

Fisher information theory for parameter estimation in single molecule microscopy: tutorial

JERRY CHAO,^{1,2} E. SALLY WARD,^{2,3} AND RAIMUND J. OBER^{1,2,*}

¹Department of Biomedical Engineering, Texas A&M University, College Station, Texas 77843, USA

²Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, Texas 77843, USA

³Department of Microbial Pathogenesis and Immunology, Texas A&M Health Science Center, College Station, Texas 77843, USA

*Corresponding author: raimund.ober@tamu.edu

Received 17 February 2016; revised 24 May 2016; accepted 26 May 2016; posted 1 June 2016 (Doc. ID 259337); published 22 June 2016

Estimation of a parameter of interest from image data represents a task that is commonly carried out in single molecule microscopy data analysis. The determination of the positional coordinates of a molecule from its image, for example, forms the basis of standard applications such as single molecule tracking and localization-based super-resolution image reconstruction. Assuming that the estimator used recovers, on average, the true value of the parameter, its accuracy, or standard deviation, is then at best equal to the square root of the Cramér–Rao lower bound. The Cramér–Rao lower bound can therefore be used as a benchmark in the evaluation of the accuracy of an estimator. Additionally, as its value can be computed and assessed for different experimental settings, it is useful as an experimental design tool. This tutorial demonstrates a mathematical framework that has been specifically developed to calculate the Cramér–Rao lower bound for estimation problems in single molecule microscopy and, more broadly, fluorescence microscopy. The material includes a presentation of the photon detection process that underlies all image data, various image data models that describe images acquired with different detector types, and Fisher information expressions that are necessary for the calculation of the lower bound. Throughout the tutorial, examples involving concrete estimation problems are used to illustrate the effects of various factors on the accuracy of parameter estimation and, more generally, to demonstrate the flexibility of the mathematical framework. © 2016 Optical Society of America

OCIS codes: (180.2520) Fluorescence microscopy; (100.2960) Image analysis; (110.3055) Information theoretical analysis; (030.5290) Photon statistics; (000.5490) Probability theory, stochastic processes, and statistics.

<http://dx.doi.org/10.1364/JOSAA.33.000B36>

1. INTRODUCTION

Single molecule microscopy [1,2] is a powerful technique that has enabled the study of biological processes at the level of individual molecules using the fluorescence microscope. The technique is achieved through the use of a suitable fluorophore to label the molecule of interest, an appropriate light source to excite the fluorophore, a properly designed microscope system to capture the fluorescence emitted by the fluorophore while minimizing the collection of extraneous light, and a quantum-efficient detector to record the fluorescence. Using experimental setups built to this high-level specification, researchers have gained significant insight into dynamic processes in living cells by observing and analyzing the behavior of individual molecules of interest [3–9].

An important aspect of quantitative data analysis in single molecule microscopy is the estimation of parameters of interest from the acquired images. Perhaps the most prominent of examples is the estimation of the positional coordinates of a molecule of interest [10,11]. By estimating the position of a

moving molecule in each image of a time sequence, the complex motion revealed by the resulting trajectory has contributed to the elucidation of biological processes at the molecular level [3,4,7–9]. Estimating the position of a molecule also represents a crucial step in the localization-based super-resolution reconstruction of subcellular structures [12–15] because the high-resolution image of a structure is generated from the positional coordinates of individual fluorophores that label the structure.

Regardless of the particular application, it is desirable that the estimator of a given parameter is unbiased, meaning that, on average, it recovers the true value of the parameter. Moreover, it is desirable that the estimator recovers the true value with high accuracy, in the sense that estimates of the parameter obtained from repeat images of the same scene can be expected to have a distribution about the true value that is characterized by a small standard deviation. Both unbiasedness and accuracy are important [16–18] because analysis based on values returned by a biased and/or inaccurate estimator can

lead to misrepresentations of the biological process or structure being investigated [19,20].

This tutorial deals with the accuracy of parameter estimation in single molecule microscopy, specifically addressing how one can calculate, for a given parameter, a lower bound on the standard deviation with which it can be determined by any unbiased estimator. This lower bound, which is also referred to as the *limit of the accuracy* [21,22] with which the parameter can be estimated, is a practically useful quantity in two ways. First, it serves as a benchmark against which the standard deviation of a particular estimator can be evaluated, indicating how much room there might be for improvement. Second, by calculating and assessing its value under different experimental settings, the limit of accuracy can be used to design an experiment that will generate image data from which a parameter of interest can be estimated with the desired standard deviation.

The objective of the tutorial is to demonstrate how the information theory-based mathematical framework developed in [21,22] can be used to calculate a limit of accuracy. This framework is characterized by its generality, in the sense that, as it makes no assumptions about problem-specific details such as the parameters to be estimated, the mathematical description of the image of an object of interest, and the type of detector that is used to capture the image, it can be used with appropriate specifications of such details to calculate limits of accuracy for a wide variety of estimation problems. Originally illustrated with limits of accuracy calculated for the estimation of the positional coordinates of an in-focus molecule that is imaged by a charge-coupled device (CCD) detector [21,22], the framework has subsequently been utilized, for example, to compute limits of accuracy for the localization of an out-of-focus molecule [8,23,24], the determination of the distance separating two molecules [25,26], and the estimation of parameters from image data produced by an electron-multiplying CCD (EMCCD) detector [27].

The limit of the accuracy for estimating a parameter is given by the square root of the Cramér–Rao lower bound [28,29] on the variance with which the parameter can be determined by any unbiased estimator. The Cramér–Rao lower bound is a well-known result from estimation theory, and its calculation requires a stochastic description of the data from which the parameter of interest is to be estimated and the computation of the Fisher information matrix [28,29] corresponding to the data description. The Fisher information matrix is essentially a quantity that provides a measure of the amount of information that the acquired data contains about the parameter that one wishes to estimate from the data. The greater the amount of information that the data carries about the parameter, the higher the accuracy (i.e., the smaller the standard deviation) with which the parameter can be estimated. The amount of information is determined by considering how the likelihood of observing the acquired data, which is stochastic in nature, changes with the value of the parameter of interest. If the likelihood of the data does not vary significantly with the value of the parameter, then the data contains relatively little information about the parameter. On the other hand, if the likelihood of the data is sensitive to changes in the value of the parameter, then the data carries a relatively large amount

of information about the parameter and will allow it to be estimated with relatively high accuracy. Given the Fisher information matrix, the Cramér–Rao lower bound corresponding to the parameter of interest is obtained through the inverse of the matrix, reflecting the expectation that a larger amount of information about the parameter should result in a smaller bound on the variance with which the parameter can be estimated.

The mathematical framework of [21,22] essentially provides the foundation for calculating the Fisher information matrix pertaining to image data generated by single molecule microscopy and, more generally, fluorescence microscopy experiments. Since the introduction of the framework, the use of Fisher information-based limits of accuracy has become prevalent in the field of single molecule microscopy. Different groups have, for example, computed Fisher information-based limits of accuracy for molecule localization involving different point spread functions that describe the image of the molecule [20,30–35], for the estimation of the orientation of a fluorescent dipole emitter [33,35–37], and for the determination of the diffusion coefficient based on the recorded trajectory of a molecule [38,39]. A method also has been demonstrated that generates information-rich point spread functions via the optimization of limits of the localization accuracy [40].

Note that, in the context of the localization of a molecule, another commonly used benchmark for assessing the accuracy of an estimator is described in [41]. This accuracy benchmark does not make use of Fisher information and is derived based on the least-squares minimization of errors. It has been adapted in various ways [42,43], such as for the localization of a molecule from image data produced by an EMCCD detector. A similar accuracy benchmark [44] also has been derived that takes into account the effect of the diffusional movement of the molecule to be localized. For a comparison of the benchmark of [41] and the Fisher information-based limit of the localization accuracy, see [17].

A limit of accuracy is computed for a specific experimental setting and a specific estimation problem. It is therefore a quantity that depends on a multitude of factors, including the value of the parameter of interest itself (e.g., the value of a molecule's positional coordinate, or the value of the distance separating two molecules), the number of photons detected from the object of interest, the number of photons detected from a background component, the wavelength of the detected photons, the numerical aperture of the objective lens used to collect the photons, the lateral magnification of the microscopy system, the pixel size of the detector used to capture the image, the type and amount of noise introduced by the detector, and even the size of the pixel array from which the parameter of interest is estimated. Importantly, a limit of accuracy also depends, as alluded to above, on the mathematical description of the image of the object of interest, which should in principle mirror reality as closely as possible. In the case of a single molecule, for example, a point spread function should be used that appropriately models the observed image of the molecule. Its dependence on many details makes the limit of accuracy a valuable tool for experimental design because one can determine the effect of a particular experimental parameter or a particular attribute of the object of interest by computing the limit of

accuracy for different values of the parameter or attribute. At the same time, it makes it a difficult task to fully illustrate the effect of any one experimental parameter or any one object attribute, given that the effect of a parameter or attribute typically has to be considered in combination with the values of the other parameters and attributes. This tutorial will nevertheless attempt to provide a relatively comprehensive view and discussion of the general effects of various parameters and attributes and, in doing so, demonstrate the flexibility of the mathematical framework, which enables the investigation of these effects in the context of different estimation problems. With an aim toward accessibility, emphasis will be placed on the usage rather than the more technical aspects of the mathematical framework. For a rigorous treatment of the framework, the reader is referred to [22].

This tutorial is organized as follows. In Section 2, the modeling of the photon detection process that underlies all fluorescence microscopy image data is described. In Section 3, different image data models are presented that correspond to different detector types, including ones that represent the commonly used CCD [45], EMCCD [46,47], and complementary metal-oxide-semiconductor (CMOS) [48,49] detectors. In Section 4, Fisher information expressions corresponding to different image data models are presented, and limits of accuracy computed using these expressions are used to illustrate the effects of factors such as the detected photon count, the effective pixel size of the acquired image, and detector noise. In Section 5, the computation of limits of accuracy is extended to the case where a parameter of interest is estimated from data that comprises multiple images. Throughout the tutorial, the commonly encountered estimation problem of single molecule localization is used as the primary example for illustrative purposes.

2. MODELING OF PHOTON DETECTION

Before the limit of the accuracy for estimating a parameter can be calculated, a mathematical description of the acquired image data is needed. Different image data models, both hypothetical and practical, can be obtained based on assumptions made about the nature of the image detector and the noise sources that corrupt the photon signal detected from the object of interest. All image data models, however, are built upon the same description of the underlying process by which photons are detected from the object of interest. Therefore, before delving into the different data models for which limits of accuracy can be calculated, it is necessary to consider this photon detection process.

A. Photon Detection as a Spatiotemporal Random Process

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. As photon emission by a fluorescent object is typically assumed to follow a Poisson process, the temporal component of the detection of the emitted photons is accordingly assumed to be a Poisson process. The intensity function that characterizes this Poisson process is referred to as the *photon detection rate* [22] and is given generally by the function of time $\Lambda_\theta(\tau)$, $\tau \geq t_0$, where $t_0 \in \mathbb{R}$ is some arbitrary time

point, and $\theta \in \Theta$ is the vector of parameters to be estimated from the resulting image data, with Θ denoting the parameter space. A specific definition of Λ_θ might, for example, be an exponential decay function that models the photobleaching of the object of interest, or a constant function when the object has relatively high photostability. It is understood that, in its definition, Λ_θ takes into account the loss of photons as determined by the optical efficiency of the imaging setup's detection system, which comprises the microscope's detection optics and the image detector. Alternatively, Λ_θ can be equivalently expressed in terms of the object's photon emission rate and an explicit optical efficiency parameter [21].

Note that, throughout this tutorial, the subscript θ is used to denote the fact that the given function or random variable can, in general, depend on the vector θ of parameters to be estimated. In the case of a function, this simply means that the function is potentially expressed in terms of parameters that are to be estimated from the image data. For example, in the case where the photon detection rate Λ_θ is modeled with an exponential decay function, the decay constant might be a parameter of interest that is included in θ . In the case where it is modeled with a constant function, the constant detection rate itself might be a parameter of interest and therefore a component of θ . For the purposes and illustrations in this section and Section 3, a precise definition for θ is not critical, so long as it is understood that the given function or random variable depends on θ . A precise definition becomes crucial in the illustrations of Sections 4 and 5, however, as the calculation of limits of accuracy requires that a Fisher information matrix is derived for a specifically defined θ . Note also that, in the specific examples of Sections 4 and 5, the subscript θ will be dropped from the notation for functions such as Λ_θ without further explanation whenever the function does not depend on θ as defined for the given estimation problem.

The spatial component of the photon detection process is specified by a 2D probability density function that describes the distribution of the location at which a photon emitted by the object of interest is detected. Referred to as a *photon distribution profile* [22], this density function is given generally by $f_{\theta,\tau}(x,y)$, $(x,y) \in \mathbb{R}^2$, $\theta \in \Theta$, $\tau \geq t_0$, where the dependence on τ denotes that, in general, the spatial density of photon detection varies with time, as in the case of a moving object of interest. To use more accurate language, $f_{\theta,\tau}$ as given here specifies the distribution at time τ of the location at which a photon is detected only in the case of a detector with an infinitely large detection area. When the detector has a finite detection area, $f_{\theta,\tau}$ specifies the distribution at time τ of the location at which a photon impacts the detector plane because, in this case, only a photon that falls within the finite detection area is actually detected.

The specific definition of $f_{\theta,\tau}$ is based on the image of the object of interest as observed using the imaging setup. By the typical approximation of the optical microscope as a laterally shift-invariant system [50], $f_{\theta,\tau}$ has the form

$$f_{\theta,\tau}(x,y) = \frac{1}{M^2} q_{z_0,\tau} \left(\frac{x}{M} - x_{0,\tau}, \frac{y}{M} - y_{0,\tau} \right), \quad (1)$$

$(x,y) \in \mathbb{R}^2$, $\theta \in \Theta$, $\tau \geq t_0$, where $M > 0$ is the lateral magnification of the imaging setup, $(x_{0,\tau}, y_{0,\tau}, z_{0,\tau})$ are the object

space positional coordinates of the object at time τ , and $q_{z_0}(x, y)$, $(x, y) \in \mathbb{R}^2$, $z_0 \in \mathbb{R}$, 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. The function q_{z_0} is referred to as an *image function* [22] and has the additional property that it integrates to one over \mathbb{R}^2 . By the assumption of lateral shift invariance, the photon distribution profile is thus a scaled and laterally shifted version of the object's unmagnified image when the object is laterally located at the origin of the object space coordinate system.

The image function q_{z_0} can, in principle, be defined for any type of object, typically as the convolution of the object with the point spread function of the microscope. In [51], for example, such a definition is given for the image function of a fluorescent microsphere. For a point-like object such as a single fluorescent protein or dye molecule, the image function is usually given by just the point spread function itself. A typical image function for an in-focus molecule is the Airy pattern from optical diffraction theory [52]. Dropping the subscript z_0 from the notation as the in-focus condition confines the molecule to the microscope's focal plane, the Airy pattern image function can be written as

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

where J_1 is the first-order Bessel function of the first kind, n_a is the numerical aperture of the objective lens, and λ is the wavelength of the photons detected from the molecule. Another commonly used point spread function model for the in-focus scenario is the 2D Gaussian function [16,41,53,54], and the corresponding image function is given by

$$q(x, y) = \frac{1}{2\pi\sigma_g^2} \cdot e^{-\frac{x^2+y^2}{2\sigma_g^2}}, \quad (x, y) \in \mathbb{R}^2, \quad (3)$$

where $\sigma_g > 0$ is the standard deviation of the component 1D Gaussian functions in the x and y directions.

For the out-of-focus scenario, the image function is given instead by a 3D point spread function. The image function corresponding to the classical point spread function of Born and Wolf [52], for example, can be written as

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

$(x, y) \in \mathbb{R}^2$, $z_0 \in \mathbb{R}$, where J_0 is the zeroth-order Bessel function of the first kind, n_a and λ are as defined in Eq. (2), and n is the refractive index of the objective lens immersion medium.

Note that the temporal and spatial components of the photon detection process are assumed to be mutually independent. Moreover, in the spatial component, the locations at which photons impact the detector plane are assumed to be mutually independent.

B. Superposition of Independent Photon Detection Processes

The spatiotemporal random process described in Section 2.A pertains to the detection of photons from a single object of interest. The formalism, however, can be extended in a straightforward manner to include the simultaneous but independent

detection of photons from other sources, such as a second object of interest or a background component like the autofluorescence of a cellular sample. Given two independent photon detection processes characterized by the photon detection rates Λ_θ^1 and Λ_θ^2 and photon distribution profiles $f_{\theta,\tau}^1$ and $f_{\theta,\tau}^2$, $\tau \geq t_0$, their superposition [22] is simply the photon detection process with photon detection rate

$$\Lambda_\theta(\tau) = \Lambda_\theta^1(\tau) + \Lambda_\theta^2(\tau), \quad \theta \in \Theta, \quad \tau \geq t_0, \quad (5)$$

and photon distribution profile

$$f_{\theta,\tau}(x, y) = \frac{\Lambda_\theta^1(\tau)}{\Lambda_\theta(\tau)} f_{\theta,\tau}^1(x, y) + \frac{\Lambda_\theta^2(\tau)}{\Lambda_\theta(\tau)} f_{\theta,\tau}^2(x, y), \quad (6)$$

$(x, y) \in \mathbb{R}^2$, $\theta \in \Theta$, $\tau \geq t_0$. To give a more specific example for $f_{\theta,\tau}$, suppose photons are detected from two in-focus and stationary objects of interest located at (x_{01}, y_{01}) and (x_{02}, y_{02}) . Using the expression of Eq. (1) without the dependence on time and z location for both $f_{\theta,\tau}^1$ and $f_{\theta,\tau}^2$ in Eq. (6), the resulting photon distribution profile for the superposition of the two photon detection processes is given by

$$f_{\theta,\tau}(x, y) = \frac{1}{M^2} \left(\frac{\Lambda_\theta^1(\tau)}{\Lambda_\theta(\tau)} q_1\left(\frac{x}{M} - x_{01}, \frac{y}{M} - y_{01}\right) + \frac{\Lambda_\theta^2(\tau)}{\Lambda_\theta(\tau)} q_2\left(\frac{x}{M} - x_{02}, \frac{y}{M} - y_{02}\right) \right), \quad (7)$$

$(x, y) \in \mathbb{R}^2$, $\theta \in \Theta$, $\tau \geq t_0$, where q_1 and q_2 are the image functions for the two objects.

3. IMAGE DATA MODELS

Building on the mathematical description of the photon detection process in Section 2, image data models can be obtained that differ in terms of the size of the detector, whether the detector is pixelated, and the types of noise that the detector introduces. The models also can differ in terms of whether a background component exists that introduces additional noise to the acquired data. Many different models are possible by taking different combinations of detector attributes and noise sources. However, they can be subsumed under four main models, namely, the fundamental data model, the Poisson data model, the CCD/CMOS data model, and the EMCCD data model. The first two models are hypothetical in nature, as they assume the use of detectors that do not exist. As will be seen in Section 4, however, they are extremely useful in that they represent ideal imaging scenarios against which the practical imaging scenarios represented by the CCD/CMOS and EMCCD data models can be compared.

A. Fundamental Data Model

The fundamental data model describes image data that is acquired under the most ideal of imaging scenarios. It assumes the use of an image detector that has an unpixelated and infinitely large photon detection area. The absence of pixelation means that the location at which a photon is detected is recorded with an arbitrarily high precision, and the infinitely large area ensures that no photon is undetected by virtue of impacting the detector plane at a location outside of the detector. The fundamental data model further assumes that the detector does not introduce noise of any kind and that there is no

background component. Therefore, the photon signal detected from the object(s) of interest is not in any way corrupted by noise.

In the fundamental data model, the data acquired over the time interval $[t_0, t]$ consists of the time points of detection $\{\tau_1, \dots, \tau_{N_0}\}$ and the locations of detection $\{(x_1, y_1), \dots, (x_{N_0}, y_{N_0})\}$ of the N_0 photons that are detected during the interval. As presented in Section 2.A, the time points are distributed according to the Poisson process with intensity function $\Lambda_\theta(\tau)$, $\tau \geq t_0$, $\theta \in \Theta$, and for $k = 1, 2, \dots, N_0$, the positional coordinate pair (x_k, y_k) is a realization of the bivariate random variable distributed according to the spatial density function $f_{\theta, \tau_k}(x, y)$, $(x, y) \in \mathbb{R}^2$, $\theta \in \Theta$, $\tau \geq t_0$.

For a concrete example of a fundamental image, consider the relatively simple case in which the positional coordinates of a stationary molecule are to be estimated. In this specific scenario, the time points of photon detection need not be simulated, as they do not play a role in the estimation of the molecule's positional coordinates. Instead, only the number of detected photons, N_0 , needs to be determined as a realization of the Poisson random variable with mean $\int_{t_0}^t \Lambda_\theta(\tau) d\tau$. The location of each of the N_0 photons is then generated as a realization of the bivariate random variable with probability density f_θ given by $f_{\theta, \tau}$ of Eq. (1) without the dependence on time τ . Assuming the molecule to be in-focus and to have an image described by the Airy pattern of Eq. (2), an example of f_θ is shown in Fig. 1(a), and an image simulated as described is shown in Fig. 1(b). For the simulation, the photon detection rate is assumed to be a constant function $\Lambda_\theta(\tau) = \Lambda_0$, $\tau \geq t_0$, where Λ_0 is a positive constant.

The most important property that distinguishes the fundamental data model from the data models that follow is the assumption of an unpixelated detection area that allows the locations of the detected photons to be recorded with arbitrarily high precision. All variations of the fundamental data model will therefore retain this property. A variation can be obtained, for example, by introducing a background component using the superposition property described in Section 2.B or by imposing a finite photon detection area. Variations of the fundamental data model will be used in Section 4 to help illustrate the different levels of accuracy benchmarking that are possible.

B. Poisson Data Model

The Poisson data model describes image data that is acquired using an image detector with a pixelated and finite photon detection area. It also allows for the existence of a background component that corrupts the photon signal detected from the object(s) of interest. In these regards, the Poisson data model considers an imaging scenario that is significantly closer to what is encountered in practice than the scenario considered by the fundamental data model. Indeed, due to these realistic assumptions, it has generally been characterized as a practical data model (e.g., [17,25,26]). Strictly speaking, however, the Poisson data model is hypothetical, as it makes the impractical assumption that the detector does not introduce noise when reading out the electrons accumulated in its pixels.

Despite its hypothetical nature, the Poisson data model can be a reasonable model in cases where the number of photons

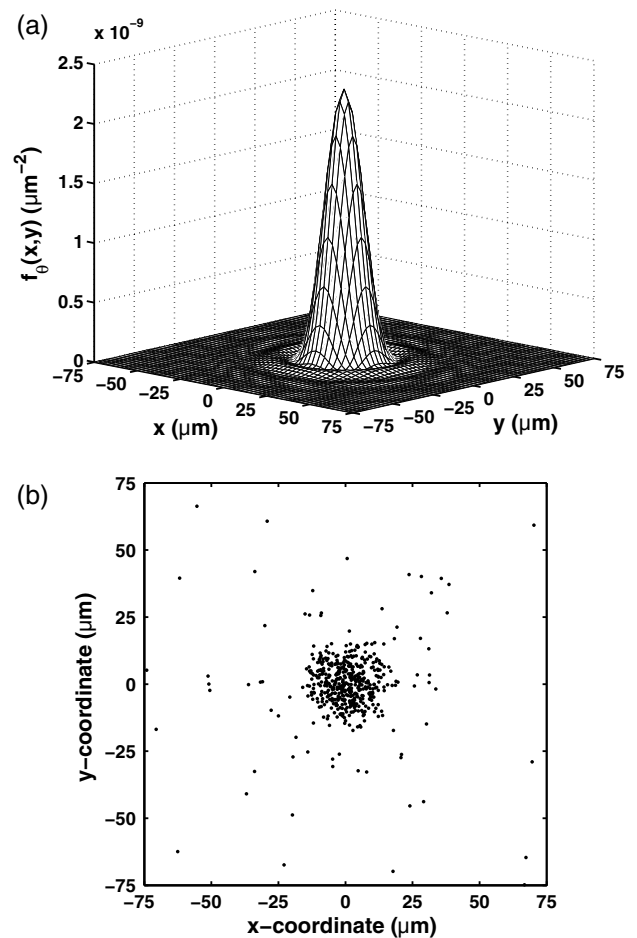


Fig. 1. Fundamental data model example. (a) Photon distribution profile f_θ for an in-focus and stationary molecule whose image is modeled with the Airy pattern. This profile assumes that the molecule is located at $(x_0, y_0) = (0 \text{ nm}, 0 \text{ nm})$ in the object space and that photons of wavelength $\lambda = 520 \text{ nm}$ are collected using an objective lens with a numerical aperture of $n_a = 1.4$ and a magnification of $M = 100$. While the profile is defined over the detector plane \mathbb{R}^2 , the plot shows its values over the $150 \mu\text{m} \times 150 \mu\text{m}$ region over which it is centered. (b) A fundamental image simulated according to the profile in (a), with each dot representing the location at which a photon is detected. The simulation assumes a constant photon detection rate of $\Lambda_0 = 10,000$ photons/s and an acquisition time of $t - t_0 = 0.05 \text{ s}$, such that the mean photon count in an image is $\Lambda_0 \cdot (t - t_0) = 500$. In this particular realization, $N_0 = 516$ photons are detected, of which 492 fall within the $150 \mu\text{m} \times 150 \mu\text{m}$ region shown, over which the image is centered.

detected in each pixel from the object(s) of interest and background component is generally expected to be large compared with the readout noise in the pixel. (For examples of experimental data analysis based on the Poisson data model, see [20,30,34,55].)

According to the Poisson data model, a K -pixel image acquired over the time interval $[t_0, t]$ consists of the numbers of electrons z_1, \dots, z_K accumulated in the K pixels during the interval. For $k = 1, \dots, K$, the electron count z_k in the k th pixel is a realization of the random variable $H_{\theta,k}$ given by

$$H_{\theta,k} = S_{\theta,k} + B_{\theta,k}, \quad \theta \in \Theta, \quad (8)$$

where $S_{\theta,k}$ and $B_{\theta,k}$ are independent Poisson random variables representing quantities resulting from independent processes. The variable $S_{\theta,k}$ represents the number of photoelectrons in the k th pixel that result from the detection of photons from the object(s) of interest and is distributed with mean

$$\mu_{\theta,k} = \int_{t_0}^t \Lambda_{\theta}(\tau) \int_{C_k} f_{\theta,\tau}(x,y) dx dy d\tau, \quad \theta \in \Theta, \quad (9)$$

where C_k is the region in the detector plane corresponding to the k th pixel, and $\Lambda_{\theta}(\tau)$ and $f_{\theta,\tau}$ are the rate of photon detection and the profile of photon distribution, as described in Section 2.A. This expression for the mean photoelectron count in the k th pixel due to the object(s) of interest is intuitively the fraction of the photon distribution profile over the k th pixel scaled by the mean number of photons that impact the detector plane over the acquisition interval. The variable $B_{\theta,k}$ represents the number of electrons in the k th pixel that arise from a background component, which, beside sources of spurious photons such as the sample's autofluorescence, can more generally include sources of spurious electrons such as the detector's dark current. It is distributed with mean

$$\beta_{\theta,k} = \int_{t_0}^t \Lambda_{\theta}^b(\tau) \int_{C_k} f_{\theta,\tau}^b(x,y) dx dy d\tau, \quad \theta \in \Theta, \quad (10)$$

which is analogous in form to Eq. (9), with $\Lambda_{\theta}^b(\tau)$, $\tau \geq t_0$, and $f_{\theta,\tau}^b(x,y)$, $(x,y) \in \mathbb{R}^2$, $\tau \geq t_0$, denoting the detection rate and distribution profile of photons from the background component. Taking some liberty with terminology for ease of presentation, Λ_{θ}^b and $f_{\theta,\tau}^b$ are referred to as a photon detection rate and a photon distribution profile despite the fact that $B_{\theta,k}$ can include electrons that do not result from the detection of photons. This is justified because $B_{\theta,k}$ is primarily made up of photoelectrons resulting from the detection of spurious photons, given that spurious electrons that originate from the detector, such as those arising from the dark current, typically represent a negligible portion of the background component. Henceforth, with the understanding that they can consist, in small part, of spurious electrons, $B_{\theta,k}$ will accordingly be viewed as representing the background photoelectron count in the k th pixel, and $\beta_{\theta,k}$ the mean of the background photoelectron count in the k th pixel.

As $H_{\theta,k}$ of Eq. (8) is the sum of two independent Poisson random variables, it is itself a Poisson random variable, with mean given by

$$\nu_{\theta,k} = \mu_{\theta,k} + \beta_{\theta,k}, \quad \theta \in \Theta, \quad (11)$$

the sum of Eqs. (9) and (10). Note that, by applying the result of Section 2.B, Eq. (11) also can be derived from a single detection process consisting of the superposition of the independent processes of object and background photon detection.

The function $\nu_{\theta,k}$ of Eq. (11) is thus the mean of the photoelectron count in the k th pixel. The terminology "photoelectrons" is used to distinguish electrons in a given pixel that are generated by the detection of photons and accumulated in the pixel during the image acquisition interval from electrons in the pixel that arise from the detector's noise processes. This distinction will help with the clarity of presentation in the

remainder of this tutorial. Note also that, in this tutorial, each detected photon is assumed, based on the physics of photon-to-electron conversion, to result in the generation of one photoelectron.

In Fig. 2(a), the mean image of the in-focus and stationary molecule of Fig. 1, computed over an 11×11 -pixel region using Eq. (11), is shown. For this mean image, the spatial distribution

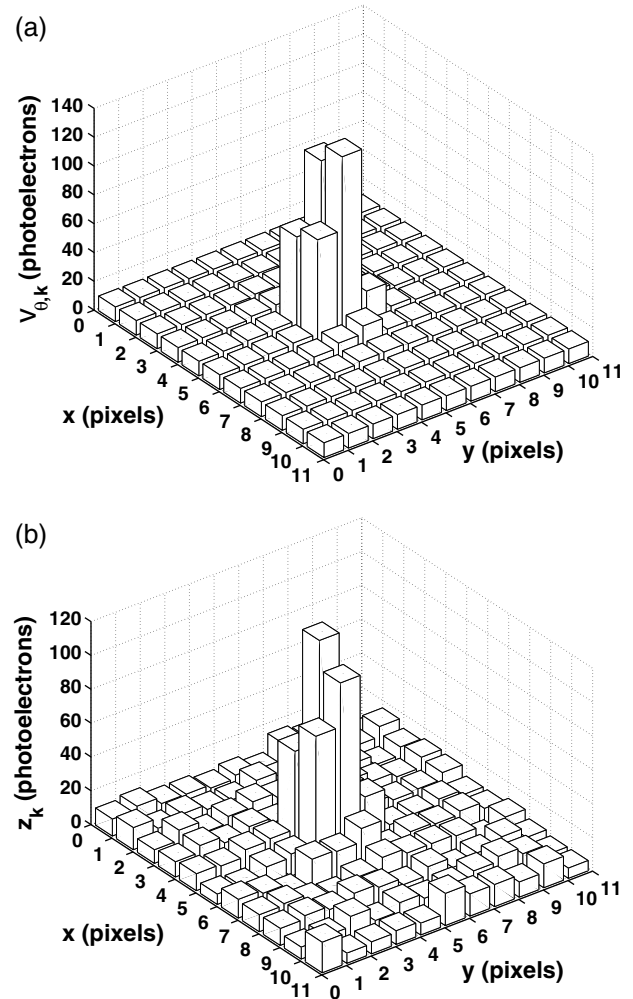


Fig. 2. Poisson data model example. (a) The mean pixelated image of the in-focus and stationary molecule of Fig. 1, computed over an 11×11 -pixel region with a pixel size of $13 \mu\text{m} \times 13 \mu\text{m}$ and a mean background photoelectron count of $\beta_{\theta,0} = 10$ per pixel (obtained as the product of the constant background photon detection rate $\Lambda_{\theta}^b = 24,200$ photons/s and the uniform spatial distribution $f_{\theta}^b(x,y) = 1/20,449 \mu\text{m}^{-2}$, integrated over each pixel and the acquisition time of $t - t_0 = 0.05$ s). The molecule is located at $(x_0, y_0) = (656.5 \text{ nm}, 682.5 \text{ nm})$ in the object space, which corresponds to (5.05 pixels, 5.25 pixels) in the image space, assuming (0, 0) in both spaces coincides with the upper-left corner of the 11×11 -pixel region. In terms of the photoelectron count due to the molecule, the mean image contains 476.15 out of the total of $\Lambda_{\theta} \cdot (t - t_0) = 500$ mean number of photoelectrons distributed over the detector plane. All other relevant parameters are as specified in Fig. 1. (b) A Poisson realization of the mean image in (a). The photoelectron count in each pixel is drawn from the Poisson distribution with mean given by the photoelectron count in the corresponding pixel in (a).

of the background photons is assumed to be uniform over the 11×11 -pixel region and to not depend on time. Specifically, it is defined as $f_{\theta}^b(x, y) = 1/A$ for (x, y) within the 11×11 -pixel region with area A , and $f_{\theta}^b(x, y) = 0$ otherwise. Moreover, the detection rate of the background photons is, as in the case of the molecule's photon detection rate, assumed to be a constant function. It is defined as $\Lambda_{\theta}^b(\tau) = \Lambda_0^b$, $\tau \geq t_0$, where Λ_0^b is a positive constant. Every pixel in the mean image therefore has the same mean background photoelectron count, and Eq. (10) simplifies to $\beta_{\theta, k} = \Lambda_0^b \cdot (t - t_0) \cdot A_{\text{pixel}}/A = \beta_{\theta, 0}$, $k = 1, \dots, K$, where A_{pixel} denotes the area of a pixel. In Fig. 2(b), a Poisson realization of the mean image is shown, providing an example of an image simulated according to the Poisson data model. The Poisson noise is particularly evident in the fluctuations of the photoelectron counts in the peripheral pixels, whose mean photoelectron counts are, in contrast, nearly identical, as seen in Fig. 2(a), by virtue of the signal from the molecule being small compared with the uniform background in the outer portions of the 11×11 -pixel region.

C. CCD/CMOS Data Model

The CCD/CMOS data model describes image data that is acquired using a CCD or CMOS detector. It builds on the Poisson data model by adding a component that accounts for the noise introduced by the detector when the photoelectrons accumulated in a pixel are read out. This additional component to the acquired data represents the primary noise source associated with a CCD or CMOS detector, and it is typically assumed to be Gaussian distributed. A K -pixel CCD or CMOS image acquired over the time interval $[t_0, t]$ therefore comprises the electron counts z_1, \dots, z_K , where the count z_k in the k th pixel, $k = 1, \dots, K$, is a realization of the random variable $H_{\theta, k}$ given by

$$H_{\theta, k} = S_{\theta, k} + B_{\theta, k} + W_k, \quad \theta \in \Theta. \quad (12)$$

In Eq. (12), $S_{\theta, k}$ and $B_{\theta, k}$ are the same Poisson random variables from Eq. (8) and W_k is a Gaussian random variable with mean η_k and variance σ_k^2 that represents the number of electrons due to the readout process of the detector. The three random variables are mutually independent, as they represent quantities resulting from independent processes. Note that the mean η_k of W_k can be used to model the detector offset for the k th pixel. Also, though in principle W_k can depend on the parameter vector θ , here the assumption is that η_k (the detector offset) and σ_k^2 are either known or can be separately determined. (For more on the determination of the detector offset and the readout noise variance, see Section 3.E.)

Based on Eq. (12), the electron count z_k in the k th pixel of a CCD or CMOS image is just the sum of a realization of a Poisson random variable with mean $\nu_{\theta, k}$ given by Eq. (11) and a realization of a Gaussian random variable with mean η_k and variance σ_k^2 . In Fig. 3(a), a CCD realization of the mean image of Fig. 2(a) is shown. This particular realization is generated by simply adding to each pixel of the Poisson realization of Fig. 2(b) a random number drawn from the Gaussian distribution with mean $\eta_0 = 0$ electrons and standard deviation $\sigma_0 = 8$ electrons. This image is a CCD realization because the same Gaussian distribution with parameters η_0 and σ_0 is used to simulate the readout noise in every pixel (i.e., $\eta_k = \eta_0$ and

$\sigma_k = \sigma_0$, $k = 1, \dots, K$), reflecting the fact that all pixels in a CCD detector share the same readout circuitry. To account for the fact that each pixel in a CMOS detector has its own dedicated readout circuitry, a CMOS realization can be obtained by using Gaussian distributions with different mean and standard deviation values to simulate the readout noise in different pixels. (For a study that accounts for the pixel-dependent readout noise of a CMOS detector, see [56].) Note that in Fig. 3(a), a view is chosen to ensure that a few negative electron counts can be seen in the image. A negative count is obtained in a pixel whenever the Gaussian readout noise is negative and is greater in magnitude than the Poisson component of the data. In experimental data analysis, this can occur after the typical preprocessing step of detector offset subtraction (which is equivalent to setting η_0 to 0 electrons) and subsequent multiplication by the digital count-to-electron count conversion factor (see Section 3.E for more

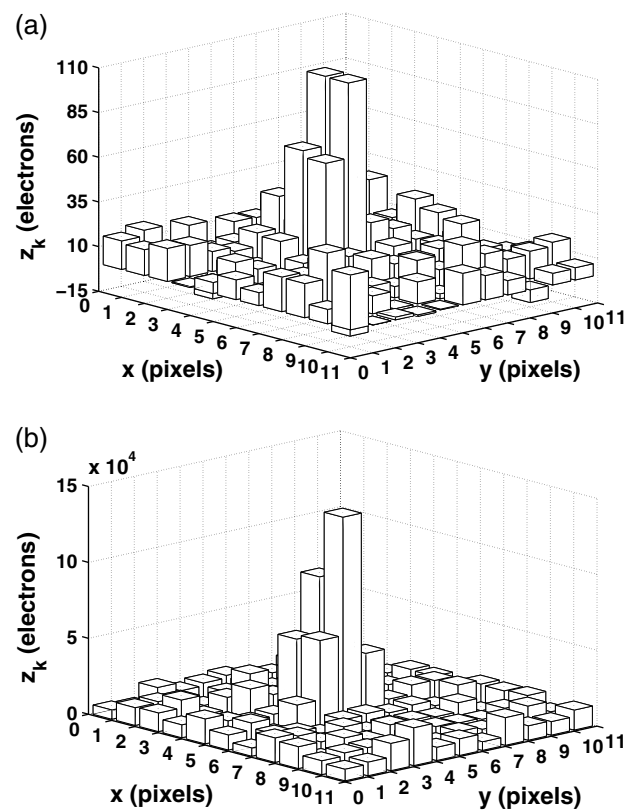


Fig. 3. CCD data model and EMCCD data model examples. (a) A CCD realization of the mean image in Fig. 2(a). The electron count in each pixel is the sum of a random number drawn from the Poisson distribution with mean given by the photoelectron count in the corresponding pixel in Fig. 2(a), and a random number drawn from the Gaussian distribution with mean $\eta_0 = 0$ electrons and standard deviation $\sigma_0 = 8$ electrons. In this particular realization, the Poisson random numbers are taken directly from the Poisson realization of Fig. 2(b). (b) An EMCCD realization of the mean image in Fig. 2(a). The electron count in each pixel is the sum of a random number drawn from the probability distribution of Eq. (14) with electron multiplication gain $g = 950$ and $\nu_{\theta, k}$ given by the photoelectron count in the corresponding pixel in Fig. 2(a), and a random number drawn from the Gaussian distribution with mean $\eta_0 = 0$ electrons and standard deviation $\sigma_0 = 24$ electrons.

on the modeling of experimental data). Note that a negative electron count is a reflection of the detector's measurement error when reading out a small number of electrons in a pixel and does not indicate an actual occurrence of a negative number of electrons.

D. EMCCD Data Model

The EMCCD data model describes image data that is acquired using an EMCCD detector. It is similar to the CCD/CMOS data model in that the data in a given pixel is modeled as the sum of three components corresponding to the three independent processes of object detection, background detection, and detector readout. The difference, however, lies in the fact that the electron counts corresponding to the object(s) of interest and the background component are no longer the Poisson-distributed numbers of photoelectrons resulting directly from their respective detection processes. Instead, they are stochastically amplified versions of those Poisson-distributed photoelectron counts and are, accordingly, modeled with random variables having probability distributions that account for the stochastic amplification. This multiplication of the photoelectrons accumulated in a pixel is what distinguishes an EMCCD detector from a CCD detector, and it importantly enables imaging under low-light conditions by rendering the readout noise small in comparison with the augmented electron count.

A K -pixel EMCCD image acquired over the time interval $[t_0, t]$ thus consists of the electron counts z_1, \dots, z_K , where the count z_k in the k th pixel, $k = 1, \dots, K$, is a realization of the random variable $H_{\theta,k}$ given by

$$H_{\theta,k} = S_{\theta,k}^a + B_{\theta,k}^a + W_k, \quad \theta \in \Theta. \quad (13)$$

As in the CCD/CMOS data model of Eq. (12), W_k in Eq. (13) is a Gaussian random variable with mean η_k and variance σ_k^2 that models the electron count contributed by the detector's readout process. The random variables $S_{\theta,k}^a$ and $B_{\theta,k}^a$, on the other hand, respectively represent the results of the stochastic amplification of the photoelectron counts modeled by the Poisson random variables $S_{\theta,k}$ and $B_{\theta,k}$ in Eq. (12). Depending on the model of amplification, different probability distributions can be used to describe $S_{\theta,k}^a$ and $B_{\theta,k}^a$. Assuming, however, that the photoelectrons accumulated in a pixel are multiplied according to a branching process [57] that generates, for each input electron in each multiplication stage, a geometrically distributed number of electrons that includes the input electron itself, the sum random variable $S_{\theta,k}^a + B_{\theta,k}^a$ has the probability mass function [27]

$$p_{\theta,k}^a(l) = \begin{cases} e^{-\nu_{\theta,k}}, & l = 0, \\ e^{-\nu_{\theta,k}} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j} \left(1 - \frac{1}{g}\right)^{l-1-j} \left(\frac{\nu_{\theta,k}}{g}\right)^{j+1}}{(j+1)!}, & l = 1, 2, \dots, \end{cases} \quad (14)$$

$\theta \in \Theta$, where $\binom{l-1}{j}$ denotes " $(l-1)$ choose j ," the binomial coefficient indexed by $l-1$ and j , $\nu_{\theta,k}$ is as defined in Eq. (11) and is the mean of the Poisson random variable $S_{\theta,k} + B_{\theta,k}$ representing the photoelectron count in the k th pixel prior to amplification, and $g > 1$ is the electron multiplication gain or, more technically, the mean gain of the branching process,

defined as the mean number of electrons that result from the multiplication of a single initial electron. Note that Eq. (14) is obtained by treating the sum $S_{\theta,k}^a + B_{\theta,k}^a$ as a single amplified electron count having originated from the single Poisson-distributed photoelectron count $S_{\theta,k} + B_{\theta,k}$ with mean $\nu_{\theta,k}$. This is equivalent to considering the amplification separately for the object and background photoelectrons.

According to Eq. (13), the electron count z_k in the k th pixel of an EMCCD image is the sum of a realization of a random variable with probability mass function given by Eq. (14) and a realization of a Gaussian random variable with mean η_k and variance σ_k^2 . An EMCCD realization of the mean image of Fig. 2(a) generated using this approach is shown in Fig. 3(b). Specifically, to obtain the electron count z_k , a random number is drawn from the probability distribution of Eq. (14) with the electron multiplication gain set to a relatively high $g = 950$ and $\nu_{\theta,k}$ given by the k th pixel of the mean image of Fig. 2(a), and added to a random number drawn from the Gaussian distribution with mean $\eta_0 = 0$ electrons and standard deviation $\sigma_0 = 24$ electrons. Here, the same Gaussian distribution with parameters η_0 and σ_0 is used to simulate the readout noise in every pixel because, just as in the case of a CCD detector, all pixels in an EMCCD detector share the same readout circuitry. Note that, to reflect the typically larger standard deviation for the readout noise of an EMCCD detector, a larger σ_0 is used here compared with σ_0 used for the CCD realization of Fig. 3(a). Note also that, due to the electron multiplication, the electron counts in the EMCCD realization are much larger than those in the CCD realization. In addition, note that, in general, negative electron counts also can arise in an EMCCD image due to negative Gaussian readout noise. Negative counts are not present in the realization of Fig. 3(b), however, and in fact are unlikely for the particular scenario considered here. This is because the effect of the combination of the high electron multiplication gain of $g = 950$ and the relatively high background level of $\beta_{\theta,0} = 10$ photoelectrons per pixel is such that even a peripheral pixel in an image will likely have an amplified electron count that is larger in magnitude than any potential negative readout noise at the given level of $\sigma_0 = 24$ electrons.

It is of interest to note that when, as is typically the case, an EMCCD detector is used with a high electron multiplication gain g , the probability distribution, given a single initial electron, of a geometrically amplified electron count is well approximated by an exponential distribution with parameter $1/g$. Based on this result, for a large g the probability distribution of Eq. (14) can be replaced by the probability density function [27]

$$p_{\theta,k}^a(u) = \begin{cases} e^{-\nu_{\theta,k}}, & u = 0, \\ \frac{e^{-\left(\frac{u}{g} + \nu_{\theta,k}\right)} \sqrt{\nu_{\theta,k} u / g} I_1(2\sqrt{\nu_{\theta,k} u / g})}{u}, & u > 0, \end{cases} \quad (15)$$

$\theta \in \Theta$, where I_1 is the first-order modified Bessel function of the first kind. This density function has frequently been used to model electron multiplication in an EMCCD detector. It is found in [58], for example, in an equivalent form that does not include the jump at $u = 0$ and does not use the modified Bessel function representation. It is also found in identical form in [59] and [43].

E. Modeling of Experimental Data

In the pixelated data models of Sections 3.B, 3.C, and 3.D, the pixel values of an image are specified in units of electrons. In practice, however, a typical CCD, CMOS, or EMCCD detector produces image data with pixel values given in units of digital counts. Moreover, the value of each pixel includes a detector offset. Therefore, in order to perform data modeling based on experimentally acquired data, a unit conversion is required, and the detector offset needs to be accounted for. The unit conversion is carried out by multiplying the value of each pixel by a digital count-to-electron count conversion factor. This conversion factor usually can be found in the specification sheet of the detector; it also can be measured using a standard methodology (e.g., [60]) that exploits the linear relationship between the mean and variance of the data acquired in a pixel. Unlike the unit conversion factor, which is not accounted for in the data models, the detector offset for the k th pixel can be represented by the mean η_k of the readout noise Gaussian random variable W_k . Both the offset η_k and the readout noise variance σ_k^2 for the k th pixel can be determined from dark frames acquired with a minimal exposure time (e.g., [45]). While the same unit conversion factor, detector offset, and readout noise variance are applicable to all the pixels of a CCD or EMCCD image, the fact that each CMOS pixel has its own readout circuitry means that every pixel of a CMOS image has its own values for these three quantities. The same methodologies for determining these quantities, however, apply on a per-pixel basis. (For an example of a study in which the per-pixel determination of these quantities is carried out for the analysis of CMOS images, see [56].) Regardless of the detector type, given that the unit conversion factor(s) and the detector offset(s) have been determined for the pixels of an image, a typical way of converting the image to units of electrons consists of removing the appropriate offset from each pixel by subtraction (thereby setting η_k to 0 electrons for each pixel k) and multiplying each resulting pixel value by the appropriate conversion factor.

4. FISHER INFORMATION AND CRAMÉR–RAO LOWER BOUND

The main purpose of the material in Sections 2 and 3 is to provide a mathematical description of the image data from which parameters of interest are to be estimated. The mathematical description of the data is critical, as it allows the determination of how accurately the parameters can be estimated from the data. The description, as seen in Sections 2 and 3, includes models for the detection of photons from the object(s) of interest and a potential background component, a model for the detector used for the image acquisition, and models for the noise sources that corrupt the photon signal detected from the object(s) of interest. The parameters to be estimated can be anything from an object's positional coordinates to the level of the background noise in the image, so long as they are appropriately incorporated into the description of the data.

According to estimation theory, the limit of the accuracy, in other words the limit to the standard deviation, with which a given parameter can be determined by an unbiased estimator is given by the square root of the Cramér–Rao lower bound on

the variance for estimating the parameter. The Cramér–Rao lower bound is, in turn, given by the main diagonal element, corresponding to the parameter, of the inverse of a square matrix quantity known as the Fisher information matrix. The Fisher information matrix is a function of the parameter, and, as it measures the amount of information that the data carries about the parameter, it is calculated based on the mathematical description of the data from which the parameter is to be estimated.

More formally, the Cramér–Rao inequality [28,29] states that the covariance matrix of any unbiased estimator $\hat{\theta}$ of the vector of parameters $\theta \in \Theta$, where Θ is the parameter space, is no smaller than the inverse of the Fisher information matrix $\mathbf{I}(\theta)$. Mathematically, this inequality is given by

$$\text{Cov}(\hat{\theta}) \geq \mathbf{I}^{-1}(\theta), \quad (16)$$

where for $\theta = (\theta_1, \theta_2, \dots, \theta_N) \in \Theta$ and $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_N)$, where N is a positive integer, both the covariance matrix $\text{Cov}(\hat{\theta})$ and the inverse Fisher information matrix $\mathbf{I}^{-1}(\theta)$ are $N \times N$ matrices. While the matrix inequality of Eq. (16) does not imply that each element of $\text{Cov}(\hat{\theta})$ is greater than or equal to its corresponding element in $\mathbf{I}^{-1}(\theta)$, it does imply that each main diagonal element of $\text{Cov}(\hat{\theta})$ is greater than or equal to its corresponding element in $\mathbf{I}^{-1}(\theta)$. This arises from the fact that the matrix $\text{Cov}(\hat{\theta}) - \mathbf{I}^{-1}(\theta)$ is positive semidefinite and, importantly, means that the variance of the estimate $\hat{\theta}_i$ of the parameter θ_i , $i = 1, \dots, N$, is bounded from below by the i th main diagonal element of the inverse Fisher information matrix. Mathematically, this can be expressed as

$$\text{Var}(\hat{\theta}_i) \geq [\mathbf{I}^{-1}(\theta)]_{ii} \geq ([\mathbf{I}(\theta)]_{ii})^{-1}, \quad i = 1, \dots, N, \quad (17)$$

where the subscript ii denotes the i th main diagonal element. It is then straightforward to see that the limit to the standard deviation with which parameter θ_i can be estimated is given by $\sqrt{[\mathbf{I}^{-1}(\theta)]_{ii}}$. Note that the second inequality in Eq. (17) implies that the square root of the inverse of the i th main diagonal element of the Fisher information matrix provides a potentially looser lower bound on the standard deviation of $\hat{\theta}_i$.

It is thus clear that the key to determining the limit of the accuracy for the estimation of a parameter lies with the calculation of the Fisher information matrix $\mathbf{I}(\theta)$. In order to calculate $\mathbf{I}(\theta)$, the likelihood function for θ , which is just the probability distribution of the data viewed as a function of θ , is needed. Given the data w with probability distribution $p_\theta(w)$, the Fisher information matrix is given by (e.g., [29])

$$\mathbf{I}(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ln p_\theta(w) \right)^T \left(\frac{\partial}{\partial \theta} \ln p_\theta(w) \right) \right], \quad \theta \in \Theta, \quad (18)$$

where T denotes the transpose. As mentioned in Section 1, the calculation of the Fisher information matrix is based on assessing how the likelihood of the observed data is affected by changes in the values of the parameters of interest. Mathematically, this assessment is realized by considering, as seen in Eq. (18), the derivative of the logarithm of the likelihood function with respect to the parameter vector θ . The expectation is taken over all possible values of the data, so that the

Fisher information is an indicator of how, on average, the likelihood function changes with θ .

Due to the differences in their mathematical descriptions, the image data models of Section 3 are associated with different probability distributions and, therefore, have different expressions for the Fisher information matrix. In the ensuing subsections, relatively high-level expressions for the Fisher information matrix will be given for the image data models, and illustrations will be presented by instantiating the high-level expressions with specific parameter estimation problems.

A. Fundamental Data Model

For the fundamental data model of Section 3.A, the probability distribution of the acquired data is represented by the sample function density for the spatiotemporal random process characterized by the photon detection rate $\Lambda_\theta(\tau)$, $\tau \geq t_0$, $\theta \in \Theta$, and the photon distribution profile $f_{\theta,\tau}(x, y)$, $(x, y) \in \mathbb{R}^2$, $\theta \in \Theta$, $\tau \geq t_0$. If a fundamental image realization consisting of N_0 photons detected over the acquisition interval $[t_0, t]$, where N_0 is Poisson-distributed with mean $\int_{t_0}^t \Lambda_\theta(\tau) d\tau$, is denoted by $\{w_1, \dots, w_{N_0}\}$, where $w_k = (\tau_k, r_k)$, $k = 1, \dots, N_0$, represents the time point of detection τ_k and the location of detection $r_k = (x_k, y_k) \in \mathbb{R}^2$ of the k th detected photon, then the sample function density is given by

$$p_\theta(w_1, \dots, w_{N_0}) = e^{-\int_{t_0}^t \Lambda_\theta(\tau) d\tau} \left(\prod_{k=1}^{N_0} f_{\theta,\tau_k}(r_k) \right) \left(\prod_{k=1}^{N_0} \Lambda_\theta(\tau_k) \right), \quad (19)$$

$\theta \in \Theta$. Using Eq. (19) in place of $p_\theta(w)$ in Eq. (18), the resulting Fisher information matrix can be expressed as

$$\mathbf{I}(\theta) = \int_{t_0}^t \int_{\mathbb{R}^2} \frac{1}{\Lambda_\theta(\tau) f_{\theta,\tau}(x, y)} \left(\frac{\partial[\Lambda_\theta(\tau) f_{\theta,\tau}(x, y)]}{\partial \theta} \right)^T \times \left(\frac{\partial[\Lambda_\theta(\tau) f_{\theta,\tau}(x, y)]}{\partial \theta} \right) dx dy d\tau, \quad \theta \in \Theta. \quad (20)$$

For details on the derivation of Eqs. (19) and (20), see [61].

Continuing the example from Section 3.A of an in-focus and stationary molecule whose photons are detected at a constant rate of $\Lambda_\theta(\tau) = \Lambda_0$, $\tau \geq t_0$, and whose photon distribution profile f_θ is given by Eq. (1) without the time dependence and with the image function q given by the Airy image function of Eq. (2), suppose the molecule's positional coordinates (x_0, y_0) and the photon detection rate Λ_0 are to be estimated from a fundamental image acquired over the time interval $[t_0, t]$. Then, for $\theta = (x_0, y_0, \Lambda_0) \in \Theta$, where $\Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$, evaluation of Eq. (20) followed by application of Eq. (17) for each parameter yields the limits of accuracy [21,22]

$$\delta_{x_0} = \sqrt{[\mathbf{I}^{-1}(\theta)]_{11}} = \frac{\lambda}{2\pi n_a \sqrt{\Lambda_0 \cdot (t - t_0)}}, \quad (21)$$

$$\delta_{y_0} = \sqrt{[\mathbf{I}^{-1}(\theta)]_{22}} = \frac{\lambda}{2\pi n_a \sqrt{\Lambda_0 \cdot (t - t_0)}}, \quad (22)$$

$$\delta_{\Lambda_0} = \sqrt{[\mathbf{I}^{-1}(\theta)]_{33}} = \sqrt{\frac{\Lambda_0}{t - t_0}}. \quad (23)$$

Several points are of note, all of which are exceptions rather than the general rule, owing to the specifics of this particular localization problem. First, all three limits of accuracy have simple analytical expressions, which would not be the case if, for example, the detector had a finite and rectangular photon detection area. Second, the limits of accuracy δ_{x_0} and δ_{y_0} for the estimation of x_0 and y_0 , respectively, are identical because the Airy pattern's circular symmetry, combined with the detector's infinite and unpixelated photon detection area, produces images that contain exactly the same information about x_0 and y_0 . Third, δ_{x_0} and δ_{y_0} are not dependent on the magnification M or the coordinates x_0 and y_0 themselves, a fact that is attributable again to the infinite and unpixelated nature of the detection area. Fourth, δ_{x_0} and δ_{y_0} have a simple inverse square root dependence on the mean detected photon count $\Lambda_0 \cdot (t - t_0)$, a special property that would not generally hold if, for example, a background component had been present.

In Fig. 4(a), δ_{x_0} of Eq. (21) [which is identical to δ_{y_0} of Eq. (22)] is shown, for three different values of the wavelength λ of the photons detected from the molecule, as a function of the expected number of detected photons $\Lambda_0 \cdot (t - t_0)$, which ranges from 100 to 10,000. For each wavelength, the lower bound on the standard deviation with which x_0 or y_0 can be estimated is seen to decrease in value and, therefore improve, as the mean detected photon count is increased. This result is expected from the inverse square root dependence on the photon count but also agrees with the intuition that the accuracy with which any parameter can be estimated should improve as more data are collected. The curves of Fig. 4(a) additionally show that, for a given mean photon count, the limit of accuracy improves with decreasing wavelength of the detected photons. This is expected as well because a shorter wavelength produces a narrower Airy pattern that allows the peak of the pattern, which gives the location of the molecule, to be estimated with better accuracy. Note that, based on Eqs. (21) and (22), the same effect also can be obtained by using a larger numerical aperture n_a .

To give a specific example, the curve corresponding to $\lambda = 520$ nm shows that, from a fundamental image like the one shown in Fig. 1(b), which has a mean detected photon count of $\Lambda_0 \cdot (t - t_0) = 500$, the x_0 and y_0 coordinates of the molecule can each be determined with a standard deviation of no smaller than 2.64 nm. By doubling the expected photon count to 1000, the curve shows that the limit of accuracy is improved to 1.87 nm. If in addition the molecule is changed to one that emits photons of wavelength $\lambda = 440$ nm, then the limit of accuracy is further improved to 1.58 nm.

Note that a limit of the accuracy with which a positional coordinate can be estimated, such as δ_{x_0} and δ_{y_0} in this example, is also referred to as a *localization accuracy measure* [17,62].

In Fig. 4(b), δ_{Λ_0} of Eq. (23) is also seen, for the specific localization problem considered here, to have an inverse square root dependence on the mean detected photon count. For example, the curve shows that, by doubling the acquisition time $t - t_0$ from 0.05 to 0.1 s, and hence doubling the mean detected photon count $\Lambda_0 \cdot (t - t_0)$ from 500 to 1000, the lower bound on the standard deviation with which the photon detection rate of $\Lambda_0 = 10,000$ photons/s can be estimated is improved from 447.2 photons/s to 316.2 photons/s.

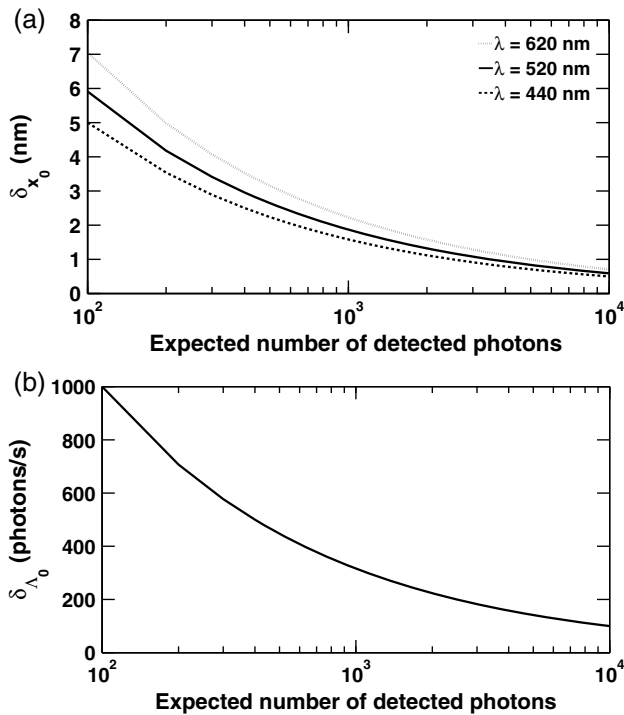


Fig. 4. Limits of accuracy under the fundamental data model. (a) Limit of accuracy δ_{x_0} for the estimation of the x_0 coordinate of an in-focus and stationary molecule from a fundamental image is plotted as a function of the expected number of photons detected in the image. Curves are shown that correspond to the detection of photons of wavelengths $\lambda = 620$ nm (dotted), 520 nm (solid), and 440 nm (dashed), which are collected by an objective lens with numerical aperture $n_a = 1.4$. The expected photon counts range from 100 to 10,000 and are obtained as the product of the constant photon detection rate $\Lambda_0 = 10,000$ photons/s and acquisition times $t - t_0$ ranging from 0.01 to 1 s. The image of the molecule is described by the Airy image function. (b) Limit of accuracy δ_{Λ_0} for the estimation of the photon detection rate $\Lambda_0 = 10,000$ photons/s from a fundamental image is shown as a function of the expected number of photons detected in the image.

It is of interest to note that, if the image of the molecule were described by the 2D Gaussian function of Eq. (3) instead of the Airy pattern, simple analytical expressions similar to those of Eqs. (21)–(23) would have been obtained for the limits of accuracy. Specifically, δ_{Λ_0} would be identical in expression, and δ_{x_0} and δ_{y_0} would each be given by $\sigma_g / \sqrt{\Lambda_0 \cdot (t - t_0)}$ [21,22]. Therefore, in terms of the results obtained with the 2D Gaussian image function, Fig. 4(b) would apply just as well to δ_{Λ_0} , and the patterns of dependence on the photon count and the width of the image function illustrated in Fig. 4(a) would apply to δ_{x_0} and δ_{y_0} . In the case of the 2D Gaussian image function, a shorter wavelength λ or a larger numerical aperture n_a would equate to a smaller standard deviation σ_g , which leads to a narrower 2D Gaussian function.

For another illustration of limits of accuracy under the fundamental data model, consider the estimation of the distance $d > 0$ separating two like molecules that are both in-focus and stationary. Suppose the rates Λ_1 and Λ_2 at which the molecules' photons are detected are constant and equal, such that

$\Lambda_1(\tau) = \Lambda_2(\tau) = \Lambda_0$, $\Lambda_0 > 0$, $\tau \geq t_0$. Note that the subscript θ is not used with the photon detection rates, as they are assumed to be known in this example and will not be estimated. Further suppose that the molecules lie on the x axis and are equidistant from the origin $(0, 0)$ of the object space coordinate system. As the processes corresponding to the detection of photons from the two molecules are independent, their superposition is described by the photon detection rate of Eq. (5), which reduces to $\Lambda(\tau) = 2\Lambda_0$, $\tau \geq t_0$, and by the photon distribution profile f_θ of Eq. (7), which likewise becomes time-independent given the photon detection rates as defined and has the molecules' positional coordinates specified by $(x_{01}, y_{01}) = (-d/2, 0)$ and $(x_{02}, y_{02}) = (d/2, 0)$ based on the given geometry. Further assuming the molecules' image functions q_1 and q_2 in Eq. (7) to each be given by the Airy pattern of Eq. (2), evaluation of the Fisher information matrix of Eq. (20), with $\theta = d \in \mathbb{R}^+$ and Λ and f_θ as described, yields the limit of accuracy [25]

$$\delta_d = \sqrt{\mathbf{I}^{-1}(\theta)} = \frac{1}{\sqrt{4\pi \cdot \Lambda_0 \cdot (t - t_0) \cdot \Gamma_0(d)}} \cdot \frac{\lambda}{n_a}, \quad (24)$$

where

$$\Gamma_0(d) = \int_{\mathbb{R}^2} \frac{1}{\frac{J_1^2(\gamma r_{01})}{r_{01}^2} + \frac{J_1^2(\gamma r_{02})}{r_{02}^2}} \left(\left(x + \frac{d}{2} \right) \frac{J_1(\gamma r_{01}) J_2(\gamma r_{01})}{r_{01}^3} - \left(x - \frac{d}{2} \right) \frac{J_1(\gamma r_{02}) J_2(\gamma r_{02})}{r_{02}^3} \right)^2 dx dy,$$

where $r_{01} = \sqrt{(x + d/2)^2 + y^2}$, $r_{02} = \sqrt{(x - d/2)^2 + y^2}$, $\gamma = 2\pi n_a / \lambda$, and J_2 is the second-order Bessel function of the first kind.

For three different wavelengths λ of the detected photons, δ_d of Eq. (24) is plotted in Fig. 5 as a function of the distance d that separates the two molecules. For each wavelength, the limit of the accuracy with which d can be estimated is seen to improve as the value of d increases, corroborating the expectation that a larger distance should be easier to determine because there is less of an overlap between the images of the two molecules. The curves in Fig. 5 further show that for a given separation distance d , decreasing the wavelength of the detected photons improves the lower bound on the standard deviation with which d can be estimated. This result is again attributable to the narrowing of the Airy pattern when the wavelength is decreased, which reduces the overlap between the patterns corresponding to the two molecules.

The limit of accuracy δ_d importantly shows that, contrary to the Rayleigh criterion [52,63], which specifies a minimum separation distance below which two point sources are deemed indistinguishable, arbitrarily small distances between two point sources can in fact be determined. With the limit of accuracy, the question becomes one of not whether but how well two point sources can be resolved, in the sense of how accurately the distance separating them can be determined. For example, whereas the Rayleigh criterion, given by $0.61\lambda/n_a$, specifies a minimum resolvable distance of 226.57 nm for $\lambda = 520$ nm and $n_a = 1.4$, the curve corresponding to $\lambda = 520$ nm in

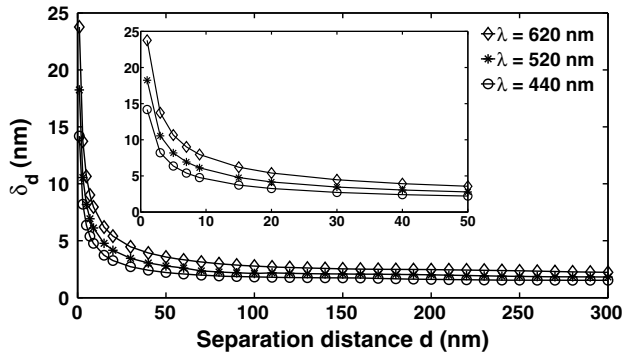


Fig. 5. Limit of the accuracy for separation distance estimation under the fundamental data model. The limit of accuracy δ_d for the estimation, from a fundamental image, of the distance d separating two in-focus and stationary molecules is plotted as a function of the distance d . Curves are shown that correspond to the detection of photons of wavelengths $\lambda = 620$ nm (\diamond), 520 nm ($*$), and 440 nm (\circ), which are collected by an objective lens with numerical aperture $n_d = 1.4$. Photons are detected from each molecule at a rate of $\Lambda_0 = 10,000$ photons/s over an acquisition time of 0.3 s, such that $\Lambda_0 \cdot (t - t_0) = 3000$ photons are, on average, detected from each molecule in a given image. The image of each molecule is described by the Airy image function. Inset shows the portion of the curves from $d = 1$ nm to $d = 50$ nm.

Fig. 5 shows that a much smaller separation distance of 20 nm can be estimated with a standard deviation of no better than 4.16 nm when $\Lambda_0 \cdot (t - t_0) = 3000$ photons are on average detected from each point source. Moreover, as can be seen from Eq. (24), this limit of accuracy can be improved, as demonstrated in Fig. 4 for the localization problem, by increasing the expected number of detected photons $\Lambda_0 \cdot (t - t_0)$.

A limit of accuracy δ_d for the estimation of a separation distance also is referred to as a *resolution measure* [25,26]. More information and results for the scenario of two in-focus molecules considered here, including δ_d derived under the pixelated data models, can be found in [25]. The resolution measure also has been extended to the more general scenario of two molecules in 3D space [26], where it has been compared with the classical axial analogue [63] of the Rayleigh criterion.

Note that the simple inverse square root dependence of δ_d of Eq. (24) on the mean number of photons $\Lambda_0 \cdot (t - t_0)$ detected from each molecule is something that can be seen from a much higher level. In the distance estimation problem as presented, the photon detection process that underlies the fundamental image is described by a photon detection rate Λ that does not depend on the parameter vector θ and a photon distribution profile f_θ that does not change with time. Given such a photon detection process, the Fisher information matrix of Eq. (20) decouples into the product of integrals

$$\mathbf{I}(\theta) = \int_{t_0}^t \Lambda(\tau) d\tau \times \int_{\mathbb{R}^2} \frac{1}{f_\theta(x, y)} \left(\frac{\partial f_\theta(x, y)}{\partial \theta} \right)^T \left(\frac{\partial f_\theta(x, y)}{\partial \theta} \right) dx dy, \quad (25)$$

$\theta \in \Theta$, where the integral over time is just the expected number of detected photons. Therefore, upon inverting the matrix and

taking the square root of each main diagonal element, the limit of the accuracy for estimating any given parameter in θ will have $1/\sqrt{\int_{t_0}^t \Lambda(\tau) d\tau}$ as part of its expression, as is seen in Eq. (24) where $\Lambda(\tau)$, $\tau \geq t_0$, is defined as the constant function $2\Lambda_0$. It is important to point out that while Eq. (25) corresponds to a common estimation problem formulation, Fisher information matrices corresponding to other problem descriptions also can lead to limits of accuracy with an inverse square root dependence on the mean detected photon count. The Fisher information matrix from which the limits of accuracy of Eqs. (21), (22), and (23) are obtained provides such an example as its associated problem description involves a photon detection rate Λ_θ that does depend on θ .

As the fundamental data model embodies the most ideal of imaging conditions, each limit of accuracy in the above examples represents, within the context of the particular estimation problem, the absolute best accuracy with which the given parameter can be determined. Put another way, a limit of accuracy obtained under the fundamental data model importantly marks the point beyond which no further improvement in accuracy is possible for estimating the parameter. It can therefore be used as the ultimate target of comparison to determine how much room there is for improving the accuracy of estimation. Given the practical limitations of an experimental setup, however, a limit of accuracy obtained under the fundamental data model often represents an unrealistic target. Instead, a tighter limit of accuracy that is obtained under a data model that assumes, for example, a pixelated detector with a finite photon detection area will provide a more realistic accuracy benchmark that can be approached in practice. This will be demonstrated in the next subsections, but, as alluded to in Section 3.A, variations of the fundamental data model can be obtained that will already yield tighter limits of accuracy that provide more realistic targets. The Fisher information matrix corresponding to variations entailing the addition of a background component or a finite photon detection area for the detector is obtained by a straightforward generalization of Eq. (20) that adds a background component using the superposition property of Section 2.B and restricts the region of integration of photon distribution profiles to the confines of the finite detection area. For an image acquired over the time interval $[t_0, t]$, this matrix is given by

$$\mathbf{I}(\theta) = \int_{t_0}^t \int_C \frac{1}{\Lambda_\theta(\tau) f_{\theta, \tau}(x, y) + \Lambda_\theta^b(\tau) f_{\theta, \tau}^b(x, y)} \times \left(\frac{\partial [\Lambda_\theta(\tau) f_{\theta, \tau}(x, y) + \Lambda_\theta^b(\tau) f_{\theta, \tau}^b(x, y)]}{\partial \theta} \right)^T \times \left(\frac{\partial [\Lambda_\theta(\tau) f_{\theta, \tau}(x, y) + \Lambda_\theta^b(\tau) f_{\theta, \tau}^b(x, y)]}{\partial \theta} \right) dx dy d\tau, \quad (26)$$

$\theta \in \Theta$, where $\Lambda_\theta^b(\tau)$, $\tau \geq t_0$ and $f_{\theta, \tau}^b(x, y)$, $(x, y) \in \mathbb{R}^2$, $\tau \geq t_0$, are the photon detection rate and distribution profile, respectively, of the background component, and C is the finite region within \mathbb{R}^2 corresponding to the photon detection area of the detector. This expression will be used in the next subsection to calculate limits of accuracy that represent more realistic benchmarks than the limit obtained with Eq. (20).

B. Poisson Data Model

In a pixelated image, the data in each pixel is obtained independently of the data in all other pixels. Given this typical and well-justified assumption, the probability distribution for the data comprising a K -pixel image is the product of the probability distributions for the data in the K pixels. Denoting the probability distribution for the random variable $H_{\theta,k}$ that models the electron count z_k in the k th pixel by $p_{\theta,k}$, $k = 1, \dots, K$, the probability distribution p_{θ} for a K -pixel image can be written as

$$p_{\theta}(z_1, \dots, z_K) = \prod_{k=1}^K p_{\theta,k}(z_k), \quad \theta \in \Theta. \quad (27)$$

Using Eq. (27) in place of $p_{\theta}(w)$ in the general expression of Eq. (18), the Fisher information matrix corresponding to a K -pixel image can be shown to be

$$\mathbf{I}(\theta) = \sum_{k=1}^K \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ln p_{\theta,k}(z) \right)^T \left(\frac{\partial}{\partial \theta} \ln p_{\theta,k}(z) \right) \right], \quad (28)$$

$\theta \in \Theta$, which is easily verified to be the sum of the Fisher information matrices corresponding to the data in the K pixels.

From the examples of Section 4.A, it can be seen that, in a typical estimation problem, parameters of interest represented by the vector θ , such as the positional coordinates of an object, the rate at which photons are detected from an object, and the distance separating two objects, are naturally all part of the specification of the detection rate Λ_{θ} and distribution profile $f_{\theta,\tau}$ of the photon detection process. It follows that, in the pixelated case, information about θ in the k th pixel of an image is contained in the object-based component $S_{\theta,k}$ of the data in the pixel [see Eq. (8)], whose mean $\mu_{\theta,k}$ is specified in terms of Λ_{θ} and $f_{\theta,\tau}$. In addition, information about θ is present in the background component $B_{\theta,k}$ of the data in the pixel when one chooses to estimate quantities that parameterize its mean $\beta_{\theta,k}$. Taken together, information about θ in the k th pixel is contained in the Poisson component $S_{\theta,k} + B_{\theta,k}$ of the data in the pixel, which has mean $\nu_{\theta,k}$ given by Eq. (11). On the other hand, information about θ is typically not contained in the detector-based noise components of the data in the pixel. Even though parameters such as the readout noise mean and variance and the EMCCD detector's electron multiplication gain can, in principle, be included in θ , they are usually not included because they can be determined separately from the object-based and background-based parameters. Given this typical form of an estimation problem where, in the k th pixel, θ parameterizes only the mean $\nu_{\theta,k}$ of the Poisson component of the data, the Fisher information matrix of Eq. (28) can be written as

$$\mathbf{I}(\theta) = \sum_{k=1}^K \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right)^T \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right) \cdot \mathbb{E} \left[\left(\frac{\partial \ln p_{\theta,k}(z)}{\partial \nu_{\theta,k}} \right)^2 \right], \quad (29)$$

$\theta \in \Theta$. For $k = 1, \dots, K$, this representation importantly expresses the Fisher information matrix corresponding to the data in the k th pixel as a product of two fundamentally different parts. The first part is a matrix of partial derivatives of the mean photoelectron count $\nu_{\theta,k}$ with respect to the parameters in θ . It is a function of the specific estimation problem, as the derivatives depend on the vector θ of the parameters to be estimated,

and through $\nu_{\theta,k}$, on the specification of the photon detection rates and distribution profiles corresponding to the object(s) of interest and a potential background component. As $\nu_{\theta,k}$ is purely a function of photon detection rates and distribution profiles, the matrix is not dependent in any way on the particular detector or type of detector that is used to capture the image data. Therefore, for a given estimation problem, the matrix is the same regardless of the detector used. This can be seen in the fact that the Poisson image of Fig. 2(b) and the CCD and EMCCD images of Fig. 3 are all realizations based on the same mean image of Fig. 2(a), which is just a pixel array of $\nu_{\theta,k}$ values. (Note that, as mentioned in Section 3.B, $\nu_{\theta,k}$ can in fact include, through the mean $\beta_{\theta,k}$ of the background component, spurious electrons that originate from the detector. However, because these spurious electrons are common to, and are typically very small in number in all standard detector types, $\nu_{\theta,k}$ can justifiably be considered to be independent of the particular detector and type of detector used.)

In direct contrast, the second part of the product is a scalar expectation term that is not dependent on the specific estimation problem but is dependent on the particular detector and type of detector used. It depends on the detector through the probability distribution $p_{\theta,k}$ of the electron count in the k th pixel, which is a function of parameters specific to a detector such as the mean and variance of the readout noise in the case of a CCD or CMOS detector, and both the mean and variance of the readout noise and the electron multiplication gain in the case of an EMCCD detector. The scalar expectation term does not depend on the specific estimation problem, in the sense that, while it is a function of $\nu_{\theta,k}$, it is dependent only on its value and not how the value is obtained through the underlying photon detection rates and distribution profiles. In other words, a given value of $\nu_{\theta,k}$, regardless of how $\nu_{\theta,k}$ or θ is defined, will always produce the same value for the scalar expectation term, provided all other parameters in the term's expression remain the same.

As will be seen in Section 4.E, the fact that it is not dependent on the specific estimation problem makes the scalar expectation term a useful tool for determining, based solely on the mean photoelectron count in a pixel, the effect that a given detector has on the information content of the pixel. Indeed, the breaking of the overall data description into an estimation problem-dependent portion and a detector noise-dependent portion also has been exploited in a nested expectation maximization algorithm for the localization of closely spaced fluorescence emitters [64].

As the scalar expectation term of Eq. (29) is the only part of the Fisher information expression that is dependent on the type of detector used, the Fisher information matrices corresponding to images captured with different types of pixelated detectors will differ by this term through the probability distributions that describe the electron counts comprising the images.

In a K -pixel image described by the Poisson data model, which assumes a hypothetical detector that introduces no readout noise, the photoelectron count z_k in the k th pixel, $k = 1, \dots, K$, is according to Eq. (8) a realization of the Poisson random variable $H_{\theta,k}$ with mean $\nu_{\theta,k}$. The probability distribution $p_{\theta,k}$ for $H_{\theta,k}$, $k = 1, \dots, K$, is thus the Poisson probability mass function

$$p_{\theta,k}(z) = \frac{\nu_{\theta,k}^z e^{-\nu_{\theta,k}}}{z!}, \quad z = 0, 1, \dots, \quad \theta \in \Theta. \quad (30)$$

Evaluating the scalar expectation term of Eq. (29) with Eq. (30) yields the Fisher information matrix

$$\mathbf{I}(\theta) = \sum_{k=1}^K \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right)^T \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right) \cdot \frac{1}{\nu_{\theta,k}}, \quad \theta \in \Theta. \quad (31)$$

By incorporating the realistic assumption of a detector with a finite and pixelated photon detection area and by allowing for the presence of a background component, a limit of accuracy computed under the Poisson data model can be significantly worse than the limit of accuracy computed under the fundamental data model. This is because the conditions assumed by the Poisson data model represent a deterioration of the quality of the acquired image data. The finite detection area means a loss of data in that photons falling outside the detection area are not detected, and the pixelation of the detection area means a loss of precision with which the location of a detected photon is recorded because the extent to which the location is known is reduced to the pixel within which the photon impacts the detection area. Meanwhile, the presence of a background component naturally results in the corruption of the photon signal detected from the object(s) of interest.

A comparison of limits of accuracy computed under the Poisson and fundamental data models is given in Fig. 6, which again uses the example of the limit of accuracy δ_{x_0} for the estimation of an in-focus and stationary molecule's x_0 coordinate. In the figure, δ_{x_0} corresponding to the Poisson data model (*), computed using Eq. (31), is seen to improve as the precision of the recording of photon location by the detector is increased by a decrease in the effective pixel size, which is, in turn, realized by increasing the magnification used for the imaging. The improvement is seen to level off, however, at more than 1.5 nm away from the fundamental data model's limit of accuracy of $\delta_{x_0} = 3.41$ nm (solid line). This substantial gap is largely attributable to the presence of background noise that is uniformly distributed over the image. By assuming the absence of this background component [by ignoring $\beta_{\theta,k}$ in Eq. (11) and using $\nu_{\theta,k} = \mu_{\theta,k}$, $k = 1, \dots, K$, in Eq. (31)], δ_{x_0} for the Poisson data model (+) can be seen to level off at a vastly improved value that is within a quarter of a nanometer of the target of 3.41 nm. Note that, for a given effective pixel size, δ_{x_0} for the Poisson data model assuming the absence of background noise represents, as expected, a better accuracy than δ_{x_0} for the Poisson data model assuming the presence of background noise.

Even with the absence of background noise and the use of a small effective pixel size, it is clear from Fig. 6 that δ_{x_0} for the Poisson data model still does not quite approach δ_{x_0} for the fundamental data model. The remaining small gap is primarily due to the detector's finite photon detection area under the Poisson data model; this is illustrated by the fact that δ_{x_0} computed for the same finite but unpixelated photon detection area and with the assumption of the absence of background noise (dotted line) has a value of 3.54 nm that is greater than δ_{x_0} for the fundamental data model. As can be seen, the value of 3.54 nm represents a more realistic target for δ_{x_0} corresponding to the Poisson data model assuming the absence of background noise. This value is calculated using the generalized Fisher

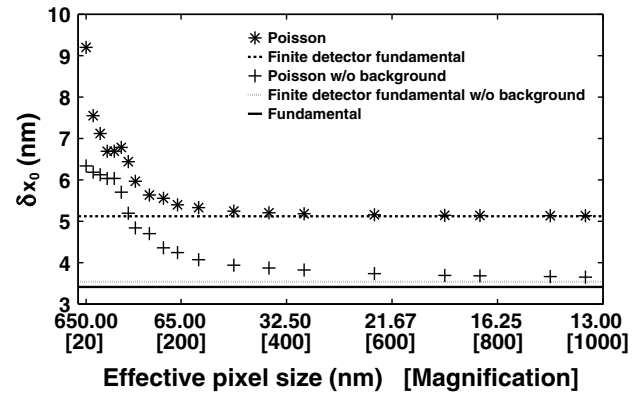


Fig. 6. Comparing limits of accuracy corresponding to different data models and variations thereof. (a) Limit of accuracy δ_{x_0} for the estimation of the x_0 coordinate of an in-focus and stationary molecule from an image is plotted as a function of the effective pixel size for the Poisson data model (*) and the Poisson data model assuming the absence of background noise (+). Also shown are the limits of accuracy corresponding to the finite detection area variation of the fundamental data model (dashed line), the finite detection area variation of the fundamental data model assuming the absence of background noise (dotted line), and the fundamental data model (solid line), which are plotted as horizontal lines, as they do not depend on the effective pixel size. For the two cases involving the Poisson data model, the physical pixel size is $13 \mu\text{m} \times 13 \mu\text{m}$, and the different effective pixel sizes are obtained by varying the lateral magnification from $M = 20$ to $M = 966.67$. The image consists of a 15×15 -pixel region at $M = 100$ and a proportionately scaled pixel region at each of the other magnifications. At each magnification, the molecule is positioned such that the center of its image, given by the Airy image function, is located at 0.05 pixels in the x direction and 0.25 pixels in the y direction with respect to the upper-left corner of the center pixel of the image. For the two cases involving the finite detection area variation of the fundamental data model, δ_{x_0} is computed using the unpixelated equivalent of the $M = 100$ scenario, though any other scenario can be used to obtain the same result. For all cases, photons are detected from the molecule at a rate of $\Lambda_0 = 6000$ photons/s over an acquisition time of $t - t_0 = 0.05$ s, such that $\Lambda_0 \cdot (t - t_0) = 300$ photons are, on average, detected from the molecule over the detector plane (and 289.5 photons are, on average, detected from the molecule over a given finite-sized image). For the two cases involving a background component, a background detection rate of $\Lambda_0^b = 22,500$ photons/s is assumed, along with a uniform background spatial distribution of $f^b(x, y) = 1/A \mu\text{m}^{-2}$, $(x, y) \in C$, where C is the region in the detector plane corresponding to the image and A is the area of C . For example, for the $M = 100$ scenario, $f^b(x, y) = 1/38,025 \mu\text{m}^{-2}$, such that there is a background level of $\beta_0 = 5$ photoelectrons per pixel over the 15×15 -pixel region. For all cases, it is assumed that photons of wavelength $\lambda = 520$ nm are collected by an objective lens with numerical aperture $n_a = 1.4$.

information matrix of Eq. (26), without the background photon detection rate Λ_θ^b and distribution profile $f_{\theta,\tau}^b$, and with the region of integration C defined to be the region in the detector plane corresponding to the image. Equation (26) is also used to compute δ_{x_0} under the same conditions but with Λ_θ^b and $f_{\theta,\tau}^b$ defined to specify the uniformly distributed background noise (dashed line) that is accounted for by the Poisson data model that assumes the presence of background

noise. The value of δ_{x_0} obtained in this case is 5.12 nm and analogously represents a tighter bound for δ_{x_0} corresponding to the Poisson data model assuming the presence of background noise.

Figure 6 thus demonstrates the various levels of accuracy benchmarking that are possible for a given estimation problem by considering the Fisher information for different data models and variations thereof. Importantly, it illustrates that, by comparing the limits of accuracy corresponding to different data models and their variations, one can obtain a clear picture of the factors responsible for, and the degrees to which they affect, the distance between a given limit of accuracy and the ultimate target as provided by the fundamental data model.

Note that the limits of accuracy in Fig. 6 are computed for $\theta = (x_0, y_0) \in \mathbb{R}^2$, and that, while not shown, the results for δ_{y_0} follow the same patterns as the results for δ_{x_0} . The above discussion therefore applies equally well to δ_{y_0} . Note also that, in the cases involving a finite photon detection area, δ_{y_0} is not identical in value to δ_{x_0} because of the asymmetric positioning of the Airy pattern within the image. While they are nearly identical in value when the image is unpixelated or has a fine effective pixelation, they can have very different values when the image has a coarse effective pixelation.

C. CCD/CMOS Data Model

According to Eq. (12), the electron count z_k , $k = 1, \dots, K$, in the k th pixel of a K -pixel CCD or CMOS image is a realization of the random variable $H_{\theta,k}$ obtained as the sum of a Poisson random variable with mean $\nu_{\theta,k}$ and a Gaussian random variable with mean η_k and variance σ_k^2 . By the independence of the Poisson and Gaussian random variables, the probability distribution $p_{\theta,k}$ for $H_{\theta,k}$, $k = 1, \dots, K$, is the convolution of a Poisson distribution and a Gaussian distribution with the given parameters. The result of the convolution is the probability density function [21,65]

$$p_{\theta,k}(z) = \frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=0}^{\infty} \frac{\nu_{\theta,k}^l}{l!} e^{-\frac{(z-l-\eta_k)^2}{2\sigma_k^2}}, \quad z \in \mathbb{R}, \quad \theta \in \Theta. \quad (32)$$

Using Eq. (32) in Eq. (29) gives the Fisher information matrix

$$\mathbf{I}(\theta) = \sum_{k=1}^K \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right)^T \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right) \times \left(\int_{-\infty}^{\infty} \frac{\left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} \frac{\nu_{\theta,k}^{l-1}}{(l-1)!} e^{-\frac{(z-l-\eta_k)^2}{2\sigma_k^2}} \right)^2}{p_{\theta,k}(z)} dz - 1 \right), \quad (33)$$

$\theta \in \Theta$, where $p_{\theta,k}$ is the density function of Eq. (32).

For the same localization problem considered in Fig. 6, limits of accuracy computed for three different levels of readout noise using the Fisher information matrix of Eq. (33) are shown in Fig. 7(a) as a function of the effective pixel size. These limits of accuracy correspond to the CCD data model, as the same mean and variance are assumed for the readout noise in every pixel of the detector. However, while the relatively high readout noise standard deviation of 6 electrons is typical of older but still commonly used CCD detectors, the readout noise standard

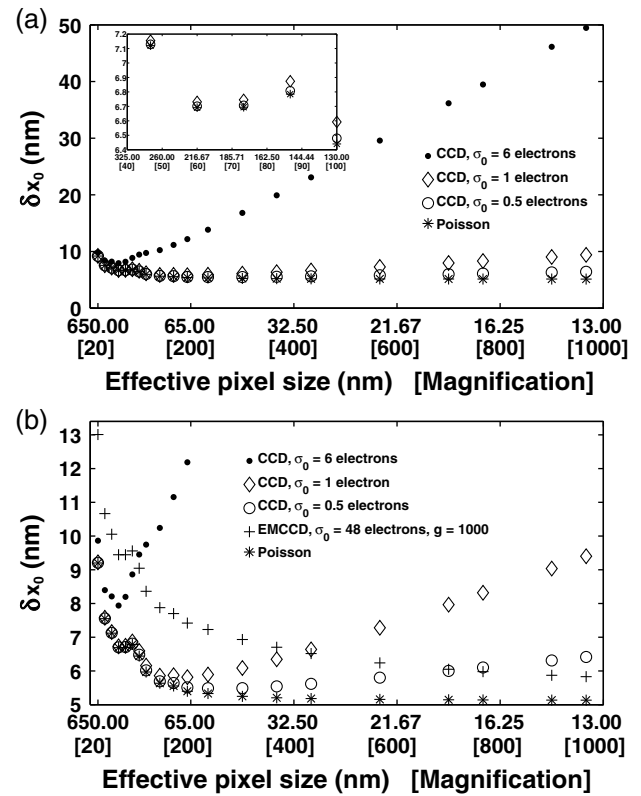


Fig. 7. Comparing limits of accuracy corresponding to the CCD and EMCCD data models. (a) For the CCD data model, the limit of accuracy δ_{x_0} for the estimation of the x_0 coordinate of an in-focus and stationary molecule from an image is plotted as a function of the effective pixel size. Limits of accuracy are shown for three different levels of readout noise in each pixel, characterized by standard deviations of $\sigma_0 = 6$ electrons (\bullet), 1 electron (\diamond), and 0.5 electrons (\circ). In each case, the mean of the readout noise in each pixel is $\eta_0 = 0$ electrons. The limit of accuracy for the Poisson data model (with background noise) from Fig. 6 ($*$) is shown as a reference for comparison because the CCD limits of accuracy are computed by adding readout noise to the same assumed conditions. All details not mentioned are therefore exactly as given in Fig. 6 for the Poisson data model. The inset provides a clearer view of some data points for five relatively large effective pixel sizes. (b) Zoomed-in version of (a) with the addition of δ_{x_0} corresponding to the EMCCD data model with readout noise of mean $\eta_0 = 0$ electrons and standard deviation $\sigma_0 = 48$ electrons in each pixel and an electron multiplication gain of $g = 1000$.

deviations of 0.5 electrons and 1 electron can be regarded as being representative of idealized versions of very-low-noise CMOS detectors that have readout circuitries with identical noise characteristics for all their pixels. The CCD limits of accuracy are computed for the same conditions used for δ_{x_0} for the Poisson data model (with background noise) from Fig. 6 but with added readout noise. Consequently, Fig. 7(a) shows that, for a given effective pixel size, they represent accuracies that are strictly worse than the Poisson limit of accuracy reproduced from Fig. 6. For the larger effective pixel sizes, this is seen more clearly in the figure's inset. More generally, this provides an example of using limits of accuracy computed under the

Poisson data model as tight benchmarks for the evaluation of limits of accuracy computed under a practical data model.

Figure 7(a) further demonstrates the deteriorative effect of readout noise by showing δ_{x_0} corresponding to the CCD data model to worsen, at a given effective pixel size, with increasing values of the readout noise standard deviation. At the larger effective pixel sizes, the pattern can be seen for the small readout noise standard deviations of 1 and 0.5 electrons in the figure's inset. Also of note is the fact that neither very large nor very small effective pixel sizes are optimal in terms of parameter estimation accuracy when a CCD detector is used. [For a clearer view of the shapes of the curves, see the zoomed-in versions in Fig. 7(b).] When the effective pixel size is large, the number of photoelectrons in each pixel will be relatively large, so that the effect of the readout noise will be minimal. This explains the fact that, at the largest effective pixel sizes, the δ_{x_0} values for the CCD data models with already low readout noise standard deviations of 1 and 0.5 electrons are nearly equal to the corresponding values of δ_{x_0} for the Poisson data model. However, such a scenario will not necessarily yield the best limit of accuracy because large pixels also mean that the locations of the detected photons are recorded with poor precision and that the resulting pixelated image represents a poorly sampled version of the continuous image described by the underlying photon distribution profiles. On the other hand, a small effective pixel size also will not necessarily yield the best limit of accuracy despite the pixelated image being a finely sampled version of the continuous image. In this scenario, the relatively few photoelectrons in each pixel will mean that the effect of readout noise becomes significant.

For a specific example, consider the CCD data model with the readout noise standard deviation of 6 electrons in Fig. 7. In this case, δ_{x_0} attains its lowest value of around 7.94 nm at an effective pixel size of approximately 216.67 nm (60× magnification). At the large effective pixel size of 650 nm (20× magnification), it has a poorer value of 9.86 nm, and, at the small effective pixel size of 43.33 nm (300× magnification), it predicts an even worse estimation accuracy of no better than 16.81 nm.

D. EMCCD Data Model

In a K -pixel EMCCD image, the electron count z_k in the k th pixel, $k = 1, \dots, K$, is according to Eq. (13) a realization of the random variable $H_{\theta,k}$ obtained as the sum of a random variable with probability mass function given by Eq. (14) and a Gaussian random variable with mean η_k and variance σ_k^2 . By the independence of the two summand random variables, the probability distribution $p_{\theta,k}$ for $H_{\theta,k}$, $k = 1, \dots, K$, is the convolution of Eq. (14) and a Gaussian distribution with mean η_k and variance σ_k^2 . The resulting probability density function is given by [27]

$$p_{\theta,k}(z) = \frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi\sigma_k}} \left[e^{-\frac{(z-\eta_k)^2}{2\sigma_k^2}} + \sum_{l=1}^{\infty} e^{-\frac{(z-l-\eta_k)^2}{2\sigma_k^2}} \right] \times \sum_{j=0}^{l-1} \frac{\binom{l-1}{j} \left(1 - \frac{1}{g}\right)^{l-1-j} \left(\frac{\nu_{\theta,k}}{g}\right)^{j+1}}{(j+1)!}, \quad z \in \mathbb{R}, \quad (34)$$

$\theta \in \Theta$. Evaluation of Eq. (29) using Eq. (34) yields the Fisher information matrix

$$\mathbf{I}(\theta) = \sum_{k=1}^K \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right)^T \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right) \times \left(\int_{-\infty}^{\infty} \frac{\left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi\sigma_k}} \sum_{l=1}^{\infty} e^{-\frac{(z-l-\eta_k)^2}{2\sigma_k^2}} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j} \left(1 - \frac{1}{g}\right)^{l-1-j} \left(\frac{\nu_{\theta,k}}{g}\right)^{j+1}}{j!g^{j+1}} \right)^2}{p_{\theta,k}(z)} dz - 1 \right), \quad (35)$$

$\theta \in \Theta$, where $p_{\theta,k}$ is the density function of Eq. (34). This expression applies specifically to the modeling of electron multiplication by a geometrically multiplied branching process because its derivation makes use of Eq. (14) for the distribution of the amplified electron count in the k th pixel. If Eq. (14) is replaced by the exponential distribution-based density function of Eq. (15) in the convolution with the Gaussian distribution with mean η_k and variance σ_k^2 , then the probability distribution $p_{\theta,k}$ for $H_{\theta,k}$, $k = 1, \dots, K$, is given by [27]

$$p_{\theta,k}(z) = \frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi\sigma_k}} \left[e^{-\frac{(z-\eta_k)^2}{2\sigma_k^2}} + \int_0^{\infty} e^{-\frac{(z-u-\eta_k)^2}{2\sigma_k^2} - \frac{u}{g}} \times \frac{\sqrt{\frac{\nu_{\theta,k}u}{g}} I_1 \left(2\sqrt{\frac{\nu_{\theta,k}u}{g}} \right)}{u} du \right], \quad z \in \mathbb{R}, \quad (36)$$

$\theta \in \Theta$. Using Eq. (36) in Eq. (29) yields the Fisher information matrix

$$\mathbf{I}(\theta) = \sum_{k=1}^K \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right)^T \left(\frac{\partial \nu_{\theta,k}}{\partial \theta} \right) \times \left(\int_{-\infty}^{\infty} \frac{\left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi\sigma_k g}} \int_0^{\infty} e^{-\frac{(z-u-\eta_k)^2}{2\sigma_k^2} - \frac{u}{g}} I_0 \left(2\sqrt{\frac{\nu_{\theta,k}u}{g}} \right) du \right)^2}{p_{\theta,k}(z)} dz - 1 \right), \quad (37)$$

$\theta \in \Theta$, where I_0 is the zeroth-order modified Bessel function of the first kind, and $p_{\theta,k}$ is the density function of Eq. (36). As the probability distribution of Eq. (15) is a good approximation of the probability distribution of Eq. (14) when the electron multiplication gain g is large, Eq. (37) can be expected to yield limits of accuracy that are similar in value to those obtained with Eq. (35) for large values of g .

Returning to the example of the localization of an in-focus and stationary molecule, Fig. 7(b) adds to Fig. 7(a) by showing δ_{x_0} for the EMCCD data model, computed using the Fisher information matrix of Eq. (35) with an electron multiplication gain of $g = 1000$ and a readout noise standard deviation of $\sigma_0 = 48$ electrons per pixel. As the EMCCD limit of accuracy also is calculated for the same conditions used for δ_{x_0} for the Poisson data model, but accounts for the effects of electron multiplication and added readout noise, it is, as in the case of the CCD limits of accuracy, strictly worse than the Poisson limit of accuracy for a given effective pixel size.

Compared with the three curves corresponding to the CCD data model, one can see that the EMCCD limit of accuracy is worse at the largest effective pixel sizes. This is expected because the relatively high numbers of photoelectrons in the large pixels render the stochastic electron multiplication unnecessary, causing it to have the net effect of introducing additional noise to the data. As the effective pixel size is decreased, the EMCCD limit of accuracy can be seen to improve, much like δ_{x_0} for the CCD detectors with the low readout noise standard deviations of 1 and 0.5 electrons. Unlike the CCD limits of accuracy, however, it continues to improve at the smallest effective pixel sizes shown and surpasses even the CCD limit of accuracy computed for the 0.5-electron readout noise standard deviation at a pixel size of around 17 nm. At this small pixel size, only about 2 photons are, on average, detected in the brightest pixel of the image (largest $\nu_{\theta,k}$ value is 2.03 photoelectrons), and less than 1 photon is on average detected in over 99% of the pixels. This shows that the EMCCD detector is particularly suited for imaging under conditions in which a very small amount of light is expected to impinge on each of its pixels. This unique property of the EMCCD detector forms the basis of the ultrahigh accuracy imaging modality proposed in [66] and will be treated in more detail in Section 4.E.

It is important to emphasize that the results of Fig. 7(b) are specific to the conditions assumed. In particular, given that an average of 289.5 photons are detected from the molecule in a given image and an average of $\beta_0 = 5$ background photons per pixel are detected at the 100 \times magnification, Fig. 7(b) shows that the smallest δ_{x_0} values are attained by the CCD detectors with the 1- and 0.5-electron readout noise standard deviations at an effective pixel size in the neighborhood of 60 nm. Given sufficiently smaller mean photon counts from the molecule and the background component, however, the EMCCD detector can be expected to attain the overall lowest value of δ_{x_0} within the range of effective pixel sizes considered and to start outperforming the very-low-noise CCD detectors at a larger effective pixel size [66].

E. Noise Coefficient

As seen in Sections 4.C and 4.D, limits of accuracy corresponding to the CCD/CMOS and EMCCD data models depend in different ways on the amount of light that is detected in their pixels. Whereas the CCD and CMOS detectors tend to provide superior parameter estimation performance when relatively many photons are detected in each of their pixels, the EMCCD detector tends to yield the best limits of accuracy when the expected photoelectron count in each of its pixels is very small. This difference in dependence on the pixel photoelectron count can be explored using a quantity based on the scalar expectation term in the Fisher information matrix expression of Eq. (29).

For $k = 1, \dots, K$, the scalar expectation term corresponding to the k th pixel of a K -pixel image is easily verified to be the Fisher information with respect to the mean $\nu_{\theta,k}$ of the Poisson-distributed photoelectron count due to the photons detected in the k th pixel. Therefore, the larger the value of this scalar term, the greater the amount of information that the data in the k th pixel contains about the mean $\nu_{\theta,k}$ of the photoelectron count

in the pixel and the lower the bound on the standard deviation with which $\nu_{\theta,k}$ can be estimated. As noted in Section 4.B, this term also is the part of the Fisher information expression for the k th pixel that, through its dependence on the probability distribution $p_{\theta,k}$ of the data in the k th pixel, accounts for the deteriorative effects of the noise introduced by the particular detector that is used to capture the image. The value of this term therefore importantly provides a detector-dependent measure of the amount of information that the data in the k th pixel contains about $\nu_{\theta,k}$. More generally, it can be seen from Eq. (29) that this scalar term is a detector-dependent scaling factor for the estimation problem-dependent matrix portion of the Fisher information expression for the k th pixel. Therefore, it also is the case that the larger its value, the greater the amount of information that the data in the k th pixel carries about the vector θ of parameters to be estimated and the lower the bound on the standard deviation with which each of those parameters can be determined based on the data in the k th pixel. It follows that, for a given level of light $\nu_{\theta,k}$ that impinges on the k th pixel, the scalar expectation term can be calculated for different detector noise settings and detector types to determine, by comparison, the detector and noise settings that yield data in the pixel with the highest information content with respect to θ . Comparisons using the expectation term are based solely on the expected amount of light $\nu_{\theta,k}$ that is detected in the k th pixel and do not depend on the particular estimation problem because, as noted in Section 4.B, the expectation term depends only on the value of $\nu_{\theta,k}$ and not on the definition of $\nu_{\theta,k}$ or θ .

To facilitate the use of the scalar expectation term in assessing the information content of the data in the k th pixel, a meaningful point of reference for comparison is taken to be the expectation term $1/\nu_{\theta,k}$ [see Eq. (31)] corresponding to the Poisson data model. For a given $\nu_{\theta,k}$, this expectation term has the largest value compared with the expectation terms for the other pixelated data models because the Poisson data model represents the most ideal of the pixelated data models by assuming the absence of detector readout noise and, therefore, an uncorrupted photoelectron count in the k th pixel. This point of reference is incorporated through the definition of a quantity called the *noise coefficient* [27]. Denoted by α_k , the noise coefficient for the k th pixel of a given detector is quite simply the ratio of the expectation term for the k th pixel of the detector to the expectation term $1/\nu_{\theta,k}$ for the corresponding pixel of the hypothetical detector assumed by the Poisson data model. It is given by

$$\alpha_k = \frac{\text{E} \left[\left(\frac{\partial \ln p_{\theta,k}(z)}{\partial \nu_{\theta,k}} \right)^2 \right]}{\frac{1}{\nu_{\theta,k}}} = \nu_{\theta,k} \cdot \text{E} \left[\left(\frac{\partial \ln p_{\theta,k}(z)}{\partial \nu_{\theta,k}} \right)^2 \right], \quad (38)$$

and note that the Fisher information matrix of Eq. (29) can easily be expressed in terms of this coefficient by substituting the expectation term with $\alpha_k/\nu_{\theta,k}$. The noise coefficient as defined conveniently takes on a minimum value of 0 and a maximum value of 1. It takes on a minimum value of 0 because $\nu_{\theta,k} > 0$ and because the expectation term, being the expected value of a squared expression, must be greater than or equal to 0. The minimum value of 0 is, therefore, attained when the

expectation term is 0, which occurs in the extreme case in which the data in the k th pixel consists of only the detector's readout noise and, hence, carries no information about θ . The noise coefficient takes on a maximum value of 1 because the expectation term, as noted above, can at most be equal to $1/\nu_{\theta,k}$. In other words, a maximum value of 1 is attained for the coefficient when the given detector is the Poisson data model-based reference detector itself, a scenario that represents the opposite extreme wherein the data in the k th pixel consists of only the photoelectron count, uncorrupted by readout noise.

For a practical detector described by the CCD/CMOS data model or the EMCCD data model, the noise coefficient thus takes on a value between 0 and 1. For a given light level $\nu_{\theta,k}$, a value close to 0, which results from a small expectation term, indicates that the data in the k th pixel is to a large extent corrupted by detector noise and, therefore, contains little information about θ . On the other hand, a value close to 1, which results from a large expectation term, indicates that the data in the k th pixel is minimally corrupted by detector noise and, therefore, contains close to the same amount of information about θ as the uncorrupted photoelectron count. Using the noise coefficient, comparison of the information content of the data in a pixel is thus reduced to a comparison of values between 0 and 1.

Noise coefficients α_k calculated using Eq. (38) with the probability density functions of Eqs. (32) and (34) for the CCD/CMOS and EMCCD data models, respectively, are shown in Fig. 8 as a function of the light level $\nu_{\theta,k}$. The noise coefficient curves corresponding to CCD detectors with 0.5- and 1-electron readout noise standard deviations can be seen to approach the maximum value of 1 at large values of $\nu_{\theta,k}$, demonstrating that, when enough photons can be detected in a pixel to render the readout noise insignificant, the data in the pixel is minimally corrupted and contains nearly the same amount of information about θ as the uncorrupted photoelectron count assumed by the Poisson data model. On the other hand, these two curves can be seen to approach

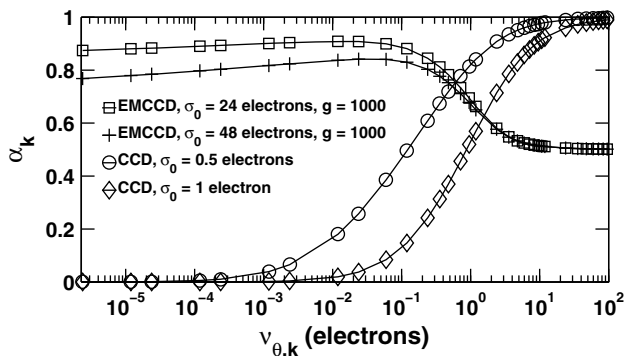


Fig. 8. Comparing noise coefficients corresponding to the CCD and EMCCD data models. The noise coefficient α_k for the k th pixel of an image is plotted as a function of the mean photoelectron count $\nu_{\theta,k}$ in the pixel. Noise coefficients are shown for two EMCCD detectors operated at an electron multiplication gain of $g = 1000$, with readout noise of standard deviations $\sigma_0 = 24$ electrons (\square) and 48 electrons ($+$) in each pixel, and for two CCD detectors, with readout noise of standard deviations $\sigma_0 = 0.5$ electrons (\circ) and 1 electron (\diamond) in each pixel. The readout noise mean is $\eta_0 = 0$ electrons in all cases.

the minimum value of 0 at values of $\nu_{\theta,k}$ that are significantly less than 1, illustrating that, when not enough photons can be detected in a pixel, the CCD detector's readout noise dominates the data in the pixel to such an extent that the data contains almost no information about θ . A comparison of the two CCD noise coefficient curves also shows that a smaller readout noise standard deviation will, as expected, result in data containing larger amounts of information about θ , particularly over the range of $\nu_{\theta,k}$ values between the convergence of the curves to 0 and the convergence of the curves to 1.

In terms of the EMCCD detector, Fig. 8 shows the noise coefficient curves corresponding to EMCCD detectors with 24- and 48-electron readout noise standard deviations and an electron multiplication gain of 1000 to both converge to the value 0.5 at large values of $\nu_{\theta,k}$. A noise coefficient value of 0.5 represents a halving of the maximum amount of information that the data in the pixel can carry about θ . As shown in [27], this halving of the maximum information content is consistent with the well-known excess noise factor [47,67,68] associated with stochastic electron multiplication, based on which it is commonly believed that the standard deviation with which a parameter can be estimated from an EMCCD image is, at best, the corresponding Poisson limit of accuracy multiplied by $\sqrt{2}$. While this is indeed approximately the case when a relatively large number of photons are detected in each pixel of an EMCCD image, there has been a misconception that the $\sqrt{2}$ penalty applies generally for all EMCCD images. The two EMCCD noise coefficient curves in Fig. 8 thus importantly show that, at values of $\nu_{\theta,k}$ less than 1, substantially more than half of the maximum information content is in fact available in the pixel. This means that, given an EMCCD image characterized by very low expected photoelectron counts in its pixels, the associated limits of accuracy are penalized, with respect to the corresponding Poisson limits of accuracy, by a factor that is potentially significantly less than $\sqrt{2}$ and substantially closer to 1. The two EMCCD noise coefficient curves further demonstrate that, for a given light level $\nu_{\theta,k}$, the information content of the data in the pixel can be pushed closer to the maximum possible by lowering the EMCCD detector's readout noise standard deviation. In [66], it is also shown that the same result can be realized by increasing the electron multiplication gain.

The noise coefficient curves of Fig. 8 thus corroborate the patterns exhibited by the CCD and EMCCD limits of accuracy in the example of Fig. 7(b). More generally, noise coefficient curves are useful because they allow an assessment over a range of light levels of the extent to which the information content of the data in a pixel is affected by a change in the detector type or a detector noise setting.

F. Effect of Additional Unknown Parameters

An important factor that can have an impact on the limit of the accuracy with which a parameter of interest (e.g., a positional coordinate of a molecule) can be estimated is the inclusion of other unknown parameters (e.g., the width of the point spread function that describes the molecule's image, the rates at which photons are detected from the molecule and the background component) in the vector θ of parameters to be estimated.

At best, the limit of the accuracy with which a parameter of interest can be estimated will not be affected by the inclusion of additional unknown parameters. In general, however, it can be expected to worsen, potentially significantly, as a result of the inclusion. Intuitively, this can be explained by the fact that having more unknowns will not contribute additional information about the parameters of interest and, if anything, can only make the overall estimation task more difficult.

Mathematically, the extent to which the limit of the accuracy for the determination of a parameter is deteriorated by the inclusion of other parameters in the estimation is related to the magnitudes of the off-diagonal elements of the Fisher information matrix. If element (i, j) of the matrix is small in magnitude, then the limit of the accuracy for estimating parameter θ_i can be expected to worsen by a relatively small amount as a result of the inclusion of parameter θ_j , and vice versa. In the special situation where the Fisher information matrix is diagonal (i.e., all off-diagonal elements are zero), the limit of the accuracy for estimating each parameter in θ is unaffected by the inclusion of the other parameters. This is, in fact, the case for the example of Section 4.A that considers the simultaneous estimation of a molecule's positional coordinates and the rate of photon detection from an image described by the fundamental data model. Note that, when the Fisher information matrix is diagonal, the second inequality in Eq. (17) becomes an equality, and the limit of the accuracy for estimating parameter θ_i is easily obtained as the square root of the inverse of the i th main diagonal element of the Fisher information matrix.

For a given problem, it is therefore useful to calculate and compare limits of accuracy corresponding to definitions of θ that include and exclude additional unknown parameters. If the limits of accuracy for estimating the parameters of interest deteriorate significantly with the inclusion of the additional unknown parameters, then one might consider independently estimating the additional unknown parameters and treating them as known parameters in the estimation of the parameters of interest. A critical assumption with this approach, however, is that the independent estimation of the additional unknown parameters is unbiased and accurate, so that these parameters effectively can be considered as known parameters. (For a study on how limits of accuracy are influenced by the inclusion of additional unknown parameters, see [69].)

G. Attainment of Limit of Accuracy

A limit of accuracy is a lower bound on the standard deviation with which a parameter of interest can be determined by any unbiased estimator. Given an unbiased estimator, the question then naturally arises as to whether it can determine a parameter of interest with a standard deviation that equals the limit of accuracy [i.e., whether equality is achieved in Eq. (17) between the variance of the estimate of the parameter and the corresponding element in the inverse Fisher information matrix]. This is often a difficult question to answer analytically; in some cases, it is possible that no unbiased estimator exists whose standard deviation attains the limit. In the context of single molecule microscopy, however, it has been analytically proven [21] that, in the special case of a stationary and in-focus molecule whose image is described by the 2D Gaussian function of

Eq. (3) and is acquired under the conditions of the fundamental data model, with the exception that the same number of photons is guaranteed to be detected in each repeat image, the maximum likelihood estimator of the molecule's x_0 and y_0 positional coordinates does, in fact, determine those parameters with standard deviations that equal their respective limits of accuracy (which happen to be identical). In terms of pixelated image data, studies using simulated data (e.g., [17,21,26,27,55]) have shown that, in many practical scenarios, the maximum likelihood estimator can determine a parameter of interest with a standard deviation that comes close to the limit of accuracy.

5. EXTENSION TO MULTIPLE IMAGES

A basic assumption that is made throughout the presentation of Sections 4.A through 4.D is that parameters of interest are estimated from data consisting of a single image. The examples given therefore involve limits of accuracy that are calculated for estimation based on a single image. In many situations, however, the data from which parameters are to be determined comprises multiple images. The data can, for example, consist of a time sequence of images that capture the trajectory of a moving object, or comprise a set of images of an object that have been simultaneously acquired from distinct focal planes within a sample. In the former case, Cramér–Rao lower bounds have been derived [70] that predict the accuracies with which parameters such as the starting positional coordinates and the speed of an object moving in a trajectory of known shape can be estimated from an image sequence. In the latter case, limits of accuracy have been calculated [8,24] to determine how well the axial coordinate z_0 of a molecule can be estimated from a set of images acquired using multifocal plane microscopy (MUM) [71,72] or a modality that implements the same principle of simultaneously imaging different focal planes within a sample [73–75]. Another example relates to the determination of distances of separation, where limits of accuracy have been computed [25,76] for cases where images in which only one of the two molecules is detected are used, potentially in conjunction with an image in which both molecules are detected, to estimate the distance separating the two molecules.

Provided that the N_{im} images comprising a set are generated by stochastically independent photon detection processes and detector noise processes, the Fisher information matrix $\mathbf{I}_{set}(\theta)$ corresponding to the set of N_{im} images is given simply by the sum of the Fisher information matrices corresponding to the individual images. Denoting the matrix for the k th image by $\mathbf{I}_k(\theta)$, $k = 1, \dots, N_{im}$, the matrix $\mathbf{I}_{set}(\theta)$ can be written as

$$\mathbf{I}_{set}(\theta) = \sum_{k=1}^{N_{im}} \mathbf{I}_k(\theta), \quad \theta \in \Theta. \quad (39)$$

Note that this result follows from the same principle that underlies the fact that, by the mutual independence of the data in its component pixels, the Fisher information matrix for a pixelated image is the sum of the Fisher information matrices corresponding to the individual pixels [see Eq. (28)]. Indeed, a set of mutually independent pixelated images may be viewed as a single image comprising all the pixels, as the corresponding

Fisher information matrix $\mathbf{I}_{\text{set}}(\theta)$ is effectively a summation of the matrices corresponding to all the pixels in the set.

The use of MUM to overcome the poor depth discrimination of conventional microscopy provides a practical illustration of the calculation of limits of accuracy with which parameters can be estimated from a set of images. The poor depth discrimination capability of a conventional microscopy setup, in which images of an object of interest are captured from a single focal plane, refers to the fact that the closer an object is axially located with respect to the focal plane, the significantly more difficult it becomes to accurately determine its axial coordinate z_0 . This phenomenon can be appreciated by looking at the limit of accuracy δ_{z_0} for the estimation of z_0 as a function of z_0 itself. An example is shown in Fig. 9, where δ_{z_0} is calculated for the case where a CCD detector is used to image an out-of-focus and stationary molecule of interest. In the figure, δ_{z_0} corresponding to a conventional microscopy setup is seen to worsen, in severe fashion, as the axial location of the molecule approaches the focal plane at $z_0 = 0$ nm from either above or below. In this example, δ_{z_0} is computed for an image acquired over the time interval $[t_0, t]$, using the Fisher information matrix of Eq. (33) by assuming a constant rate of detection $\Lambda(\tau) = \Lambda_0$, $\tau \geq t_0$, of photons from the molecule, and a distribution profile for those photons given by f_θ of Eq. (1), without the time dependence and with the image function q_{z_0} given by the Born and Wolf image function of Eq. (4). The parameter vector is given by $\theta = (x_0, y_0, z_0) \in \mathbb{R}^3$, and a time-independent and uniformly distributed background component is assumed.

Different solutions to the depth discrimination problem have been proposed and demonstrated [71,77–80], and MUM is one such solution. For a two-plane MUM solution in which a pair of images acquired from two different focal planes are used for the estimation of θ , the Fisher information matrix according to Eq. (39) is given by $\mathbf{I}_{\text{set}}(\theta) = \mathbf{I}_1(\theta) + \mathbf{I}_2(\theta)$, where $\mathbf{I}_1(\theta)$ and $\mathbf{I}_2(\theta)$ are the Fisher information matrices corresponding to the images captured from focal planes 1 and 2. Assuming an equal splitting of the photons collected by the objective lens between the two images, and taking focal plane 1 to be the focal plane of the conventional setup, $\mathbf{I}_1(\theta)$ is given as described for the Fisher information matrix for the conventional setup, with the rates of photon and background detection reduced by 50%. The expression for $\mathbf{I}_2(\theta)$ is of the same form as $\mathbf{I}_1(\theta)$ but includes two modifications. First, assuming that focal plane 2 is positioned at a distance of Δz_f above focal plane 1, the axial coordinate of the molecule with respect to focal plane 2 is given by $z_0 - \Delta z_f$. Second, the lateral magnification associated with focal plane 2 is different from the lateral magnification associated with focal plane 1 and is, in this example, determined using a geometrical optics-based formula from [76].

Figure 9 shows that δ_{z_0} computed using $\mathbf{I}_{\text{set}}(\theta)$ for a two-plane MUM setup with a plane spacing of $\Delta z_f = 450$ nm does not exhibit any severe deterioration at either focal plane. In fact, though it is not better than δ_{z_0} for the conventional setup at all values of z_0 , δ_{z_0} for the two-plane MUM setup remains relatively constant and small throughout the 1.5- μm axial range shown.

The improvement that a MUM setup offers over a conventional setup is dependent on the number of focal planes used,

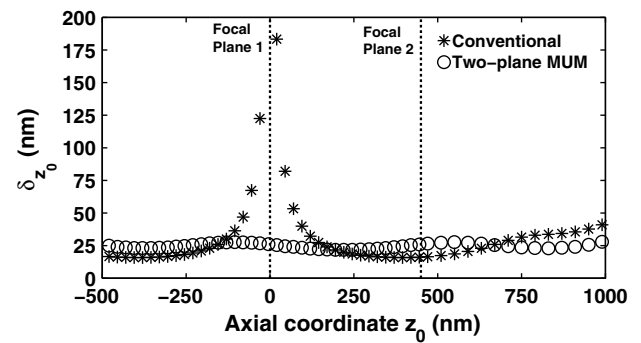


Fig. 9. Limits of the axial coordinate estimation accuracy for a conventional microscopy setup and a two-plane MUM setup. The limit of accuracy δ_{z_0} for the estimation of the z_0 coordinate of an out-of-focus and stationary molecule from a conventional image (*) or a pair of two-plane MUM images (O) is plotted as a function of the coordinate z_0 . Focal planes 1 and 2 of the MUM setup are denoted by vertical dashed lines and are located at $z_0 = 0$ nm and $z_0 = 450$ nm, respectively. The value of z_0 is specified with respect to focal plane 1, which coincides with the focal plane of the conventional setup. In both setups, an image consists of an 11×11 -pixel region, acquired using a CCD detector with a pixel size of $13 \mu\text{m} \times 13 \mu\text{m}$ and readout noise with mean $\eta_0 = 0$ electrons and standard deviation $\sigma_0 = 6$ electrons at each pixel. The image of the molecule is described by the Born and Wolf image function. For the conventional setup, the molecule is laterally positioned such that the center of its image is located at 5.05 pixels in the x direction and 5.25 pixels in the y direction. Photons are detected from the molecule at a rate of $\Lambda_0 = 10,000$ photons/s over an acquisition time of $t - t_0 = 0.1$ s, such that $\Lambda_0 \cdot (t - t_0) = 1000$ photons are on average detected from the molecule over the detector plane (and 952.3 photons are on average detected from the molecule over a given image when the molecule is in focus). A background detection rate of $\Lambda_0^b = 24,200$ photons/s is assumed, along with a uniform background spatial distribution of $f^b(x, y) = 1/20,449 \mu\text{m}^{-2}$, such that there is a background level of $\beta_0 = 20$ photoelectrons per pixel over the 11×11 -pixel image. Photons of wavelength $\lambda = 520$ nm are collected by an objective lens with magnification $M = 100$ and numerical aperture $n_a = 1.4$. The refractive index of the immersion medium is $n = 1.518$. For the MUM setup, the collected molecule and background photons are split equally between the images corresponding to the two focal planes. Therefore, the photon detection rate is $\Lambda_0 = 5000$ photons/s per image, and the background detection rate is $\Lambda_0^b = 12,100$ photons/s per image. For the image corresponding to focal plane 1, all other details are as given for the conventional setup. For the image corresponding to focal plane 2, the magnification is $M = 98.18$, calculated using a formula from [76] for a plane spacing of $\Delta z_f = 450$ nm and a tube length of 160 mm. The lateral position of the image of the molecule and the background spatial distribution are accordingly adjusted based on this slightly smaller magnification.

the spacings between the focal planes, and other experimental parameters. (For a Fisher information-based analysis of the effects of these factors on the limit of accuracy δ_{z_0} for a MUM setup, see [62].)

6. CONCLUSION

By taking into account the multitude of factors that influence the accuracy with which a parameter of interest can be estimated from an image, the information theory-based limit of

accuracy represents a particularly useful tool for estimator evaluation and experimental design in single molecule microscopy. This tutorial has demonstrated the use of the mathematical framework introduced in [21,22] over a decade ago to calculate limits of accuracy for an estimation problem. Specifically, it has described the photon detection process, various image data models, and various Fisher information expressions that constitute the ingredients for the computation of limits of accuracy. Importantly, the tutorial has used specific examples of estimation problems to illustrate the effects that different experimental parameters and object attributes have on the accuracy for estimating a parameter of interest. These examples demonstrate that, by calculating and comparing limits of accuracy corresponding to different detector types and different values for parameters such as the expected number of detected photons, the wavelength of the detected photons, the effective pixel size, and the detector readout noise level, one can arrive at appropriate choices of detector, fluorophore, magnification, readout setting, and other aspects of an experimental setup that will yield the desired level of accuracy for the estimation of a parameter of interest from the acquired image data. The material presented in the tutorial represents a bringing together of some major results produced over the last decade or so that have made use of the mathematical framework. These results exemplify the ability of the framework to accommodate the analysis of different estimation problems. As the use of the Fisher information-based limit of accuracy becomes increasingly more common in the field of single molecule microscopy, it is expected that this framework will continue to serve as a useful blueprint for its calculation.

Funding. National Institutes of Health (NIH) (R01 GM085575).

REFERENCES

- W. E. Moerner and D. P. Fromm, "Methods of single-molecule fluorescence spectroscopy and microscopy," *Rev. Sci. Instrum.* **74**, 3597–3619 (2003).
- N. G. Walter, C.-Y. Huang, A. J. Manzo, and M. A. Sobhy, "Do-it-yourself guide: how to use the modern single-molecule toolkit," *Nat. Methods* **5**, 475–489 (2008).
- P. R. Smith, I. E. G. Morrison, K. M. Wilson, N. Fernández, and R. J. Cherry, "Anomalous diffusion of major histocompatibility complex class I molecules on HeLa cells determined by single particle tracking," *Biophys. J.* **76**, 3331–3344 (1999).
- R. Iino, I. Koyama, and A. Kusumi, "Single molecule imaging of green fluorescent proteins in living cells: E-cadherin forms oligomers on the free cell surface," *Biophys. J.* **80**, 2667–2677 (2001).
- R. J. Ober, C. Martinez, X. Lai, J. Zhou, and E. S. Ward, "Exocytosis of IgG as mediated by the receptor, FcRn: an analysis at the single-molecule level," *Proc. Natl. Acad. Sci. USA* **101**, 11076–11081 (2004).
- J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, "Probing gene expression in live cells, one protein molecule at a time," *Science* **311**, 1600–1603 (2006).
- T. Dange, D. Grünwald, A. Grünwald, R. Peters, and U. Kubitschek, "Autonomy and robustness of translocation through the nuclear pore complex: a single-molecule study," *J. Cell Biol.* **183**, 77–86 (2008).
- S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, "High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells," *Biophys. J.* **95**, 6025–6043 (2008).
- M. A. Thompson, J. M. Casolari, M. Badieirostami, P. O. Brown, and W. E. Moerner, "Three-dimensional tracking of single mRNA particles in *Saccharomyces cerevisiae* using a double-helix point spread function," *Proc. Natl. Acad. Sci. USA* **107**, 17864–17871 (2010).
- H. Deschout, F. Cella Zanacchi, M. Mlodzianowski, A. Diaspro, J. Bewersdorf, S. T. Hess, and K. Braeckmans, "Precisely and accurately localizing single emitters in fluorescence microscopy," *Nat. Methods* **11**, 253–266 (2014).
- A. Small and S. Stahlheber, "Fluorophore localization algorithms for super-resolution microscopy," *Nat. Methods* **11**, 267–279 (2014).
- E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science* **313**, 1642–1645 (2006).
- S. T. Hess, T. P. K. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.* **91**, 4258–4272 (2006).
- M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods* **3**, 793–795 (2006).
- M. Heilemann, S. van de Linde, M. Schüttel, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer, "Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes," *Angew. Chem. Int. Ed.* **47**, 6172–6176 (2008).
- M. K. Cheezum, W. F. Walker, and W. H. Guilford, "Quantitative comparison of algorithms for tracking single fluorescent particles," *Biophys. J.* **81**, 2378–2388 (2001).
- 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**, 23352–23373 (2009).
- Z. Shen and S. B. Andersson, "Bias and precision of the fluoroBancroft algorithm for single particle localization in fluorescence microscopy," *IEEE Trans. Signal Process.* **59**, 4041–4046 (2011).
- X. Michalet, "Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium," *Phys. Rev. E* **82**, 041914 (2010).
- S. Liu, E. B. Kromann, W. D. Krueger, J. Bewersdorf, and K. A. Lidke, "Three dimensional single molecule localization using a phase retrieved pupil function," *Opt. Express* **21**, 29462–29487 (2013).
- R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophys. J.* **86**, 1185–1200 (2004).
- S. Ram, E. S. Ward, and R. J. Ober, "A stochastic analysis of performance limits for optical microscopes," *Multidimens. Syst. Signal Process.* **17**, 27–57 (2006).
- S. Ram, E. S. Ward, and R. J. Ober, "How accurately can a single molecule be localized in three dimensions using a fluorescence microscope?" *Proc. SPIE* **5699**, 426–435 (2005).
- S. Ram, J. Chao, P. Prabhat, E. S. Ward, and R. J. Ober, "A novel approach to determining the three-dimensional location of microscopic objects with applications to 3D particle tracking," *Proc. SPIE* **6443**, 64430D (2007).
- S. Ram, E. S. Ward, and R. J. Ober, "Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy," *Proc. Natl. Acad. Sci. USA* **103**, 4457–4462 (2006).
- J. Chao, S. Ram, A. V. Abraham, E. S. Ward, and R. J. Ober, "A resolution measure for three-dimensional microscopy," *Opt. Commun.* **282**, 1751–1761 (2009).
- J. Chao, E. S. Ward, and R. J. Ober, "Fisher information matrix for branching processes with application to electron-multiplying charge-coupled devices," *Multidimens. Syst. Signal Process.* **23**, 349–379 (2012).
- C. R. Rao, *Linear Statistical Inference and Its Applications* (Wiley, 1965).
- S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993), Vol. 1.
- F. Aguet, D. Van De Ville, and M. Unser, "A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles," *Opt. Express* **13**, 10503–10522 (2005).
- L. Holtzer, T. Meckel, and T. Schmidt, "Nanometric three-dimensional tracking of individual quantum dots in cells," *Appl. Phys. Lett.* **90**, 053902 (2007).
- S. R. P. Pavani and R. Piestun, "Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system," *Opt. Express* **16**, 22048–22057 (2008).

33. F. Aguet, S. Geissbühler, I. Märki, T. Lasser, and M. Unser, "Super-resolution orientation estimation and localization of fluorescent dipoles using 3-D steerable filters," *Opt. Express* **17**, 6829–6848 (2009).
34. M. A. Thompson, M. D. Lew, M. Badieirostami, and W. E. Moerner, "Localizing and tracking single nanoscale emitters in three dimensions with high spatiotemporal resolution using a double-helix point spread function," *Nano Lett.* **10**, 211–218 (2010).
35. S. Stallinga and B. Rieger, "Position and orientation estimation of fixed dipole emitters using an effective Hermite point spread function model," *Opt. Express* **20**, 5896–5921 (2012).
36. M. R. Foreman and P. Török, "Fundamental limits in single-molecule orientation measurements," *New J. Phys.* **13**, 093013 (2011).
37. A. Agrawal, S. Quirin, G. Grover, and R. Piestun, "Limits of 3D dipole localization and orientation estimation for single-molecule imaging: towards Green's tensor engineering," *Opt. Express* **20**, 26667–26680 (2012).
38. A. J. Berglund, "Statistics of camera-based single-particle tracking," *Phys. Rev. E* **82**, 011917 (2010).
39. X. Michalet and A. J. Berglund, "Optimal diffusion coefficient estimation in single-particle tracking," *Phys. Rev. E* **85**, 061916 (2012).
40. Y. Shechtman, S. J. Sahl, A. S. Backer, and W. E. Moerner, "Optimal point spread function design for 3D imaging," *Phys. Rev. Lett.* **113**, 133902 (2014).
41. R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescent probes," *Biophys. J.* **82**, 2775–2783 (2002).
42. T. Quan, S. Zeng, and Z.-L. Huang, "Localization capability and limitation of electron-multiplying charge-coupled, scientific complementary metal-oxide semiconductor, and charge-coupled devices for super-resolution imaging," *J. Biomed. Opt.* **15**, 066005 (2010).
43. K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, "Optimized localization analysis for single-molecule tracking and super-resolution microscopy," *Nat. Methods* **7**, 377–381 (2010).
44. H. Deschout, K. Neyts, and K. Braeckmans, "The influence of movement on the localization precision of sub-resolution particles in fluorescence microscopy," *J. Biophoton.* **5**, 97–109 (2012).
45. J. R. Janesick, *Scientific Charge-Coupled Devices* (SPIE, 2001).
46. J. Hynecek, "Impactron—a new solid state image intensifier," *IEEE Trans. Electron Devices* **48**, 2238–2241 (2001).
47. J. Hynecek and T. Nishiwaki, "Excess noise and other important characteristics of low light level imaging using charge multiplying CCDs," *IEEE Trans. Electron Devices* **50**, 239–245 (2003).
48. E. R. Fossum, "CMOS image sensors: electronic camera-on-a-chip," *IEEE Trans. Electron Devices* **44**, 1689–1698 (1997).
49. M. Bigas, E. Cabruja, J. Forest, and J. Salvi, "Review of CMOS image sensors," *Microelectron. J.* **37**, 433–451 (2006).
50. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, 1996).
51. J. Chao, T. Lee, E. S. Ward, and R. J. Ober, "Fluorescent microspheres as point sources: a localization study," *PLoS ONE* **10**, e0134112 (2015).
52. M. Born and E. Wolf, *Principles of Optics* (Cambridge University, 1999).
53. B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," *Appl. Opt.* **46**, 1819–1829 (2007).
54. S. Stallinga and B. Rieger, "Accuracy of the Gaussian point spread function model in 2D localization microscopy," *Opt. Express* **18**, 24461–24476 (2010).
55. C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," *Nat. Methods* **7**, 373–375 (2010).
56. F. Huang, T. M. P. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, M. W. Davidson, D. Toomre, and J. Bewersdorf, "Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms," *Nat. Methods* **10**, 653–658 (2013).
57. T. E. Harris, *The Theory of Branching Processes* (Prentice-Hall, 1963).
58. A. G. Basden, C. A. Haniff, and C. D. Mackay, "Photon counting strategies with low-light-level CCDs," *Mon. Not. R. Astron. Soc.* **345**, 985–991 (2003).
59. M. H. Ulbrich and E. Y. Isacoff, "Subunit counting in membrane-bound proteins," *Nat. Methods* **4**, 319–321 (2007).
60. M. Newberry, "Pixel response effects on CCD camera gain calibration," Technical Note (Mirametrix, 1998), http://www.mirametrix.com/tech_note_ccdgain.htm.
61. S. Ram, "Resolution and localization in single molecule microscopy," Ph.D. thesis (University of Texas at Arlington, 2007).
62. A. Tahmasbi, S. Ram, J. Chao, A. V. Abraham, F. W. Tang, E. S. Ward, and R. J. Ober, "Designing the focal plane spacing for multifocal plane microscopy," *Opt. Express* **22**, 16706–16721 (2014).
63. S. Inoué, "Foundations of confocal scanned imaging in light microscopy," in *Handbook of Biological Confocal Microscopy*, J. B. Pawley, ed., 3rd ed., (Springer, 2006), pp. 1–19.
64. A. Krull, A. Steinborn, V. Ananthanarayanan, D. Ramunno-Johnson, U. Petersohn, and I. M. Tolić-Nørrelykke, "A divide and conquer strategy for the maximum likelihood localization of low intensity objects," *Opt. Express* **22**, 210–228 (2014).
65. D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White, "Compensation for readout noise in CCD images," *J. Opt. Soc. Am. A* **12**, 272–283 (1995).
66. J. Chao, S. Ram, E. S. Ward, and R. J. Ober, "Ultra-high accuracy imaging modality for super-localization microscopy," *Nat. Methods* **10**, 335–338 (2013).
67. K. Matsuo, M. C. Teich, and B. E. A. Saleh, "Noise properties and time response of the staircase avalanche photodiode," *IEEE Trans. Electron Devices* **32**, 2615–2623 (1985).
68. J. N. Hollenhorst, "A theory of multiplication noise," *IEEE Trans. Electron Devices* **37**, 781–788 (1990).
69. Z. Lin, Y. Wong, and R. J. Ober, "Influence of prior knowledge on the accuracy limit of parameter estimation in single-molecule fluorescence microscopy," in *Proceedings of IEEE International Symposium on Circuits and Systems* (IEEE, 2013), pp. 1304–1307.
70. Y. Wong, J. Chao, Z. Lin, and R. J. Ober, "Effect of time discretization of the imaging process on the accuracy of trajectory estimation in fluorescence microscopy," *Opt. Express* **22**, 20396–20420 (2014).
71. P. Prabhat, S. Ram, E. S. Ward, and R. J. Ober, "Simultaneous imaging of different focal planes in fluorescence microscopy for the study of cellular dynamics in three dimensions," *IEEE Trans. Nanobiosci.* **3**, 237–242 (2004).
72. P. Prabhat, Z. Gan, J. Chao, S. Ram, C. Vaccaro, S. Gibbons, R. J. Ober, and E. S. Ward, "Elucidation of intracellular recycling pathways leading to exocytosis of the Fc receptor, FcRn, by using multifocal plane microscopy," *Proc. Natl. Acad. Sci. USA* **104**, 5889–5894 (2007).
73. P. M. Blanchard and A. H. Greenaway, "Simultaneous multiplane imaging with a distorted diffraction grating," *Appl. Opt.* **38**, 6692–6699 (1999).
74. S. Abrahamsson, J. Chen, B. Hajj, S. Stallinga, A. Y. Katsov, J. Wisniewski, G. Mizuguchi, P. Soule, F. Mueller, C. D. Darzacq, X. Darzacq, C. Wu, C. I. Bargmann, D. A. Agard, M. Dahan, and M. G. L. Gustafsson, "Fast multicolor 3D imaging using aberration-corrected multifocus microscopy," *Nat. Methods* **10**, 60–63 (2013).
75. M. F. Juetter, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, "Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples," *Nat. Methods* **5**, 527–529 (2008).
76. J. Chao, S. Ram, E. S. Ward, and R. J. Ober, "A comparative study of high resolution microscopy imaging modalities using a three-dimensional resolution measure," *Opt. Express* **17**, 24377–24402 (2009).
77. H. P. Kao and A. S. Verkman, "Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position," *Biophys. J.* **67**, 1291–1300 (1994).
78. M. Speidel, A. Jonás, and E.-L. Florin, "Three-dimensional tracking of fluorescent nanoparticles with subnanometer precision by use of off-focus imaging," *Opt. Lett.* **28**, 69–71 (2003).
79. A. Greengard, Y. Y. Schechner, and R. Piestun, "Depth from diffracted rotation," *Opt. Lett.* **31**, 181–183 (2006).
80. Y. Sun, J. D. McKenna, J. M. Murray, E. M. Ostap, and Y. E. Goldman, "Parallax: high accuracy three-dimensional single molecule tracking using split images," *Nano Lett.* **9**, 2676–2682 (2009).