T \mu\text{m}$  and covariance  $P_0 = 10I_{2 \times 2} \text{nm}^2$ . Detected locations of the photons emitted from the molecule in the image space are simulated using Eq. (3.7) with the parameters given in Section 4.1. Shown in the second row are the percentage differences between the standard deviation of the diffusion coefficient and first order drift coefficient estimates and the square roots of their corresponding CRLBs.

## 5.2. Fisher information matrix for non-Gaussian measurement noise.

So far, for computational purposes and taking advantage of the Kalman filter formulation, we have focused on computing the Fisher information matrix and CRLB only for Gaussian measurements. Although the Gaussian assumption is very useful in some applications, there are many cases for which this assumption can be problematic in practice due to the fact that the Gaussian model is often not a suitable approximation for an analytical image profile. As mentioned earlier, from optical diffraction theory, a typical point spread function for an in-focus molecule is given by the Airy profile. Also, for the out-of-focus scenario, the image function is given by a classical model of Born and Wolf [6].

Here, we computed the Fisher information matrix of both the first order drift and diffusion coefficients for the Airy measurements case and compared the results with the Fisher information matrix obtained for the case that the Airy profile is approximated by a 2D Gaussian profile. The typical approximation of the Airy profile with  $\alpha := 2\pi n_a/\lambda$  by a 2D Gaussian profile with standard deviation  $\sigma$  yields a value of  $\sigma = 1.323/\alpha$  [21]. We only focused on the one photon case, since computing the integrals of the Fisher information expression for the Airy profile case numerically requires a large number of samples and it is computationally expensive (see Section SM13 in the supplementary material for the detailed computational procedure). As shown in Fig. 10, the difference between the Fisher information matrices of these two different profiles can be significant. As can be seen, this difference goes to 0 for the diffusion coefficient and remains constant for the estimation of the drift coefficient. The other main difference is that the CRLB decreases in the diffusion estimation case, whereas it increases for the drift estimation as the size of the drift increases.

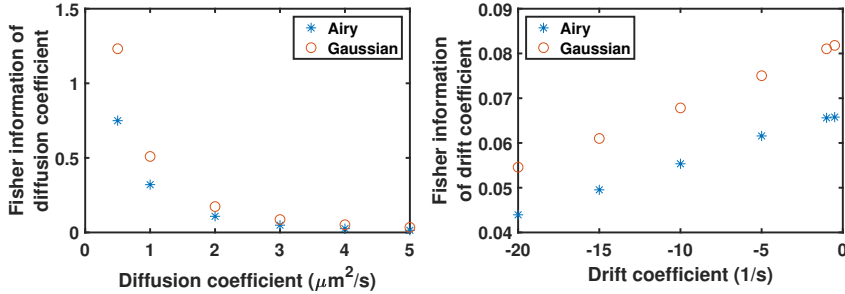


FIG. 10. Fisher information matrix for Airy measurement noise versus Gaussian measurement noise. Fisher information matrix of diffusion and first order drift coefficients for the Airy measurement noise with parameter  $\alpha = 2\pi n_a/\lambda$  given in Section 4.1 and by a 2D Gaussian profile with standard deviation  $\sigma = 1.323/\alpha$ , in case that we have one photon with arrival time of  $\tau_1 = 20$  ms.

**5.3. CRLB and Fisher information matrix for different sets of time points.** To examine further the CRLB for parameter estimation for a moving single molecule with a stochastic trajectory, we calculated the square root of the CRLB for the simulated trajectories with the same parameters as in Fig. 9, and different time points drawn from a Poisson process with a mean value which ranges from 250 to 1250. In Fig. 11, we have plotted, for a given photon count, the median and standard deviation of the different simulations of the square root of the CRLB for both the diffusion and drift estimates. As in prior analyses the results vary significantly for the two scenarios. While in both cases the median decreases with increasing photon count (and thereby increasing acquisition time), the standard deviations of the square roots of the CRLB expressions behave very differently. In case of the diffusion parameter the standard deviations of the CRLB expressions are almost insignificant, indicating that the specifics of the photon detection times do not have a major impact on the standard deviation with which the diffusion coefficient can be estimated. The situation for the estimation of the drift parameter is, however, very different. Here the corresponding standard deviations are relatively high and in fact increase with the number of photons that are acquired. This shows that the standard deviation with which the drift coefficient can be estimated, in contrast to the diffusion coefficient, is highly dependent on the specific time points at which the emitted photons are

detected.

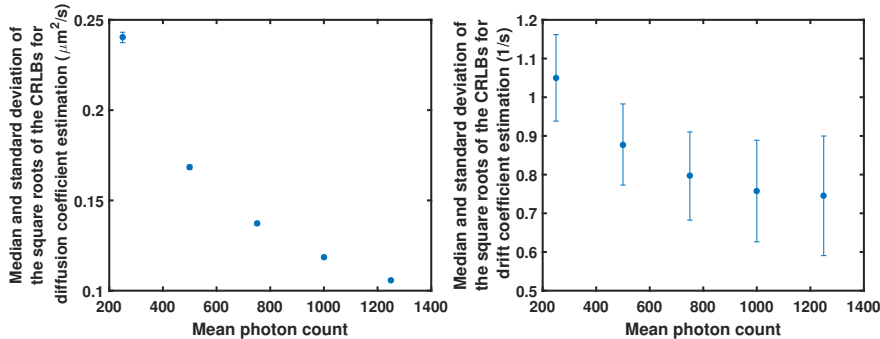


FIG. 11. Analysis of the square root of the CRLB of the diffusion coefficient and first order drift coefficient estimates for different sets of Poisson distributed time points. Medians and standard deviations of the square roots of the CRLBs of the diffusion coefficient and the first order drift coefficient estimates are shown by the circles and error bars, respectively, for the simulated trajectories with the same parameters as in Fig. 9, and different time points drawn from a Poisson process with the same mean value, which ranges from 250 to 1250.

We also show the Fisher information matrices (and Fisher information matrix increments, i.e., the amount of information obtained by detecting one additional photon) for Poisson distributed time points and for equally distributed time points in Fig. 12. For this purpose, we simulated two data sets of single molecule trajectories with Gaussian measurements, the first containing a trajectory of a molecule simulated using Eqs. (SM4.1), where the time points are drawn from a Poisson process with mean 250 in the time interval  $[0, 50]$  ms, and the second containing 250 equally spaced time points in the time interval  $[0, 50]$  ms. We then calculated the Fisher information matrix increments and Fisher information matrix (sum of the increments) on the diffusion coefficient estimation for both data sets. As can be seen, the Fisher information matrix increments, after initial iterations, are constant for the case of equally spaced time points. However, for different realizations of Poisson time points, the Fisher information matrix increments are different from each other. When the time difference between two successive time points decreases (increases), the Fisher information matrix increment of the diffusion coefficient decreases (increases), and conversely, the corresponding CRLB increases (decreases).

#### REFERENCES

- [1] A. V. ABRAHAM, S. RAM, J. CHAO, E. S. WARD, AND R. J. OBER, *Quantitative study of single molecule location estimation techniques*, Opt. Express, 17 (2009), pp. 23352–23373.
- [2] N. H. AKHTAR, O. PAIL, A. SARAN, L. TYRELL, AND S. T. TAGAWA, *Prostate-specific membrane antigen-based therapeutics*, Adv. Urol., 2012 (2012).
- [3] T. T. ASHLEY AND S. B. ANDERSSON, *Method for simultaneous localization and parameter estimation in particle tracking experiments*, Phys. Rev. E, 92 (2015), p. 052707.
- [4] A. M. BASHAROV, *Derivation of kinetic equations from non-Wiener stochastic differential equations*, J. Phys. Conf. Ser., 478 (2013), p. 012011.
- [5] A. J. BERGLUND, *Statistics of camera-based single-particle tracking*, Phys. Rev. E, 82 (2010), p. 011917.
- [6] M. BORN AND E. WOLF, *Principles of optics*, Cambridge Univ. Press, p. 1999.
- [7] V. BRIANE, C. KERVRANN, AND M. VIMOND, *Statistical analysis of particle trajectories in living cells*, Phys. Rev. E, 97 (2018), p. 062121.

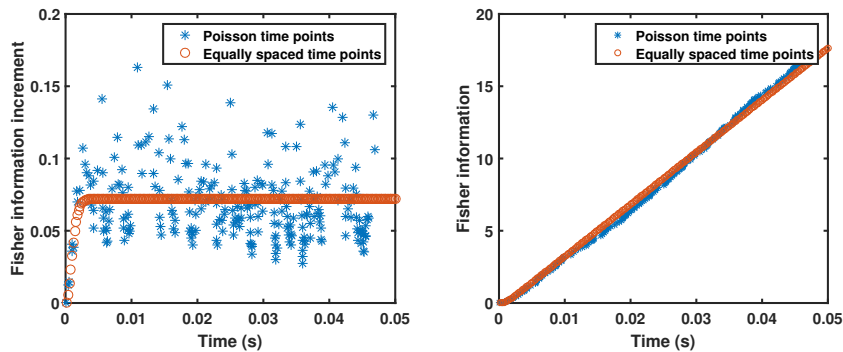


FIG. 12. Fisher information analysis of single molecule trajectories simulated using Poisson distributed and equally spaced time points. Shown in the left are the Fisher information matrix increments on the diffusion coefficient estimation for data sets of two trajectories, first containing a trajectory of a molecule simulated using Eqs. (SM4.1), where the time points are drawn from a Poisson process with mean 250 in the time interval  $[0, 50]$  ms, and second containing 250 equally spaced time points in the time interval  $[0, 50]$  ms, with the parameters given in Fig. (9). Shown in the right is the Fisher information matrix (sum of the increments) for both trajectories.

- [8] C. P. CALDERON, *Motion blur filtering: A statistical approach for extracting confinement forces and diffusivity from a single blurred trajectory*, Phys. Rev. E, 93 (2016), p. 053303.
- [9] 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, 10 (2015), p. e0137633.
- [10] 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, 117 (2013), pp. 15701–15713.
- [11] J. CHAO, E. S. WARD, AND R. J. OBER, *Fisher information theory for parameter estimation in single molecule microscopy: tutorial*, J. Opt. Soc. Amer. A, 33 (2016), pp. B36–B57.
- [12] V. A. DIPIPPO, W. C. OLSON, H. M. NGUYEN, L. G. BROWN, R. L. VESSELLA, AND E. COREY, *Efficacy studies of an antibody-drug conjugate PSMA-ADC in patient-derived prostate cancer xenografts*, The Prostate, 75 (2015), pp. 303–313.
- [13] M. FRIEDRICH, T. RAUM, R. LUTTERBUESE, M. VOELKEL, P. DEEGEN, D. RAU, R. KISCHEL, P. HOFFMANN, C. BRANDL, J. SCHUHMACHER, P. MUELLER, R. FINNERN, M. FUERGUT, D. ZOPF, J. W. SLOOTSTRA, P. A. BAEUERLE, B. RATTEL, AND F. KUFER, *Regression of human prostate cancer xenografts in mice by AMG 212/BAY2010112, a novel PSMA/CD3-Bispecific BiTE antibody cross-reactive with non-human primate antigens*, Mol. Cancer Ther., 11 (2012), p. 26642673.
- [14] A. H. JAZWINSKI, *Stochastic processes and filtering theory*, Acad. Press, New York, USA, (1970).
- [15] L. JIANG, S. S. SINGH, AND S. YILDIRIM, *Bayesian tracking and parameter learning for non-linear multiple target tracking models*, IEEE Trans. Signal Process., 63 (2015), pp. 5733–5745.
- [16] N. KANTAS, A. DOUCET, S. S. SINGH, J. MACIEJOWSKI, AND N. CHOPIN, *On particle methods for parameter estimation in state-space models*, Statist. Sci., 30 (2015), pp. 328–351.
- [17] C. MANZO AND M. F. GARCIA-PARAJO, *A review of progress in single particle tracking: from methods to biophysical insights*, Rep. Progr. Phys., 78 (2015), p. 124601.
- [18] X. MICHALET, *Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium*, Phys. Rev. E, 82 (2010), p. 041914.
- [19] X. MICHALET AND A. J. BERGLUND, *Optimal diffusion coefficient estimation in single-particle tracking*, Phys. Rev. E, 85 (2012), p. 061916.
- [20] P. C. NELSON, C. ZURLA, D. BROGIOLI, J. F. BEAUSANG, L. FINZI, AND D. DUNLAP, *Tethered particle motion as a diagnostic of DNA tether length*, J. Phys. Chem. B, 110 (2006), p. 1726017267.
- [21] R. J. OBER, S. RAM, AND E. S. WARD, *Localization accuracy in single-molecule microscopy*,

- Biophys. J., 8 (2004), pp. 1185–1200.
- [22] H. QIAN, M. P. SHEETZ, AND E. L. ELSON, *Single particle tracking. Analysis of diffusion and flow in two-dimensional systems*, Biophys. J., 60 (1991), pp. 910–921.
  - [23] S. RAM, *Resolution and localization in single molecule microscopy*, Ph. D. thesis, University of Texas at Arlington/University of Texas Southwestern Medical Center at Dallas. 2007.
  - [24] S. RAM, E. S. WARD, AND R. J. OBER, *A stochastic analysis of performance limits for optical microscopes*, Multidimens. Syst. Signal Process., 17 (2006), pp. 27–57.
  - [25] 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, 93 (2016), p. 042401.
  - [26] H. RISKEN, *The Fokker-Planck equation: methods of solution and applications*, Springer, Berlin, Germany, (1996).
  - [27] M. J. SAXTON, *Single-particle tracking: connecting the dots*, Nat. Methods, 5 (2008), pp. 671–672.
  - [28] M. J. SAXTON, *Two-dimensional continuum percolation threshold for diffusing particles of nonzero radius*, Biophys. J., 99 (2010), pp. 1490–1499.
  - [29] M. J. SAXTON AND K. JACOBSON, *Single-particle tracking: applications to membrane dynamics*, Annu. Rev. Biophys. Biomol. Struct., 26 (1997), pp. 373–399.
  - [30] T. B. SCHÖN, A. WILLS, AND B. NINNESS, *System identification of nonlinear state-space models*, Automatica, 47 (2011), pp. 39–49.
  - [31] Z. SCHUSS, *Theory and applications of stochastic processes: an analytical approach*, Springer, New York, USA, (2009).
  - [32] D. L. SNYDER AND M. I. MILLER, *Random point processes in time and space*, Springer Verlag, New York, USA, (1991).
  - [33] R. L. STREIT, *Poisson point processes*, Springer, Boston, USA, (2010).
  - [34] 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. Express, 8 (2017), pp. 1332–1355.
  - [35] 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*, Proc. SPIE. 2017, p. 10070: 100700J.
  - [36] 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., 59 (2011), pp. 895–911.