Estimation of the diffusion constant from intermittent trajectories with variable position uncertainties

Peter K. Relich,[1] Mark J. Olah,[1] Patrick J. Cutler,[2] and Keith A. Lidke[1]
[1]*Department of Physics and Astronomy, University of New Mexico, Mexico*
[2]*Department of Pathology, University of New Mexico, Mexico*
(Received 13 August 2015; revised manuscript received 19 January 2016; published 4 April 2016; corrected 13 April 2016)

The movement of a particle described by Brownian motion is quantified by a single parameter, D, the diffusion constant. The estimation of D from a discrete sequence of noisy observations is a fundamental problem in biological single-particle tracking experiments since it can provide information on the environment and/or the state of the particle itself via the hydrodynamic radius. Here, we present a method to estimate D that takes into account several effects that occur in practice, important for the correct estimation of D, and that have hitherto not been combined together for an estimation of D. These effects are motion blur from the finite integration time of the camera, intermittent trajectories, and time-dependent localization uncertainty. Our estimation procedure, a maximum-likelihood estimation with an information-based confidence interval, follows directly from the likelihood expression for a discretely observed Brownian trajectory that explicitly includes these effects. We begin with the formulation of the likelihood expression and then present three methods to find the exact solution. Each method has its own advantages in either computational robustness, theoretical insight, or the estimation of hidden variables. The Fisher information for this likelihood distribution is calculated and analyzed to show that localization uncertainties impose a lower bound on the estimation of D. Confidence intervals are established and then used to evaluate our estimator on simulated data with experimentally relevant camera effects to demonstrate the benefit of incorporating variable localization errors.

DOI: 10.1103/PhysRevE.93.042401

I. INTRODUCTION

Single-particle tracking (SPT) is a method to observe and classify the motion of individual particles as *trajectories*: estimates of a particle's position in a sequence of discrete measurement times. In the field of biological microscopy, SPT has been used for finding and analyzing protein motion in heterogeneous environments such as the cellular membrane [1,2] and cytoplasm [3,4]. SPT trajectory information can be used to resolve variations in the individual motion of molecules that would otherwise be lost in ensemble imaging techniques.

In the analysis of trajectories, the pure Brownian motion model is often the first model used to describe a trajectory in the absence of prior information about particle behavior. The behavior of a single particle dominated by Brownian motion can be described by a normal distribution with the variance term proportional to a single physical scale parameter, D, the diffusion constant, which makes Brownian motion the simplest model for describing stochastic motion. More complicated behavior could potentially be modeled as Brownian motion with discrete changes in the diffusion constant that could be identified with change point analysis [5]. The estimation of the diffusion constant of a particle from discrete, noisy, and possibly short particle trajectories is a fundamental problem in single-particle tracking.

In this paper, we focus on the likelihood distribution of D. We present a maximum-likelihood-based approach for estimating the diffusion constant of a particle conditional on information from an SPT trajectory that includes the individual localization error for each position in the trajectory, the time of the observation, and the camera integration time. Our approach is based on a direct solution to the likelihood equation for the observation of a particular trajectory. The need for such an estimation procedure has evolved out of the rapid progress that has been made in SPT analysis techniques over the past few years [6–10]. On the experimental side of this problem, fluorescent probes are subjected to the phenomenon of fluorescence intermittency, where probes stochastically switch between emission states over various time scales. The effects of stochastic switching of emission states typically result in interframe blinking or a variability in photon emissions from frame to frame that is poorly described by a single Poisson distribution. The phenomena of fluorescence intermittency has guided the development of emitter localization techniques so that they can not only accurately resolve the location of an emitter to tens of nanometers, but also reliably estimate the localization error of single molecules [11]. Even in the case in which a probe is specifically designed to minimize the effect of fluorescence intermittency (e.g., giant quantum dots [12]), the signal-to-noise ratio of a particle can still vary significantly from frame to frame in an image sequence due to environmental factors such as varying background signals or photobleaching of a redundantly labeled receptor. All of these factors cause the point emitter localization estimation accuracy to fluctuate from frame-to-frame. We have therefore developed an estimator that takes into account the available localization variance estimates to accurately weight the influence of each localization data point in the estimate.

A. Background and related work

Historically, one of the primary techniques for estimating the diffusion constant from trajectories relied on a linear regression of the mean-squared displacement (MSD) of the tracked particle coordinates as a function of time lag [13]. In the absence of measurement errors, the observed MSD

for pure Brownian motion scales linearly with time lag and intersects at the origin, allowing the direct recovery of the diffusion constant from a linear regression on the well-sampled data points. It has been shown that a regression of the MSD with an offset parameter can be interpreted to account for the cumulative effects of static [14] and dynamic measurement errors [15]. If the MSD is built using the same data points for multiple time lags, the correlation between MSD values must also be taken into account in the regression [13,16,17]. Although it seems theoretically possible to include individual localization error in the MSD regression, to date this has not been described.

A separate line of work has focused on maximum-likelihood approaches to the estimation procedure. A maximum-likelihood estimator works by finding the maximum of a likelihood function $\mathcal{L}(D) = P(O|D)$ that gives the probability of observing a particular trajectory O, given a diffusion constant D. Ideally this probability should incorporate both the variable localization errors of the trajectory and the effect of motion-blur. The motion-blur effect arises from the fact that each localization is performed on data that are acquired over some nonzero exposure time. Typically camera sensors integrate the signal over the exposure time, resulting in a blurring of the particle image. This blurring has important numerical effects on the likelihood function [18]. A specific solution to the likelihood function has been accurately derived that incorporates the effects of motion-blur but with the caveat that only a single global localization error estimate is used as an input or estimated [17,19]. This maximum-likelihood estimator (MLE) is a more robust alternative to the MSD-based estimators because it can incorporate information from each jump directly into the likelihood, unlike the MSD analysis that averages over most of the information before subsequently fitting with a linear regression. Subsequent work has extended the MLE approach to deal with nonuniformly spaced or intermittent trajectories [20], however the associated software implementation did not incorporate the effects of motion blur. Maximum-likelihood estimators are not the only class of diffusion estimators that have evolved recently; continued development on displacement-based estimators has resulted in an estimator that incorporates the effects of covariances between sequentially observed displacements [21].

In this work, we provide a generalized solution to the likelihood function, incorporating variable localization errors and variable displacement periods, which results in an improvement in estimation accuracy for short trajectories, trajectories with large variations in localization accuracy, and trajectories with intermittently spaced measurements. In Sec. II we formulate the diffusion likelihood function to directly incorporate the effects of motion-blur, variable localization errors, and intermittent or nonuniformly spaced observations in time. We present three independent solutions to this likelihood function. The first derivation, the recursive method (Sec. III A), is a sequential integration of the nuisance parameters and provides the fastest numerical implementation. The second derivation, the Laplace method (Sec. III B), utilizes a second-order Taylor expansion to express the likelihood as a multivariate Gaussian in the basis of integration. The Laplace method additionally returns the maximum-likelihood values of the true positions given a sampled D. The third derivation, the Markov method (Sec. III C), calculates the characteristic function in order to express the likelihood in the basis of displacements. The Markov method allows us to verify that the generalized form of the expression derived in Ref. [19] is the same distribution as the expressions derived in this paper. The Markov method was also instrumental in determining the coefficients necessary to reduce the computational complexity of all the methods (Appendix C). Each of these derivations leads to an independent, numerically accurate computational algorithm for estimating the likelihood of D (Sec. IV), making full use of all the information contained in a noisy trajectory. The resulting likelihood calculation allows for robust computations in specific problems, such as a maximum-likelihood estimator, confidence intervals, or change point analysis.

The likelihood distribution is then applied to an information-based analysis (Sec. V) to find a robust parametrization of the likelihood distribution. We chose to parametrize the likelihood distribution by a log-normal approximation consisting of the maximum-likelihood estimate and the observed information. The log-normal approximation subsequently allows us to develop a simple, approximate confidence interval on D estimates. We tested the methods against an experimentally relevant simulation (Sec. VI B) with varying background values to show the performance improvement when dealing with trajectories with naturally generated variable localization errors. The significance of the observed information as a likelihood distribution parameter is investigated more deeply in Sec. VII A, and the predictive power of the expected information is investigated in Sec. VII B. Finally, we use the intuition gained from the previous sections to evaluate the effectiveness of the MLE analysis in Sec. VII C. The corresponding parameters calculated from our likelihood distribution are compared to parameters calculated from the likelihood distribution described in Ref. [20]. It is shown that our improved likelihood distribution results in an MLE estimator that is more robust than prior MLE implementations under experimentally relevant conditions.

II. PROBLEM FORMULATION

If a diffusing particle is accurately and exactly observed at a discrete sequence of N positions $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$ at times t_i, then $P(\mathbf{X}|D)$, the probability of sequence $\mathbf{X}$ given diffusion constant D and initial position $\mathbf{x}_1$, is

$$P(\mathbf{X}|D) = \prod_{i=1}^{N-1} P(\mathbf{x}_{i+1}|\mathbf{x}_i). \tag{1}$$

In Eq. (1), $P(\mathbf{x}_{i+1}|\mathbf{x}_i) = P(\mathbf{x}_{i+1}|\mathbf{x}_i, D)$ is the probability density of each discrete jump from $\mathbf{x}_i \to \mathbf{x}_{i+1}$ over time step $\delta t_i = t_{i+1} - t_i$ given diffusion constant D. We will omit explicit representation of D whenever we refer to the transition probability density, $P(\mathbf{x}_{i+1}|\mathbf{x}_i)$, because it is assumed here that a diffusion process parametrized by a single D is the only generator of positional displacements for a perfectly measured trajectory.

When measured experimentally, however, the true positions $\mathbf{X}$ are never known exactly, but they are conditionally related to observed positions $\mathbf{O}$ by some set of probability density

functions describing the localization uncertainty. If the measurements are obtained from signals that are collected at instantaneous periods, such as a camera with small exposure times relative to the elapsed periods between the image acquisitions ($t_\varepsilon \ll \delta t_i$), then the conditional expected value of each $\mathbf{o}_i$ is a function of the corresponding $\mathbf{x}_i$. Therefore, the conditional probability density of each $\mathbf{o}_i$ is described by $P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{v}_i)$, where here we take $\mathbf{v}_i$ to mean the variance of a normally distributed uncertainty. For instantaneous measurements, $\mathbf{O}$ is related to $\mathbf{X}$ in the joint distribution

$$P(\mathbf{O},\mathbf{X}|\mathbf{V},D) = \prod_{i=1}^{N} P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{v}_i) \prod_{j=1}^{N-1} P(\mathbf{x}_{j+1}|\mathbf{x}_j). \quad (2)$$

In many experiments, the exposure time, t_ε, of the camera is a significant fraction of the period between image acquisitions so that Eq. (2) is no longer an accurate description of the joint distribution. When exposure times are non-negligible, $\mathbf{O}$ no longer represents a set of instantaneous measurements but rather a set of time-integrated measurements. The integration of the image signal over a finite exposure period consequently implies that the conditional expectation value of each $\mathbf{o}_i$ must be described by at least two of the instantaneous positions from $\mathbf{X}$, i.e., one position prior to and one position after the acquisition of the associated $\mathbf{o}_i$. For instantaneous measurements, each $\mathbf{o}_i$ is described as a function of the corresponding $\mathbf{x}_i$. To maintain a convention so that Eq. (2) is a special case of a more general distribution, we will define the integrated measurements for each $\mathbf{o}_i$ as functions of both $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$. In our description of the problem, the acquisition of $\mathbf{o}_i$ begins instantaneously after the true but unknown $\mathbf{x}_i$ and ends after an elapsed time t_ε before the true but unknown $\mathbf{x}_{i+1}$. Under this experimental model, $P(\mathbf{O},\mathbf{X}|\mathbf{V},D)$, the combined likelihood of the observed positions $\mathbf{O}$ and the actual positions $\mathbf{X}$ is a product of the observation probability densities $P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{x}_{i+1},\mathbf{v}_i,D)$ and the diffusion transition probability densities $P(\mathbf{x}_{i+1}|\mathbf{x}_i)$ for each of the N observed positions and displacements,

$$P(\mathbf{O},\mathbf{X}|\mathbf{V},D) = \prod_{i=1}^{N} P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{x}_{i+1},\mathbf{v}_i,D)P(\mathbf{x}_{i+1}|\mathbf{x}_i). \quad (3)$$

For simplicity, we omit explicit representations of $\mathbf{V}$ in all subsequent probability distributions as it is assumed that the localization estimator returns both $\mathbf{V}$ and $\mathbf{O}$ together. Furthermore, we will simplify the representation of the observational probability distributions with the definition $P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{x}_{i+1}) = P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{x}_{i+1},\mathbf{v}_i,D)$. Note that our convention for establishing Eq. (3) implies that $N+1$ elements of $\mathbf{X}$ are required to describe N elements of $\mathbf{O}$. Since $\mathbf{X}$ is unknown for experimental data, we integrate Eq. (3) over all possible $\mathbf{X}$ to marginalize out the dependence on $\mathbf{X}$, and we write the diffusion likelihood as an integral over the space of all $\mathbf{X}$ values,

$$\begin{aligned} P(\mathbf{O}|D) &= \int d\mathbf{X}\, P(\mathbf{O},\mathbf{X}|D) \\ &= \int d\mathbf{X} \prod_{i=1}^{N} P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{x}_{i+1})P(\mathbf{x}_{i+1}|\mathbf{x}_i). \end{aligned}$$

Experimental data typically involve trajectories with two or three spatial dimensions. For diffusion in an isotropic medium with the assumption that particle uncertainties are given as normal distributions with no covariance among the spatial dimensions, the probability distribution of a particular displacement in each dimension is separable. Thus, if Υ is the number of dimensions, then

$$P(\mathbf{O}|D) = \prod_{n=1}^{\Upsilon} P(O_n|D). \quad (4)$$

Hence, it is sufficient to only consider the estimation problem in the one-dimensional (1D) case $O = \{o_i\}_{i=1}^{N}$, which is represented as

$$P(O|D) = \int_{\mathbb{R}^{N+1}} dX \prod_{i=1}^{N} P(o_i|x_i,x_{i+1})P(x_{i+1}|x_i). \quad (5)$$

A. Accounting for the effects of exposure time integration

Equation (5) is the fundamental description of the likelihood of diffusion constant D given observations O. Unfortunately, solving for this expression explicitly is difficult because Eq. (5) has every o_i term defined as conditionally dependent on both x_i and x_{i+1}. The conditional dependence arises because the estimate of o_i's position is typically made from data collected over an exposure time $0 < t_\epsilon \leqslant (t_{i+1} - t_i)$ between time points t_i and t_{i+1}, which are associated with x_i and x_{i+1}, respectively. If the observational apparatus is a camera sensor, the signal will be integrated over the frame, resulting in a motion-blurred image of the moving particle. If X is known, then the probability density of an o_i is conditional upon the particle's true position before the photon acquisition at the beginning of the frame (x_i) and the particle's true position after the photon acquisition at the beginning of the subsequent frame (x_{i+1}) (Appendix B). The probability density calculated in Appendix B is useful for performing motion-blurred diffusion simulations without trajectory subsampling.

In the case in which the exposure time $t_\epsilon \ll \delta t_i$, the motion-blur effect is no longer present, so the observed location o_i depends only on position x_i as in Eq. (2). The corresponding 1D representation is

$$P(O|D) = \int_{\mathbb{R}^N} dX \prod_{i=1}^{N} P(o_i|x_i) \prod_{j=1}^{N-1} P(x_{j+1}|x_j). \quad (6)$$

Without the additional dependence on x_{i+1}, the methods required to solve the integral in Eq. (6) are simpler. To use this simpler representation for any nonzero t_ϵ, we will transform Eq. (5) into a form that resembles Eq. (6), and we seek functions $\mathcal{M}(o_i,x_i)$ and $\mathcal{T}(x_{j+1},x_j)$ such that

$$\begin{aligned} P(O|D) &= \int_{\mathbb{R}^{N+1}} dX \prod_{i=1}^{N} P(o_i|x_i,x_{i+1})P(x_{i+1}|x_i) \\ &= \int_{\mathbb{R}^N} dX \prod_{i=1}^{N} \mathcal{M}(o_i,x_i) \prod_{j=1}^{N-1} \mathcal{T}(x_{j+1},x_j). \quad (7) \end{aligned}$$

The function $\mathcal{T}(x_{i+1},x_i)$ stands for the transition function; it is simply $P(x_{i+1}|x_i)$, the probability of a particle diffusing

with constant D moving from x_i to x_{i+1}, over time δt_i. The function $\mathcal{M}(o_i,x_i)$ stands for the measurement function and it encapsulates the net effect of both the measurement localization error and the motion-blur. The details of the representation equivalence of Eq. (7) are important for correctness, but they also unnecessarily complicate the exposition, and so they can be found in Appendix C. Prior work [15,16,19] investigated the motion-blur effects of exposure time integration, and it was found that the effect can be represented as an effective decrease in variance of the measurement localization error, dependent on diffusion constant D and exposure time t_ϵ. Our derivation agrees with the effective correction factors investigated and derived in Refs. [15,16,19], but it also provides a form for the diffusion likelihood that is directly amenable to the solution techniques we employ in Secs. III A–III C.

The result of the transformation of Eq. (7) is that the effective measurement function $\mathcal{M}_i$ and the transition function $\mathcal{T}_i$ take the form of normalized Gaussians. We use the notation

$$\mathcal{N}(a,a_0,\eta) = \frac{1}{\sqrt{2\pi\eta}} \exp\left[-\frac{(a-a_0)^2}{2\eta}\right]$$

to represent the normalized Gaussian function with variance $\eta = \sigma^2$ centered around mean a_0 considered as a function of a, a_0, and η. Using this notation, we can succinctly represent the measurement and transition functions as

$$\mathcal{T}_i = \mathcal{T}_i(x_{i+1},x_i) = \mathcal{N}(x_{i+1},x_i,\omega_i(D)) \quad \text{for } 1 \leqslant i \leqslant N-1, \tag{8}$$

$$\mathcal{M}_i = \mathcal{M}_i(o_i,x_i) = \mathcal{N}(o_i,x_i,\varepsilon_i(D)) \quad \text{for } 1 \leqslant i \leqslant N. \tag{9}$$

The transition functions [Eq. (8)] are unaffected by the motion-blur transformation, and their Gaussian representation follows directly from the normally distributed displacements of diffusive processes, hence the variance is $\omega_i(D) = 2D\delta t_i$.

For the measurement functions [Eq. (9)], the variance $\varepsilon_i(D)$ is the variance due to measurement error, v_i, combined with a correction for the effect of motion-blur that is dependent on the diffusion constant D and exposure time t_ϵ, giving $\varepsilon_i(D) = v_i - Dt_\varepsilon/3$. The factor of $1/3$ comes from the limit of continuous integration of photon emissions over averaged Brownian trajectories (Appendix C). It is important to note that the independence of t_ε and δt_i allows for gaps in the trajectories, since δt_i could span a duration of multiple frames but t_ε is the fixed exposure time of a single frame.

The result is that Eq. (7) allows us to express the likelihood function exactly in a simple form that deals directly with variable localization error, motion-blur effects, and missing or irregularly spaced trajectory localizations,

$$P(O|D) = \int_{\mathbb{R}^N} dX \prod_{i=1}^{N} \mathcal{M}_i \prod_{j=1}^{N-1} \mathcal{T}_j. \tag{10}$$

III. LIKELIHOOD DERIVATIONS

A. Recursive method

The notation for the transition and measurement functions allows us to define the likelihood function $\mathcal{L}(D)$ by writing Eq. (10) in a form that emphasizes the dependencies on each marginalized position x_i,

$$\begin{aligned}\mathcal{L}(D) &= P(O|D)\\ &= \int dx_N\, \mathcal{M}_N \int dx_{N-1}\, \mathcal{M}_{N-1}\mathcal{T}_{N-1} \cdots\\ &\quad \int dx_2\, \mathcal{M}_2\mathcal{T}_2 \int dx_1\, \mathcal{M}_1\mathcal{T}_1.\end{aligned} \tag{11}$$

The form of Eq. (11) leads to a direct recursive solution, taking into account the properties of integrals over products of normalized Gaussian functions. Define $\mathcal{L}_i$ as the subintegrand of $\mathcal{L}(D)$ considering only the first i observations,

$$\begin{aligned}\mathcal{L}_1(D,x_2) &= \int \mathcal{M}_1\mathcal{T}_1 dx_1\,,\\ \mathcal{L}_i(D,x_{i+1}) &= \int \mathcal{M}_i\mathcal{T}_i\mathcal{L}_{i-1} dx_i\,, \quad 2 \leqslant i \leqslant N-1,\\ \mathcal{L}(D) = \mathcal{L}_N(D) &= \int \mathcal{M}_N\mathcal{L}_{N-1} dx_N\,.\end{aligned}$$

Solving the integrals $\mathcal{L}_i$ sequentially, we find the resulting integrand is always in the form of a Gaussian with recursive variables dependent on the previous integrand. The recursive variables can be found by manipulating the quadratic components of the Gaussian functions (Appendix A) so that the initial variables are defined as

$$\mu_1 = o_1, \quad \eta_1 = \varepsilon_1 + \omega_1, \quad \text{and} \quad \alpha_1 = \eta_1 + \varepsilon_2, \tag{12}$$

and for $2 \leqslant i \leqslant N-1$, the recursive variables are

$$\begin{aligned}&\mu_i = \frac{\mu_{i-1}\varepsilon_i + \eta_{i-1}o_i}{\alpha_{i-1}}, \quad \eta_i = \frac{\eta_{i-1}\varepsilon_i}{\alpha_{i-1}} + \omega_i, \quad \text{and}\\ &\alpha_i = \eta_i + \varepsilon_{i+1}.\end{aligned} \tag{13}$$

Finally, this allows us to express our integrands $\mathcal{L}_i$ as

$$\begin{aligned}\mathcal{L}_1 &= \mathcal{N}(x_2,\mu_1,\eta_1),\\ \mathcal{L}_i &= \mathcal{N}(x_{i+1},\mu_i,\eta_i)\prod_{k=1}^{i-1}\mathcal{N}(o_{k+1},\mu_k,\alpha_k), \quad 2 \leqslant i \leqslant N-1,\\ \mathcal{L}(D) = \mathcal{L}_N &= \prod_{k=1}^{N-1}\mathcal{N}(o_{k+1},\mu_k,\alpha_k).\end{aligned} \tag{14}$$

Equation (14) is the final form of the recursive solution for $\mathcal{L}(D)$, which is simply the product of $N-1$ normalized Gaussians each of which has parameters that come from a recursive relationship on o_i, ε_i, and ω_i.

B. Laplace method

The Laplace method allows for an independent solution for Eq. (10), based on integrating the second moment of the Taylor expansion of the exponential component of a function. Given that the second moment of a Taylor expansion is

quadratic, the resulting function under the integral is always a Gaussian function. Another caveat to the Laplace method is that the Taylor expansion has to occur about the peak of the exponential, so that the first moment of the Taylor expansion goes to 0. To perform the Laplace method, we express our likelihood $\mathcal{L}(D) = P(O|D)$ in terms of exponential and nonexponential components,

$$\mathcal{L}(D) = \int dX\, f(X) = \int dX\, h(X) \exp[-g(X)],$$

where $f(x)$ is simply the integrand of Eq. (10),

$$f(X) = h(X)\exp[-g(X)] = \prod_{i=1}^{N} \mathcal{M}_i \prod_{j=1}^{N-1} \mathcal{T}_j. \tag{15}$$

Thus, using Eqs. (9) and (8), we see that $h = h(X)$ is independent of X and $g(X)$ is quadratic in X,

$$h = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\varepsilon_i}} \prod_{i=1}^{N-1} \frac{1}{\sqrt{2\pi\omega_i}},$$

$$g(X) = \sum_{i=1}^{N} \frac{(o_i - x_i)^2}{2\varepsilon_i} + \sum_{i=1}^{N-1} \frac{(x_{i+1} - x_i)^2}{2\omega_i}.$$

The maximum-likelihood estimate $\widehat{X}$ of the actual positions X, given D and O, will be wherever the integrand is maximized, and since $g(X) \geqslant 0$, we have $\widehat{X} = \mathrm{argmax}_X f(X) = \mathrm{argmin}_X g(X)$. Now, given that $g(X)$ is quadratic, a second-order Taylor expansion of g about $\widehat{X}$ is exact and the Laplace method will provide an exact solution for $\mathcal{L}(D)$ as the integral can be shown to take the form of a standard Gaussian integral. Furthermore, since h is independent of X, we know that $-\nabla\nabla \ln f(X) = \nabla\nabla g(X) = M$, where M can be thought of as the inverse of the covariance matrix for the multivariate Gaussian, or equivalently as the Hessian matrix of $-\ln f(X)$. Substituting M for $\nabla\nabla g(X)$ leaves us with a Gaussian integral with the solution

$$\mathcal{L}(D) = f(\widehat{X}) \int dX\, \exp\left[-\frac{1}{2}(X - \widehat{X})^{\intercal} M (X - \widehat{X})\right]$$

$$= f(\widehat{X}) \sqrt{\frac{(2\pi)^N}{\det M}}. \tag{16}$$

We also require $\widehat{X}$ to compute Eq. (16), which can be solved for with the relation $-\nabla \ln f(\widehat{X}) = 0$ (Appendix D), giving

$$\widehat{X} = M^{-1}\Theta, \tag{17}$$

where $\Theta := \{\theta_i = o_i/\varepsilon_i\}_{i=1}^{N}$.

C. Markov method

In this subsection, we present a derivation of the likelihood function $\mathcal{L}(D)$ utilizing a technique that was named and generalized by Chandrasekhar [22]. Markov's method allows us to transform $P(O|D)$ [Eq. (10)] from a function of the N observed positions $O = \{o_i\}_{i=1}^{N}$ into a function of the $N-1$ discrete steps (displacements) between subsequent observations

$$S = \{s_i = o_{i+1} - o_i\}_{i=1}^{N-1}.$$

This is possible because the spatial invariance of diffusion means $\mathcal{L}(D)$ depends not on the absolute spatial positions O, but only their relative displacements, S. Thus we should expect that $P(O|D)$ can also be expressed as $P(S|D)$. However, the definition of $P(O|D)$ in Eq. (11) cannot be directly transformed into a function on S. This is where Markov's method allows us to solve for a function $P(S|D) = P(O|D)$ for a given D value. For a particular fixed S and dS of interest, the value of $P(S|D)dS$ gives the probability that variable S' is within the bounds

$$S - \tfrac{1}{2}dS \leqslant S' \leqslant S + \tfrac{1}{2}dS\,. \tag{18}$$

More formally, $P(S|D)dS$ is the integral over the hypervolume dS around the point of interest S. Letting S' represent the variable integrated over,

$$P(S|D)dS = \int_{S-\frac{1}{2}dS}^{S+\frac{1}{2}dS} dS'\, P(O|D).$$

The issue remains that $P(O|D)$ is expressed in a basis of O rather than of S, and integrating with respect to bounds in a different basis is nontrivial. To circumvent the integration issue, Markov utilized a product of Dirichlet integrals [22] to expand the limits of integration to all space. The resulting expression is then factored and the limit of an infinitesimal dS is taken to produce the following expression:

$$P(S|D)dS = \frac{dS}{(2\pi)^{N-1}} \int d\boldsymbol{\rho}\, \exp[-\imath\, \boldsymbol{\rho}^{\intercal} S] \Lambda(\boldsymbol{\rho}), \tag{19}$$

where we can interpret $\Lambda(\boldsymbol{\rho})$ as the characteristic function of $P(O|D)$ in the S' basis, and it has the form

$$\Lambda(\boldsymbol{\rho}) = \int dS'\, \exp[\imath\, \boldsymbol{\rho}^{\intercal} S'] P(O|D) = \exp\left[-\frac{1}{2}\boldsymbol{\rho}^{\intercal}\Sigma\boldsymbol{\rho}\right], \tag{20}$$

where the covariance matrix Σ is symmetric tridiagonal, with nonzero elements

$$\Sigma_{i,i} = \omega_i + \varepsilon_i + \varepsilon_{i+1},$$

$$\Sigma_{i,i+1} = \Sigma_{i+1,i} = -\varepsilon_{i+1}. \tag{21}$$

The expression in Eq. (20) is well known as the characteristic function of a multivariate Gaussian. Substituting Eq. (20) into Eq. (19), integrating the expression, and then factoring out dS gives

$$\mathcal{L}(D) = P(S|D) = \frac{1}{\sqrt{(2\pi)^{N-1}\det\Sigma}} \exp\left[-\frac{1}{2}S^{\intercal}\Sigma^{-1}S\right]. \tag{22}$$

IV. NUMERICAL LIKELIHOOD CALCULATION

We have presented three independent solutions to the experimental diffusion likelihood $\mathcal{L}(D)$ [Eq. (11)]: the recursive method (Sec. III A), the Laplace method (Sec. III B), and the Markov method (Sec. III C). While each method requires separate consideration, several features are common to all of the implementations. The separability of the problem allows us to estimate diffusion constants for any dimensional inputs using the 1D algorithms [Eq. (4)]. The inputs to the

algorithms are as follows: (i) the observed particle locations, $\mathbf{O} = \{\mathbf{o}_i\}_{i=1}^N$; (ii) the observation times, $T = \{t_i\}_{i=1}^N$; (iii) the measurement standard error for each observation, $\sqrt{\mathbf{V}} = \{\sqrt{\mathbf{v}_i}\}_{i=1}^N$; (iv) the exposure of each frame, t_ϵ; and (v) one or more diffusion constants D at which to evaluate the likelihood. The output for each D value is $\ln[\mathcal{L}(D)]$. The logarithm of the likelihood makes the computation of products and exponentials much faster, and it avoids the problem of numerical underflow for very small values of $\mathcal{L}(D)$. Additionally, because the logarithm is a strictly monotonically increasing function, $\text{argmax}_D \mathcal{L}(D) = \text{argmax}_D \ln[\mathcal{L}(D)]$, the maximum-likelihood estimate is identical for the log-likelihood.

A. Recursive method

The recursive algorithm follows directly from the recursively defined variables [Eqs. (12) and (13)] and the expression of $\mathcal{L}(D)$ as a product of Gaussians [Eq. (14)]. The recursive expressions for α_i, η_i, and μ_i are causal [the i terms depend only on the $(i-1)$ terms], enabling their computation in a simple for-loop over N. Noting that the logarithm of a normalized Gaussian is

$$\ln \mathcal{N}(a,b,v) = -\frac{1}{2}\left[\ln(2\pi) + \ln(v) + \frac{(a-b)^2}{v}\right], \quad (23)$$

we apply Eq. (23) directly to Eq. (14) to arrive at a computationally efficient form for the recursive solution of the log-likelihood,

$$\begin{aligned}\ln \mathcal{L}(D) &= \sum_{i=1}^{N-1} \ln \mathcal{N}(o_{i+1},\mu_i,\alpha_i) \\ &= -\frac{1}{2}\left[(N-1)\ln(2\pi) + \sum_{i=1}^{N-1}\ln(\alpha_i) + \sum_{i=1}^{N-1}\frac{(o_{i+1}-\mu_i)^2}{\alpha_i}\right].\end{aligned}$$

Of all the methods, the recursive method is the simplest to implement and the most computationally efficient and numerically stable.

B. Laplace method

The computational core of the Laplace method centers around the Hessian matrix M [Eq. (D3)]. This matrix is symmetric tridiagonal, which means all nonzero elements are on the main diagonal and the diagonals immediately above and below. Using M, we can solve the linear system $\widehat{X} = M^{-1}\Theta$ [Eq. (17)] to obtain the maximum-likelihood estimates $\widehat{X}$ for the true particle locations. Typically, solving large linear systems is expensive, but since M is tridiagonal, there are algorithms to solve this system in linear time [23].

Given a solution for $\widehat{X}$, we can use the definition of $f(X)$ in Eq. (15) along with Eq. (23) to compute

$$\ln f(\widehat{X}) = -\frac{1}{2}\left[\sum_{i=1}^{N}\ln(2\pi\varepsilon_i) + \sum_{i=1}^{N}\frac{(o_i-\widehat{x}_i)^2}{\varepsilon_i} + \sum_{i=1}^{N-1}\ln(2\pi\omega_i) + \sum_{i=1}^{N-1}\frac{(\widehat{x}_{i+1}-\widehat{x}_i)^2}{\omega_i}\right]. \quad (24)$$

Finally, we can compute the log-likelihood using the Laplace solution of Eq. (16), finding that

$$\begin{aligned}\ln \mathcal{L}(D) &= \ln f(\widehat{X}) + \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln\det M \\ &= -\frac{1}{2}\left[(N-1)\ln(2\pi) + \sum_{i=1}^{N-1}\ln(\omega_i) + \sum_{i=1}^{N-1}\frac{(\widehat{x}_{i+1}-\widehat{x}_i)^2}{\omega_i} + \sum_{i=1}^{N}\ln(\varepsilon_i) + \sum_{i=1}^{N}\frac{(o_i-\widehat{x}_i)^2}{\varepsilon_i} + \ln\det M\right]. \quad (25)\end{aligned}$$

The log-determinant of the tridiagonal matrix M can also be computed in linear time [23].

C. Markov method

Finally, the Markov method computation, like the Laplace method, is centered around matrix computations. In this case, the matrix of interest is the $(N-1)$-dimensional covariance matrix Σ [Eq. (21)], which also happens to be symmetric tridiagonal, so the same linear-time algorithms used in the Laplace method are applicable.

For the Markov method computation, we first solve the linear system $\Phi = \Sigma^{-1}S$, then we apply this solution along with the tridiagonal log-determinant algorithm to compute the logarithm of the likelihood expression from Eq. (22), giving

$$\ln \mathcal{L}(D) = -\tfrac{1}{2}[(N-1)\ln(2\pi) + S^\intercal\Phi + \ln\det\Sigma].$$

V. THE MLE AND THE OBSERVED INFORMATION

Typically, normal approximations are applied to likelihood distributions to generate simple confidence intervals about the MLE [24]. However, in the case of diffusion estimation, a log-normal approximation is preferred over the normal approximation of the likelihood distribution because the log transformation of D extends the acceptable interval from $[0,\infty)$ in D space to $(-\infty,\infty)$ in $\ln D$ space; this prevents the possibility of confidence intervals that allow negative values of D. The log-normal approximation is generally a more accurate parametrization of $\mathcal{L}(D)$ than the normal approximation. The log-normal approximation comes about from a second-order Taylor expansion about some fixed value $\ln\hat{D}$ in $\ln D$ space; the normal approximation comes from a second-order Taylor expansion about $\hat{D}$ in D space.

If $\ln\hat{D}$ is defined as the log of the MLE, then the first derivative of the Taylor expansion vanishes for all $\hat{D} > 0$ so that the second-order Taylor expansion can be described

by two parameters, $\hat{D}$ and $\tilde{\mathcal{K}}(\ln \hat{D}) = -\frac{\partial^2}{\partial \ln \hat{D}^2} \ln \tilde{\mathcal{L}}(\ln \hat{D})$. The parameter $\tilde{\mathcal{K}}(\ln \hat{D})$ is strictly positive and is known as the observed Fisher information [25] when $\hat{D}$ is the MLE. Here $\ln \tilde{\mathcal{K}}(\ln \hat{D})$ is simply referred to as the observed information and is dimensionless in $\ln D$ space. The likelihood is invariant to transformation of variables [24], but the observed information is not. For all software implementations [26], we calculate the observed information in $\ln D$ space as

$$\tilde{\mathcal{K}}(\ln \hat{D}) = \hat{D}^2 \mathcal{K}(\hat{D}),$$

where the solution to $\mathcal{K}(\hat{D})$, the observed information in D space, is derived in Appendix E.

The log-normal approximation is described by the log-likelihood ratio statistic (LLR) [27] as

$$-2 \ln \left(\frac{\mathcal{L}(D)}{\mathcal{L}(\hat{D})} \right) \approx \tilde{\mathcal{K}}(\ln \hat{D})(\ln D - \ln \hat{D})^2. \tag{26}$$

When the log-normal approximation accurately describes the true likelihood, the LLR described by Eq. (26) is χ-square distributed; which implies that the square root of the statistic is normally distributed. For the normally distributed statistic, the probability that an interval from $-\infty$ up to some value A contains the $\ln D$ that generated the corresponding trajectory is

$$\begin{aligned} P(\ln D \leqslant A) &\approx \int_{-\infty}^{A} d(\ln D)\, \mathcal{N}(\ln D, \ln \hat{D}, \tilde{\mathcal{K}}(\ln \hat{D})^{-1}) \\ &= \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{A - \ln \hat{D}}{\sqrt{2\tilde{\mathcal{K}}(\ln \hat{D})^{-1}}} \right) \right]. \end{aligned}$$

Since the log-normal approximation is symmetric in $\ln D$ space, the corresponding $C(100\%)$ confidence interval in D space defined by end points $B\pm$ is

$$B_{\pm} = \exp\left[\ln \hat{D} \pm \sqrt{2\tilde{\mathcal{K}}(\ln \hat{D})^{-1}} \operatorname{erf}^{-1}(C) \right]. \tag{27}$$

VI. ANALYSIS METHODS

In the following sections, we establish various metrics for quantifying the performance of the MLE and the associated confidence intervals. Section VI A describes the importance of the expectation of the observed information, which can be used to quantify the average estimation quality of all trajectories that share a set of fixed parameters. Section VI B is a description of the diffusion simulation we developed for testing our likelihood distribution against the likelihood distribution proposed in Ref. [20]. The implementation of the MLE analysis given our likelihood distribution and the distribution in Ref. [20] is described in Sec. VI C, and conditions for recognizing estimator failure are described in Sec. VI D.

A. The expected information

The expected Fisher information [28], which we abbreviate as the *information*, is a measure of our ability to estimate $\hat{D}$ for all O realizations of a trajectory when all other parameters are held fixed. Here, the information is defined as

$$\tilde{\mathcal{I}}(\ln \hat{D}) = \langle \tilde{\mathcal{K}}(\ln \hat{D}) \rangle = \int dS\, \tilde{\mathcal{K}}(\ln \hat{D}) P(S|\hat{D}),$$

where the measurement variances V are fixed, and S are the step displacements. Trajectories with different localization variances will have different information values. A solution to $\mathcal{I}(\hat{D})$ from the recursive method is described in Appendix F.

The information is plotted over various trajectory conditions as a function of $\hat{D}$ with fixed parameters V, T, and t_ε. The information in $\ln D$ space is bounded by a maximum value of $(N-1)/2$ for any trajectory of length N; the maximum information is obtained when measurement errors are negligible. When the information is maximal, the MLE analysis will work for all trajectories that share the same fixed parameters; conversely, an information of 0 implies that the corresponding likelihood functions have no unique maximum, and hence the MLE estimator will fail. The information, as a function of $\hat{D}$ when all other trajectory parameters are fixed, takes on a sigmoid shape where finite information values correspond to $\hat{D}$ values that can be returned by an MLE for a given set of measurement parameters. Furthermore, information plots are useful in predicting when the additional precision from sophisticated likelihood distributions are significant or negligible. In this paper, we find values of $\hat{D}$ where the information is asymptotically equal to $(N-1)/2$ so that the effects of correlated measurement errors have a negligible effect on the shape of the likelihood distribution, highlighting regions where approximated likelihood distributions perform asymptotically as well.

B. Gaussian lattice background diffusion simulation

To demonstrate the robustness of the MLE estimate and the corresponding log-normal parametrization, we developed a diffusion simulator [26] that simulates all of the measurement errors that accompany a camera-recorded fluorophore without relying on the probability densities from our derivations in Appendix B. To do this, individual photon emission times are drawn from exponential wait times with a predefined emission rate of 200 photons per full frame exposure so that the resulting images of the fluorophore have Poisson-distributed photon counts. The particle is allowed to diffuse freely so that an emitted photon position is randomly chosen from a normally distributed deviation from the actual position of the particle at the emission time. The standard deviation of the photons from the true particle position was set at 100 nm. Fluorescence intermittency is simulated via a continuous two-state process, where the particle is either in an emission state or a dark state. Photon emissions are only possible during the emission state; when the particle is in the dark state, there are no photon emissions. The transition between emission and dark states occurs independently of the camera frame time, so a particle could go dark midframe, which allows for the possibility of very dim localizations. The transition rate was 0.2 per frame to transition from the dark state to the emission state and 0.05 per frame to transition from the emission state to the dark state.

An important feature we incorporated into our simulation was a background characterized by equally spaced out-of-focus emitters. This simulates the common problem of

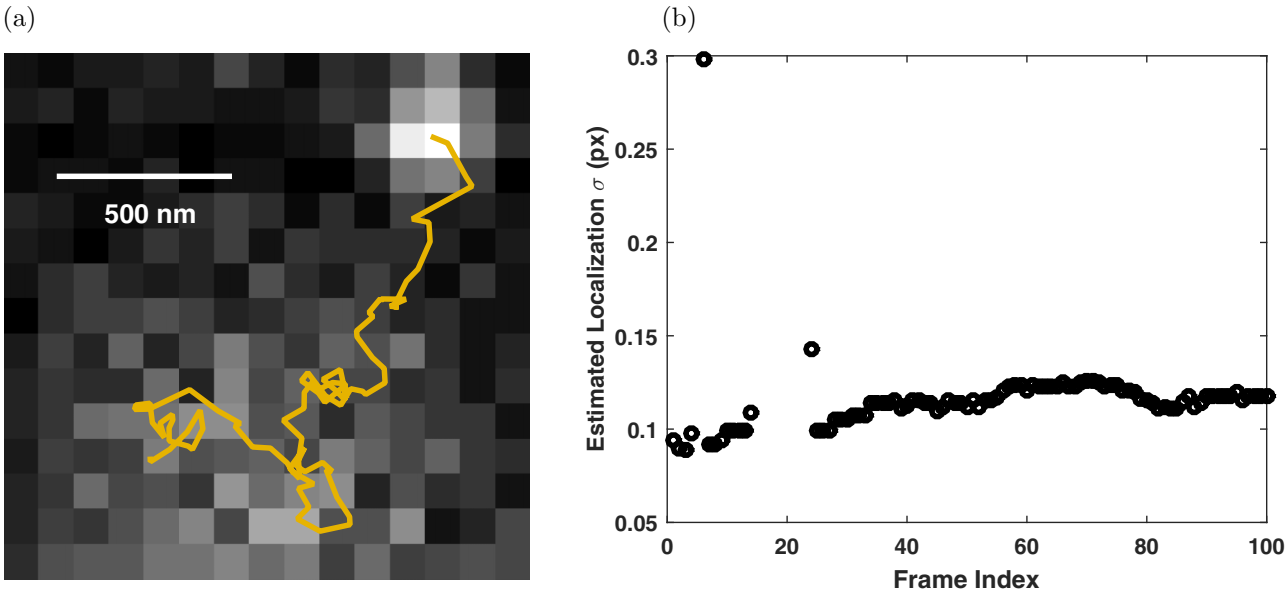

FIG. 1. A sample trajectory created with our diffusion simulations. Part (a) is a sample trajectory (solid line) overlaid on a single realization of the out-of-focus fluorescence. A noise realization of the moving particle can be seen in the upper right corner of the trajectory. The particle in this instance is moving at 10^{-1} μm^2 s^{-1} at an acquisition rate of 100 Hz so that this trajectory realization spans a greater region of the background. Part (b) shows the localization errors of the localized particle at each frame. Since there is fluorescence intermittency, some frames do not have localizations and therefore an associated localization error is not represented. The frames with significantly higher uncertainties occur when the particle transitions between the bright and dark states during a frame acquisition.

observing in-focus molecules in the presence of out-of-focus background fluorescence. The out-of-focus emitters have an effective σ of 500 nm with a mean emission rate of 2000 photons per frame. The lattice was spaced so that each background object was 2 μm apart. Our intention for choosing this type of background pattern is to guarantee that a diffusing particle will randomly traverse regions of high and low background noise, ensuring that the associated fit precision will vary naturally over a trajectory (see Fig. 1).

All trajectory trials were simulated over an exposure period of 100 camera frames at a frame rate of 100 Hz to simulate the shorter trajectories that would benefit the most from our ability to use all available localization error information. The pixel size of the image is 100 nm and the diffusion coefficient is simulated from 10^{-4} to 10^{-1} μm^2 s^{-1}. Every trajectory began at a random position to ensure that a particular set of background values was not favored.

The variance of the position estimates for diffusion analysis is calculated from the theoretical Cramer-Rao lower bound (CRLB), the minimal variance of an unbiased estimator of the localization parameters. The inputs for the CRLB calculation were as follows: particle positions calculated from the averaged position of the photon locations of a given frame, the expected model background value of a 7×7 pixel region surrounding the particle position, and the expected photon emissions, which in this instance is 200 photons times the fraction of the frame time spent in the emission state. The CRLB calculations are based on the supplemental text and software from [11]. The localized positions are then calculated by taking the average position of the photons and perturbing that position by a random displacement drawn from a normal distribution with a variance equal to the localization CRLB. With this approach, we can assume that the hypothetical localization method is theoretically optimal without focusing on specific localization software.

The Gaussian lattice simulation was run with 20 log-spaced D parameters from 10^{-4} to 10^{-1} μm^2 s^{-1} with 10^4 trajectories sampled for each D. The trajectories were analyzed individually to produce corresponding MLE and observed information values. The MLEs with positive observed information values were then used to determine 68% and 95% confidence intervals from the log-normal parametrization [Eq. (27)].

C. MLE analysis approaches

The MLE, observed information, and confidence intervals were calculated from three types of analysis approaches. The first approach is the likelihood distribution as described in Sec. III with variable localization errors and trajectory intermittencies, and it will be referred to as variable error analysis (VEA). The second approach follows the analysis in Ref. [20], and it will be referred to as the mean error analysis (MEA). In our computation of the respective likelihood distributions, both the MEA and VEA methods are performed using a MATLAB implementation of the recursive method (Sec. IV A), except that the MEA version only uses the effective mean localization error $\sqrt{\langle V \rangle}$. In our simulations, v_i are drawn from a path-dependent distribution, so we defined a different $\langle v_i \rangle = \sum_{i=1}^{N} v_i / N$ for each trajectory. The existence of occasional dim localizations in the trajectories can result in very poor values for the reduced error, leading to significant errors in the MEA. An intrepid researcher, limited to an

MEA approach, would likely filter out localizations that were exceptionally bad; therefore, we implemented a third analysis approach for comparison, the filtered mean error analysis (FMEA). For the FMEA, any localization error that was greater than half a pixel squared was omitted so that the resulting trajectory has the reasonable but *ad hoc* restriction $v_i < 0.5\ \mathrm{px}^2$.

For the corresponding analysis, the value of $\langle V \rangle$ used for all figures was determined from the FMEA approach to be approximately $1.37 \times 10^{-4}\ \mu\mathrm{m}^2$ so that the corresponding localization error is 12 nm. Furthermore, the reference time unit in the figures (τ) was set to the length of a single frame 0.01 s.

D. Critical failures

MLE analysis is not applicable for all trajectories that have been recorded from SPT experiments. If the observed information at $\hat{D}$ is not positive, then it means either the analysis software [26] has not found the correct MLE, or the likelihood function has a maximum at $D = 0$, or it is flat to numerical precision in the neighborhood of the maximum. In any of these instances, the log-normal approximation is no longer an accurate parametrization of the corresponding likelihood distribution. We deal with these trajectories by classifying them as "critical failures" and they are omitted from subsequent MLE analysis [Fig. 5(b)].

The MLE analysis software [26] runs a bounded optimization to find $\hat{D}$ in the interval $(10^{-8}, 10^{8})$, which ensures that $\hat{D} > 0$ in our analysis. Once $\hat{D}$ is returned, the corresponding observed information is calculated. If $\hat{D}$ does not correspond to a unique MLE or the actual MLE is defined at $D < 10^{-8}$, then the observed information at the estimator's $\hat{D}$ will not be positive. The trajectories that are associated with a $\tilde{\mathcal{K}}(\ln \hat{D}) \leqslant 0$ are defined as critical failures and are not considered because the returned $\hat{D}$ values for these trajectories are guaranteed to make the results of any hypothesis testing worse. The critical failure rate, the fraction of trajectories unsuitable for MLE analysis, is a metric for performing MLE analysis reliability for a given set of trajectory conditions.

VII. RESULTS

A. Noise-free trajectory parametrization

When there is no measurement error, $\mathcal{K}(\hat{D}) = \mathcal{I}(\hat{D})$, which means that there is an asymptotic limit where general statements can be made regarding both the quality of the approximate confidence intervals and the effect of trajectory intermittencies on the estimator quality. In the absence of measurement error, the likelihood expression reduces to

$$\mathcal{L}_0(D) = \prod_{i=1}^{N-1} \mathcal{N}(o_{i+1}, o_i, \omega_i) \quad \forall\ \varepsilon_i = 0,$$

where the subscript 0 in $\mathcal{L}_0(D)$ denotes the distinction that all ε_i in the data are negligible. In this limit, the Taylor expansion of $\mathcal{L}_0(D)$ in $\ln D$ space can be refactored into a simple expression for the exact LLR [the approximate LLR is defined in Eq. (26)],

$$-2 \ln\left(\frac{\mathcal{L}_0(D)}{\mathcal{L}_0(\hat{D})}\right) = 2\tilde{\mathcal{I}}_0(\ln \hat{D})\left[\frac{\hat{D}}{D} - \ln\left(\frac{\hat{D}}{D}\right) - 1\right]. \quad (28)$$

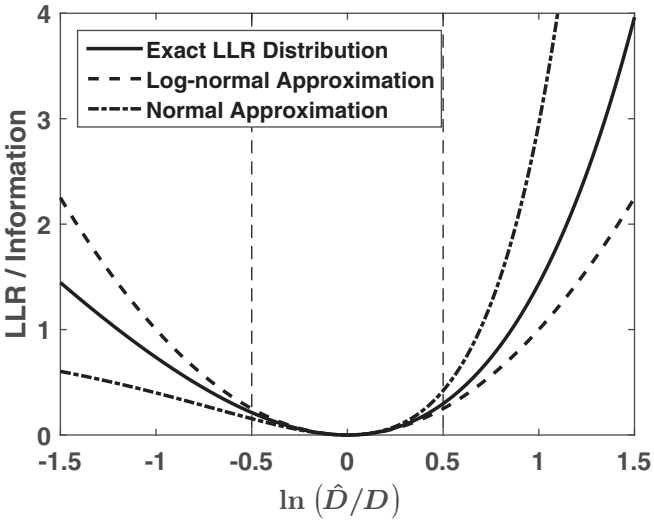

FIG. 2. The exact LLR as described by Eq. (28) (solid line), the log-normal approximation described by Eq. (26) (dashed line), and the corresponding normal approximation (dash-dotted line) in the ideal, no measurement error limit. When $|\ln(\hat{D}/D)| < 0.5$, the log-normal approximation is a good fit to the actual log-likelihood and is within 3% of the true value. The log-normal approximation is always closer to the LLR than the normal approximation.

The quantity in the brackets in Eq. (28) is known as the Stein loss [29] and perfectly describes the shape of $\mathcal{L}_0(D)$. When there is no measurement error, the expected information in $\ln \hat{D}$ space, $\tilde{\mathcal{I}}_0(\ln \hat{D})$, describes the width of the distribution for both the log-normal approximation in Eq. (26) and the exact LLR in Eq. (28). Also, $\tilde{\mathcal{I}}_0(\ln \hat{D}) = (N-1)/2$ regardless of δt_i. This implies that the only parameter that affects the estimate quality in the absence of measurement error is the number of observations, N. The information value of noiseless trajectories, $(N-1)/2$, is an upper-bound on the possible observed information value for noisy trajectories of length N. Consequently, the independence of the information from δt_i implies that the effect of trajectory intermittencies on the quality of an estimator must be intrinsically linked to measurement error. In other words, information is a function of the number of particle observations that can be reduced by measurement error at each observation. Intermittencies reduce the influence of measurement error because the larger temporal spacing between observations from intermittency results in a larger expected squared displacement of the diffusing particle.

The distributions from Eq. (26), Eq. (28), and a normal approximation [described by precision $\mathcal{I}(\hat{D})$ and mean $\hat{D}$] are divided by $\tilde{\mathcal{I}}(\ln \hat{D})$ and plotted in Fig. 2 to determine a good theoretical bound for the approximate confidence intervals. When the log-normal-based confidence intervals have a range $|\ln(\hat{D}/D)| < 0.5$, then the log-normal approximation is in very good agreement with the exact likelihood. The deviance between the exact LLR and the log-normal approximation at both end points grows substantially when $|\ln(\hat{D}/D)| > 1$, so that ranges outside of ± 1 in dimensionless $\ln(\hat{D}/D)$ space are not expected to satisfy confidence interval tests. The normal approximation deviates from the LLR faster than the log-normal approximation, so its usefulness is limited to narrower confidence intervals.

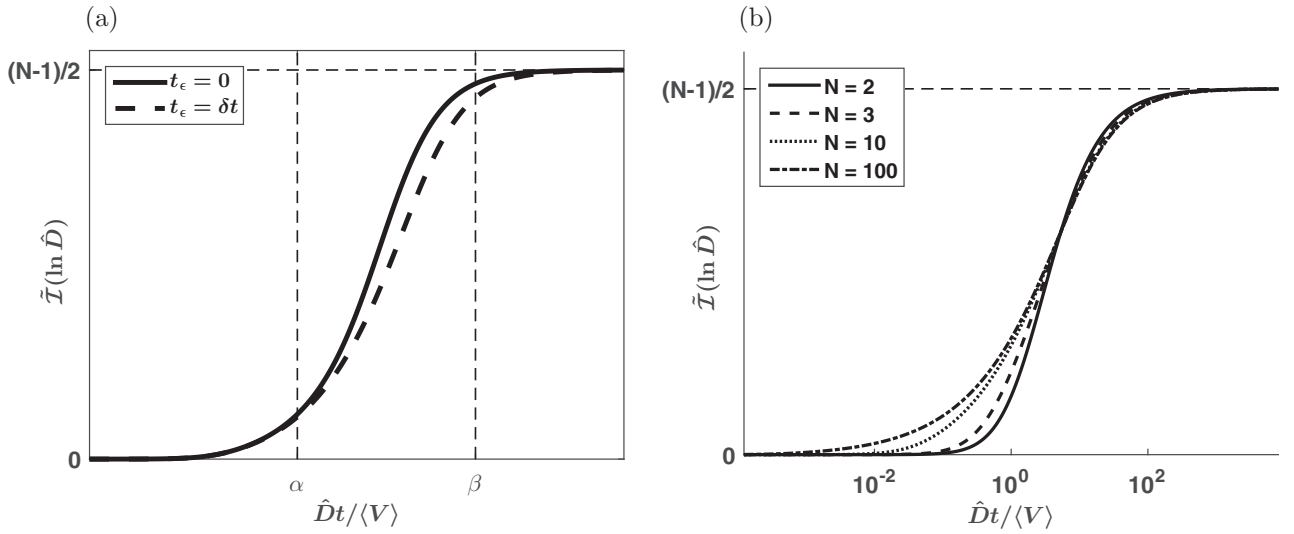

FIG. 3. Plots of the information at various ratios of $\hat{D}t/\langle V\rangle$. Part (a) is a qualitative representation of the information without dynamic error (solid line) and with dynamic error from a full frame exposure time (dashed line). Part (a) has three intervals delimited by the approximate values α and β: when $\hat{D}t/\langle V\rangle > \beta$, the information is in good agreement with the no error scenario; when $\alpha < \hat{D}t/\langle V\rangle < \beta$, the associated observed information values have the highest variance; when $\hat{D}t/\langle V\rangle < \alpha$, the information is so poor that a different form of analysis should be considered for meaningful interpretations. Part (b) is the information curve with various trajectory lengths with data acquired under a full frame exposure time. The approximate value of α would shift further to the left for longer trajectories. The approximate value of β appears to be constant since all curves asymptote at the same value, so the negligible error limit is dictated only by the magnitude of the measurement error. The crossover in information quality per displacement in (b) when $\langle V\rangle = \hat{D}t_\epsilon/6$ is due to the effects of correlated observations with full frame exposure times.

B. The expected information for noisy trajectories

To better understand the results from our simulations (Sec. VI B), it is necessary to understand the behavior of the information in the presence of measurement error. We plot the information, $\tilde{\mathcal{I}}(\ln \hat{D})$, to understand how the log-normal parametrizations will typically behave given an arbitrary O. Although our likelihood derivations can incorporate nonscalar V and trajectory intermittency, we will first restrict ourselves to the constant error, $\langle V\rangle$, and nonintermittent observation cases.

In the constant error case, the information scales with $\langle V\rangle/(\hat{D}t)$ proportional to an inverse quadratic function [17] as observed in Fig. 3. The information curves asymptote to values of $(N-1)/2$ when $\hat{D}t/\langle V\rangle \to \infty$, and to 0 when $\hat{D}t/\langle V\rangle \to 0$. Due to the inverse quadratic shape in Fig. 3(a), we see three distinct regimes where the information behaves differently on a local level. When $\hat{D}t/\langle V\rangle \gg 1$, the observed error is nearly equal to the expected information, so the Stein loss described in Eq. (28) is a better representation of the likelihood distribution than the log-normal approximation. When $\hat{D}t/\langle V\rangle \approx 1$, the observed information for individual trajectories fluctuates about the expected information with a larger variance. The log-normal approximation becomes a better parametrization than the Stein loss when the effect of measurement error becomes significant because the higher-order terms of the Taylor expansion on $\tilde{\mathcal{L}}(\ln D)$ are no longer represented by the information. When $\langle V\rangle$ becomes large, the observed information approaches zero, which implies that MLE analysis may not be possible for many trajectories with those parameter values.

Trajectories with variable localization errors will have information curves that deviate from the curve calculated from the $\langle V\rangle$ and in general do not follow an inverse quadratic function. A method that only characterizes a $\langle V\rangle$ where the full V is available is expected to deviate in quality from a method that considers V for variance estimates and confidence intervals. The expected deviation is demonstrated in Fig. 4, where

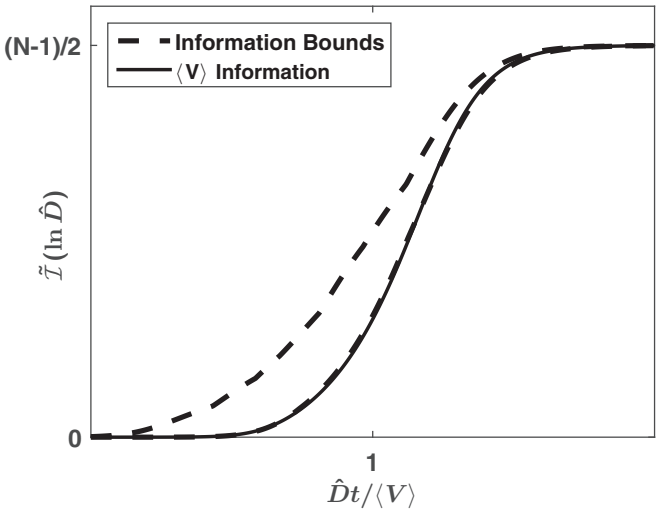

FIG. 4. A plot of the largest and smallest information values (dashed lines) calculated from 1000 sets of uniformly distributed localization errors for trajectories of length N and the corresponding information curve (solid line) when all localization errors in a trajectory are equal to $\sqrt{\langle V\rangle}$. Information values generated from a set of variable localization errors are larger than the information when all localization errors are represented by the scalar $\sqrt{\langle V\rangle}$.

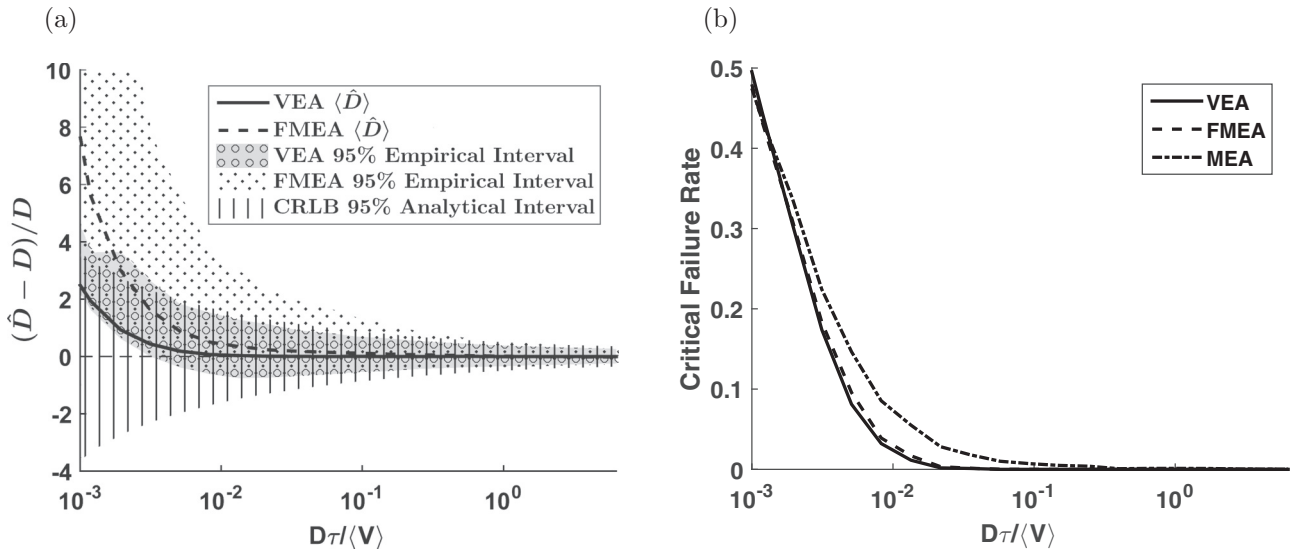

FIG. 5. Plots of the MLE distributions as well as the critical failure rate for the ensemble of trajectories corresponding to a particular D value. Part (a) compares the performance of the VEA and FMEA approaches relative to the expected performance of an unbiased, optimal estimator. The CRLB for the unbiased optimal estimator is calculated for nonintermittent trajectories with constant localization error and a length of 100 observations. The CRLB is then used as a fiducial 95% interval (vertical lines) calculated as a deviation of $1.96 \times \sqrt{\text{CRLB}}$ about a mean relative error of 0. The VEA approach is represented by its average deviation from the simulated D (solid line) and an interval that contains 95% of the returned $\hat{D}$ values (circles). The FMEA approach is represented by its average deviation from the simulated D (dashed line) and an interval that contains 95% of the returned $\hat{D}$ values (diamonds). The MEA approach was biased for all represented conditions and had a variance orders of magnitude larger than the VEA method, and it was omitted from this plot as a result. The VEA approach is superior to the CRLB fiducial until the bias manifests itself at $D\tau/\langle V \rangle < 10^{-2}$. Part (b) shows the critical failure rates of all three approaches. The critical failure rate is the fraction of trajectories that did not return a positive observed information from each sampled D value. The critical failure rate shows that when $D\tau/\langle V \rangle < 10^{-3}$, over 50% of the simulated trajectories were unsuitable for MLE analysis.

the maximum and minimum values of the Fisher information are plotted for trajectory ensembles with random localization errors drawn from a uniform distribution and categorized by the same average $\langle V \rangle$. For the ratios of $\hat{D}t/\langle V \rangle$, the solid curve returns an information value comparable to the lowest possible information realization given a random distribution of localization errors. This implies that in most cases, the returned diffusion estimate from a likelihood distribution that only recognizes a single $\langle V \rangle$ will have a lower precision than a likelihood distribution that accounts for the entire vector V associated with the trajectory. Furthermore, intermittencies are considered inseparable from variable localization errors because a likelihood distribution that describes a trajectory with constant error and intermittencies can be refactored to describe a trajectory with variable error and no intermittency.

C. Simulation results

In Fig. 5(a), the relative error between the mean $\hat{D}$ estimates and the simulated D parameters were plotted to verify the quality of the approaches. The MEA approach was completely omitted from Fig. 5(a) because its bias was orders of magnitude greater than the other methods for smaller but relevant $D\tau/\langle V \rangle$ values. A fiducial CRLB interval of relative errors for theoretically unbiased estimates from trajectories with length 100 and constant variance $\langle V \rangle$ was plotted in Fig. 5(a) for comparison. The CRLB interval in D space is somewhat misleading because it was calculated from the normal approximation, so it has intervals that extend into negative D space when $D\tau/\langle V \rangle < 10^{-1}$; this is an artifact from parametrizing the interval in a basis with boundaries. The FMEA interval quickly diverges from the CRLB interval, which implies that in this simulation, the ensemble of trajectories cannot be parametrized well by a single $\langle V \rangle$ when $D\tau/\langle V \rangle < 10^{-1}$. The VEA approach is more robust with a smaller variance because it incorporates more information from the trajectory and is practically an unbiased estimator until $D\tau/\langle V \rangle < 10^{-2}$, at which point a noticeable bias manifests. Figure 5(b) shows the critical failure rate of all three approaches, that is, the percentage of negative observed information quantities returned from the ensemble of trajectories after the MLE was estimated (Sec. VI D). The trajectories with a negative observed information were discarded because the returned $\hat{D}$ was not a true maximum of the likelihood; in this case, a meaningful analysis cannot be performed with only a second-order Taylor expansion on these particular trajectories, and in such instances a Bayesian approach may be preferable. The presence of dim localizations adversely impacts the MEA, so the critical failure rate is noticeably larger than the other methods when $D\tau/\langle V \rangle \approx 10^{-2}$. The FMEA and VEA approaches have statistically comparable failure rates, which implies that the loss of information from discarding data in the FMEA approach is preferable to the mischaracterization of poor localizations in the MEA approach.

From Eq. (27), confidence interval performances at 68% and 95% are shown in Fig. 6. The results show that the

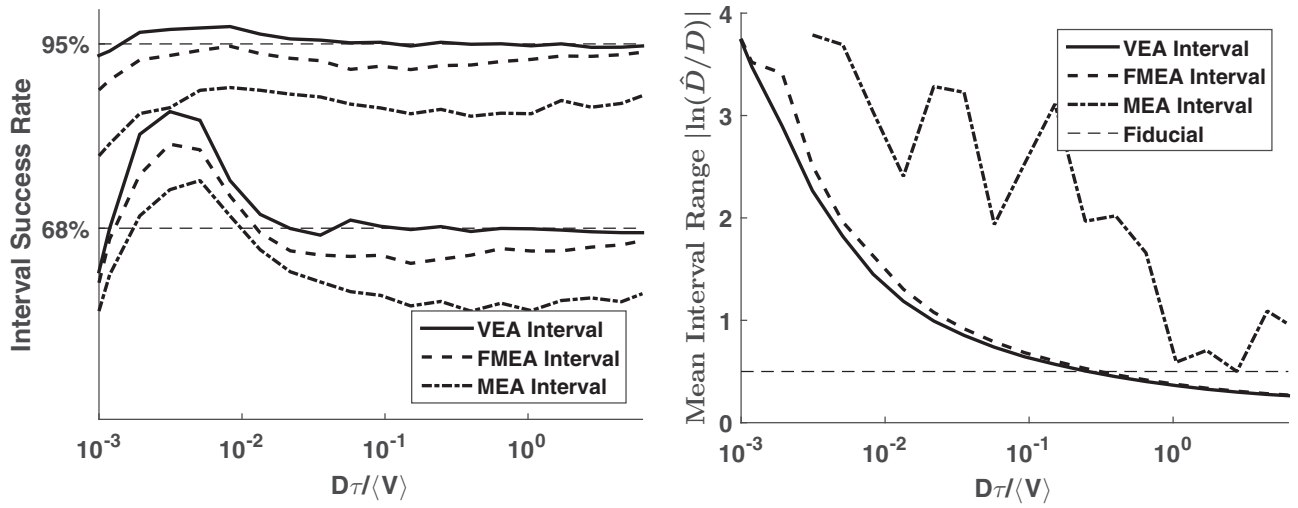

FIG. 6. Plots of confidence interval analysis for the VEA, MEA, and FMEA approaches. Confidence intervals were drawn for every trajectory, and an interval was considered successful if the true D parameter was inside the interval. Part (a) shows the success rates of both the 68% and 95% interval for all three methods. The VEA confidence interval is in good agreement with the corresponding success rate until $D\tau/\langle V\rangle < 10^{-2}$. In the region where the VEA interval is successful, the FMEA interval is on average 4% lower for both confidence values, but asymptotes to agreement as $D\tau/\langle V\rangle$ gets large. The MEA interval is on average 10% lower than the expected success rates and does not asymptote to the correct success rate. Part (b) shows the range of the 68% interval from $\hat{D}$ in $\ln D$ space. The MEA interval is adversely affected by dim localizations and did not approach a stable value with 10^4 trajectory samples for any of the simulated values of D. The mean interval ranges for both the VEA and FMEA approaches are similar, but the VEA approach was narrower. The range of the good agreement value observed in Fig. 2 is plotted as a fiducial range for the theoretical good agreement value of the log-normal approximation to the actual likelihood distribution.

confidence intervals are in good agreement with the results observed in Sec. VII A (Fig. 2) for interval ranges that are within the dimensionless limit of $|\ln(\hat{D}/D)| < 1$ [Fig. 6(b)]. The VEA method performs well at trajectory lengths just under 100 observations for values of $D\tau/\langle V\rangle > 10^{-2}$. The FMEA and MEA intervals appeared to generally underestimate the expected confidence level and should thus be treated with skepticism when making statements regarding the estimate if variable localization errors are disregarded in diffusion analysis. Interestingly enough, Fig. 6(b) shows that the ranges of the confidence intervals of the VEA and FMEA approaches are within good agreement with their corresponding likelihood distributions, as seen in Fig. 2 when $D\tau/\langle V\rangle > 10^{-1}$. When $D\tau/\langle V\rangle < 10^{-3}$ for trajectories with $N < 100$, the mean dimensionless interval ranges for all approaches are greater than 2, which corresponds to the poor performance of the confidence intervals seen in Fig. 6(a) as well as the asymmetric mismatch between the log-normal approximation and the true log likelihood, as seen in Fig. 2.

VIII. DISCUSSION AND CONCLUSION

We have developed a method for reliable diffusion estimates for single-particle trajectory analysis that can fully incorporate the realistic effects of noisy data and variable localization errors. The most important contributions of this paper are as follows: (i) the explicit solutions to a likelihood expression incorporating variable localization error that leads directly to efficient computational algorithms [Eq. (10)]; (ii) the confirmation that a log-normal approximation to the likelihood function around the MLE enables calculation of more accurate confidence intervals for uncertainty quantification on a per trajectory basis than was previously possible; and (iii) a demonstration of experimentally relevant conditions where our likelihood expression is noticeably more reliable than the likelihood expression derived in previous work.

The paper offers three detailed analytical solutions and associated computational algorithms to the likelihood expression in order to provide a basis for future developments in single-particle trajectory analysis. Although the recursive method was used exclusively when performing the analysis in Sec. VII, there exists published literature for the Fisher information in the basis of the Markov method [30]. Additionally, the Laplace method has the advantage that the expected true positions $\hat{X}$ are computed along with the likelihood; this may be useful for those interested in estimating a MLE trajectory for subsequent analysis beyond simple estimates of D. Furthermore, the Markov method is crucial for deriving the terms ϵ_i in the components $\mathcal{M}_i$ given knowledge of the true underlying probability distribution.

We applied a log-normal approximation to the likelihood distribution in order to generate simple confidence intervals for uncertainty quantification. The log-normal approximation was shown in simulation (Fig. 2) to be a more accurate approximation to the true log-likelihood than the normal approximation, thus allowing us to use the point-estimate MLE and observed information to compute accurate confidence intervals. A dimensionless value of $|\ln(\hat{D}/D)| < 1$ was found to be a useful threshold below which the confidence intervals are reliable (Fig. 6). Thus, the simple log-normal

approximation to the shape of the likelihood function is only useful a bounded distance away from $\hat{D}$. When the value of $D\tau$ is very small compared with $\langle V \rangle$, it is possible that the likelihood has a maximum arbitrarily close to 0, in which case the first derivative fails to vanish at the MLE and the observed information becomes meaningless; this behavior was observed in simulation and corresponded to our analysis of the expected Fisher information.

The analysis of the expected information was crucial for showing that there exists a resolution limit on the smallest $\hat{D}$ that can be reliably calculated given a particular V (and τ); it was also shown that this limit could be extended by increasing the trajectory length. The effect of motion-blur was shown to reduce the information as well. Finally, we have shown that the expected information from previously reported likelihood expressions that only incorporate a scalar localization variance $\langle V \rangle$ is always less than the expected information from our improved likelihood expression that fully utilizes all of the vector-valued localization variances, V.

Incorporating variable localization errors into the likelihood expression greatly improves the MLE estimators' precision, robustness, and ability to report reliable confidence intervals under conditions similar to those encountered in real microscopy experiments. Our direct analytical approach to evaluation of the likelihood expression is computationally simple and generally useful for extension to future work on this topic. A Bayesian approach could further improve the limits of estimation by incorporating prior knowledge and making use of the full posterior distribution of the parameter estimate, but the MLE-based point estimate approach we describe already provides a large improvement upon what was originally considered the lower limits of $\hat{D}$ estimation for experimental trajectories [17], and more importantly it is able to accurately detect when diffusion estimates for individual trajectories are reliable by using the observed information and corresponding confidence intervals.

ACKNOWLEDGMENTS

We wish to acknowledge Stanly Steinberg, Michael Wester, and Marjolein Meddens for reading our manuscript and providing enlightening discussions and helpful comments. Financial support for this work was provided primarily by National Science Foundation Grant No. 0954836. Additional support was provided by The New Mexico Spatiotemporal Modeling Center: NIH P50GM085273 (K.A.L.), NIH Grant No. 1R01GM100114 (K.A.L.,M.J.O.), and NIH Grant No. 1R01NS071116 (K.A.L., M.J.O.).

K.A.L., P.J.C., and P.K.R. conceived the project. P.J.C. initiated the formulation of the estimation problem as a MLE given a set of observations. P.K.R. derived the recursive, Markov, and Laplace methods and wrote the initial MATLAB implementations. M.J.O. derived the efficient algorithms for the three solution methods and the C++ and MATLAB implementations of the methods, and helped to simplify the presentation. P.K.R. established the information-based analysis. P.K.R. wrote the diffusion simulation software and performed the subsequent analysis. All authors contributed to the writing and editing of the manuscript.

APPENDIX A: GAUSSIAN PROPERTIES

The normalized Gaussian function is represented as

$$\mathcal{N}(a,a_0,\eta) = \frac{1}{\sqrt{2\pi\eta}} \exp\left[-\frac{(a-a_0)^2}{2\eta}\right].$$

The Gaussian function is symmetric with respect to the first two position parameters, so that $\mathcal{N}(a,b,v) = \mathcal{N}(b,a,v)$, and the variance of the function is given by the third parameter. Also, the normalization factor ensures that the Gaussian integrated over all space with respect to either of its position parameters is unity,

$$\int_{-\infty}^{\infty} da\ \mathcal{N}(a,b,v) = \int_{-\infty}^{\infty} db\ \mathcal{N}(a,b,v) = 1.$$

As a corollary, a normalized Gaussian has a useful scaling identity, for $q > 0$,

$$\mathcal{N}(a,b,v) = q\,\mathcal{N}(qa,qb,q^2 v).$$

Next, consider the case of the product of two normalized Gaussians sharing a common position parameter, which can be rewritten as a product of two normalized Gaussians where the common parameter only appears in one of the two,

$$\mathcal{N}(x,\mu_1,\eta_1)\mathcal{N}(x,\mu_2,\eta_2) = \mathcal{N}(\mu_1,\mu_2,\eta_1+\eta_2)\mathcal{N}(x,\mu',\eta'), \tag{A1}$$

where

$$\mu' = \frac{\mu_1\eta_2 + \mu_2\eta_1}{\eta_1+\eta_2} \quad \text{and} \quad \eta' = \frac{\eta_1\eta_2}{\eta_1+\eta_2}.$$

Using Eq. (A1), the integral of the product of two normalized Gaussians over a shared position parameter is itself a normalized Gaussian in the other two (unintegrated) position parameters,

$$\begin{aligned}&\int dx \mathcal{N}(x,\mu_1,\eta_1)\mathcal{N}(x,\mu_2,\eta_2)\\ &\quad = \mathcal{N}(\mu_1,\mu_2,\eta_1+\eta_2)\int dx \mathcal{N}(x,\mu',\eta')\\ &\quad = \mathcal{N}(\mu_1,\mu_2,\eta_1+\eta_2).\end{aligned} \tag{A2}$$

With two successive applications of Eq. (A1), the product of three Gaussians that share a common position parameter becomes

$$\begin{aligned}&\mathcal{N}(x,\mu_1,\eta_1)\mathcal{N}(x,\mu_2,\eta_2)\mathcal{N}(x,\mu_3,\eta_3)\\ &\quad = \mathcal{N}(\mu_1,\mu_2,\eta_1+\eta_2)\mathcal{N}(\mu_3,\mu',\eta'+\eta_3)\mathcal{N}(x,\mu'',\eta''),\\ &\text{where}\quad \mu'' = \frac{\mu'\eta_3 + \mu_3\eta'}{\eta'+\eta_3},\\ &\text{and}\quad \eta'' = \frac{\eta'\eta_3}{\eta'+\eta_3}.\end{aligned} \tag{A3}$$

Finally, Eq. (A3) allows the integral of three normalized Gaussians over a shared position parameter to reduce to

$$\begin{aligned}&\int dx\ \mathcal{N}(x,\mu_1,\eta_1)\mathcal{N}(x,\mu_2,\eta_2)\mathcal{N}(x,\mu_3,\eta_3)\\ &\quad = \mathcal{N}(\mu_1,\mu_2,\eta_1+\eta_2)\mathcal{N}(\mu_3,\mu',\gamma),\\ &\text{where}\quad \gamma = \frac{\eta_1\eta_2+\eta_1\eta_3+\eta_2\eta_3}{\eta_1+\eta_2}.\end{aligned} \tag{A4}$$

APPENDIX B: THE PROBABILITY DENSITY OF A TIME-AVERAGED POSITION

Given D, the probability density of a transition from point a to point b separated by a time T is $P(b|a) = \mathcal{N}(b,a,2DT)$. If an intermediate point, $y(t)$, sampled at a time $t < T$ is considered, the joint probability density of a transition from a to $y(t)$ and then from $y(t)$ to b is

$$\begin{aligned} P(y(t),a,b) &= \mathcal{N}(y(t),a,2Dt)\mathcal{N}(b,y(t),2D(T-t)) \\ &= \mathcal{N}(b,a,2DT)\mathcal{N}\left(y(t),a\left(1-\frac{t}{T}\right)\right. \\ &\quad \left. + b\frac{t}{T},2D\frac{t}{T}(T-t)\right). \end{aligned}$$

The probability density of the variable $y(t)$, preconditioned on the end points a and b, is then

$$P(y(t)|b,a) = \mathcal{N}\left(y(t),a\left(1-\frac{t}{T}\right)+b\frac{t}{T},2D\frac{t}{T}(T-t)\right). \tag{B1}$$

Equation (B1) is the probability density for what is known as a Brownian bridge [31], or Brownian motion with preconditioned end points. It has a mean and covariance defined as

$$\begin{aligned} \langle y(t)\rangle &= a\left(1-\frac{t}{T}\right)+b\frac{t}{T} \\ \mathrm{cov}[y(t),y(s)] &= 2D\left(s-\frac{st}{T}\right) \quad \text{for} \quad s \leqslant t \leqslant T. \end{aligned}$$

It is now of interest to find the probability density for a quantity that describes an integrated average of $y(t)$ such that

$$\bar{y} = \frac{1}{t_\varepsilon}\int_{t=0}^{t_\varepsilon} dt\; y(t).$$

Since each $y(t)$ is a normally distributed random variable, $y(t)$ is a Gaussian process, and the time-averaged integral of a Gaussian process, $\bar{y}$, is also a normally distributed random variable. Therefore, from Isserlis' theorem [32], only the first two moments of $\bar{y}$ are needed to determine its probability distribution. The first moment is $\langle\bar{y}\rangle = (1-\alpha)a + \alpha b$, where for notational convenience we define $\alpha = t_\epsilon/2T$. It follows that the second moment is

$$\langle\bar{y}^2\rangle = \langle\bar{y}\rangle^2 + \frac{1}{t_\varepsilon^2}\int_{t=0}^{t_\varepsilon} dt \int_{s=0}^{t_\varepsilon} ds\; \mathrm{cov}[y(t),y(s)].$$

Therefore, the variance for $\bar{y}$ is

$$\langle\bar{y}^2\rangle - \langle\bar{y}\rangle^2 = 2Dt_\epsilon\left[\frac{1}{3}-\frac{\alpha}{2}\right].$$

Given that $\bar{y}$ must be normally distributed and its first two moments are known,

$$\begin{aligned} P(\bar{y}|a,b) &= \mathcal{N}(\bar{y},\langle\bar{y}\rangle,\langle\bar{y}^2\rangle - \langle\bar{y}\rangle^2) \\ &= \mathcal{N}\left(\bar{y},(1-\alpha)a+\alpha b,2Dt_\varepsilon\left[\frac{1}{3}-\frac{\alpha}{2}\right]\right). \end{aligned}$$

For single-molecule localization, a time-averaged position is related to a localized observation by a normal Gaussian function so that

$$\begin{aligned} P(o|a,b) &= \int d\bar{y}\; P(o|\bar{y})P(\bar{y}|a,b) \\ &= \mathcal{N}\left(o,(1-\alpha)a+\alpha b,v+2Dt_\varepsilon\left[\frac{1}{3}-\frac{\alpha}{2}\right]\right). \end{aligned} \tag{B2}$$

APPENDIX C: TIME-AVERAGED PROBABILITIES TO MEASUREMENT FUNCTIONS

In Sec. III C, Markov's approach showed the following relationship for a multivariate Gaussian: $\Sigma_{i,j} = \langle s_i s_j\rangle$, where Σ is the covariance matrix of a multivariate Gaussian function describing the vector of random variables $S = \{s_i = o_{i+1} - o_i\}_{i=1}^{N-1}$. If the moments of S are known, the parameter ϵ_i for the simpler expression can be derived given that $\langle s_i s_{i+1}\rangle = -\epsilon_{i+1}$ was shown to be true for the formalism with the simpler expression as derived in Sec. III C. Starting from Eq. (B2) for an arbitrary o_i,

$$\begin{aligned} &P(o_i|x_i,x_{i+1}) \\ &\quad = \mathcal{N}\left(o_i,(1-\alpha_i)x_i+\alpha_i x_{i+1},v_i+2Dt_\epsilon\left[\frac{1}{3}-\frac{\alpha_i}{2}\right]\right), \end{aligned}$$

where v_i is defined as the localization variance and $\alpha_i = t_\epsilon/(2\delta t_i)$. From the properties of a Gaussian function with 0 mean,

$$\begin{aligned} &\langle o_i - (1-\alpha_i)x_i - \alpha_i x_{i+1}\rangle = 0, \\ &\langle[o_i - (1-\alpha_i)x_i - \alpha_i x_{i+1}]^2\rangle = v_i + 2Dt_\epsilon\left[\frac{1}{3}-\frac{\alpha_i}{2}\right], \end{aligned} \tag{C1}$$

which implies

$$\begin{aligned} \langle o_i - x_i\rangle &= 0, \\ \langle o_i - x_i|X\rangle &= \alpha_i(x_{i+1}-x_i), \\ \langle(o_i-x_i)^2\rangle &= v_i + 2Dt_\epsilon\tfrac{1}{3}, \\ \langle o_{i+1}-o_i\rangle &= \langle s_i\rangle = 0. \end{aligned} \tag{C2}$$

Given Eq. (C2), $\langle s_i s_{i+1}\rangle = -\varepsilon_{i+1} = 2Dt_\epsilon/6 - v_{i+1}$. Additionally, it follows that

$$\langle s_i^2\rangle = w_i + \epsilon_i + \epsilon_{i+1} = 2D\delta t_i - 4Dt_\epsilon\tfrac{1}{6} + v_i + v_{i+1}.$$

Therefore, $\varepsilon_i(D) = v_i - 2Dt_\epsilon/6$, which is the variance correction discovered in earlier diffusion estimation papers [15,18,19].

APPENDIX D: LAPLACE METHOD MLE POSITIONS

Recalling the objective function in the Laplace method,

$$\begin{aligned} -\ln[f(\mathbf{X})] &= \sum_{i=1}^{N}\left[\frac{1}{2}\ln(2\pi\varepsilon_i) + \frac{(o_i-x_i)^2}{2\varepsilon_i}\right] \\ &\quad + \sum_{i=1}^{N-1}\left[\frac{1}{2}\ln(2\pi\omega_i) + \frac{(x_{i+1}-x_i)^2}{2\omega_i}\right], \end{aligned} \tag{D1}$$

the gradient of Eq. (D1) is

$$\begin{aligned}
-\frac{\partial \ln f}{\partial x_1} &= \frac{(x_1 - o_1)}{\varepsilon_1} + \frac{(x_1 - x_2)}{\omega_1},\\
-\frac{\partial \ln f}{\partial x_i} &= \frac{(x_i - o_i)}{\varepsilon_i} + \frac{(x_i - x_{i-1})}{\omega_{i-1}} + \frac{(x_i - x_{i+1})}{\omega_i},\\
-\frac{\partial \ln f}{\partial x_N} &= \frac{(x_N - o_N)}{\varepsilon_N} + \frac{(x_N - x_{N-1})}{\omega_{N-1}},
\end{aligned} \tag{D2}$$

where $i \in 2 : N-1$. The Hessian $-\ln\nabla\nabla f(\widehat{X}) = M$ of Eq. (D1) has the nonzero elements

$$\begin{aligned}
M_{1,1} &= -\frac{\partial^2 \ln f}{\partial x_1 \partial x_1} = \frac{1}{\varepsilon_1} + \frac{1}{\omega_1},\\
M_{i,i} &= -\frac{\partial^2 \ln f}{\partial x_i \partial x_i} = \frac{1}{\varepsilon_i} + \frac{1}{\omega_{i-1}} + \frac{1}{\omega_i},\\
M_{N,N} &= -\frac{\partial^2 \ln f}{\partial x_N \partial x_N} = \frac{1}{\varepsilon_N} + \frac{1}{\omega_{N-1}},\\
M_{i,i+1} &= -\frac{\partial^2 \ln f}{\partial x_i \partial x_{i+1}} = -\frac{1}{\omega_i},\\
M_{i,i-1} &= -\frac{\partial^2 \ln f}{\partial x_i \partial x_{i-1}} = -\frac{1}{\omega_{i-1}}.
\end{aligned} \tag{D3}$$

Setting the gradient in Eq. (D2) equal to 0 and moving the constants to the left-hand side of the equation gives

$$\begin{aligned}
\frac{o_1}{\varepsilon_1} &= \frac{\widehat{x}_1}{\varepsilon_1} + \frac{(\widehat{x}_1 - \widehat{x}_2)}{\omega_1},\\
\frac{o_i}{\varepsilon_i} &= \frac{\widehat{x}_i}{\varepsilon_i} + \frac{(\widehat{x}_i - \widehat{x}_{i-1})}{\omega_{i-1}} + \frac{(\widehat{x}_i - \widehat{x}_{i+1})}{\omega_i},\\
\frac{o_N}{\varepsilon_N} &= \frac{\widehat{x}_N}{\varepsilon_N} + \frac{(\widehat{x}_N - \widehat{x}_{N-1})}{\omega_{N-1}}.
\end{aligned}$$

With additional factoring, the expression becomes

$$\begin{aligned}
\frac{o_1}{\varepsilon_1} &= \widehat{x}_1\left(\frac{1}{\varepsilon_1} + \frac{1}{\omega_1}\right) + \widehat{x}_2\left(\frac{-1}{\omega_1}\right),\\
\frac{o_i}{\varepsilon_i} &= \widehat{x}_i\left(\frac{1}{\varepsilon_i} + \frac{1}{\omega_{i-1}} + \frac{1}{\omega_i}\right) + \widehat{x}_{i-1}\left(\frac{-1}{\omega_{i-1}}\right) + \widehat{x}_{i+1}\left(\frac{-1}{\omega_i}\right),\\
\frac{o_N}{\varepsilon_N} &= \widehat{x}_N\left(\frac{1}{\varepsilon_N} + \frac{1}{\omega_{N-1}}\right) + \widehat{x}_{N-1}\left(\frac{-1}{\omega_{N-1}}\right).
\end{aligned}$$

The factored expression on the right can be expressed in terms of a vector product of the Hessian matrix and the maximum likelihood of the true positions, $M\widehat{X}$. Multiplying by the inverse Hessian matrix results in the solution $\widehat{X} = M^{-1}\Theta$, where the components of Θ are $\theta_i = o_i/\varepsilon_i$.

APPENDIX E: AN OBSERVED INFORMATION CALCULATION

The observed information is necessary for calculating reliable confidence intervals around the MLE. With the recursive solution, the observed information can be solved recursively in linear time,

$$\begin{aligned}
\mathcal{K}(\hat{D}) &= \frac{\partial^2}{\partial \hat{D}^2}\sum_{i=1}^{N-1}\frac{1}{2}\ln(2\pi\alpha_i) + \frac{(o_{i+1} - \mu_i)^2}{2\alpha_i}\\
&= \sum_{i=1}^{N-1}\frac{1}{2\alpha_i}\left\{\frac{\partial^2\alpha_i}{\partial\hat{D}^2} - \left(\frac{\partial\alpha_i}{\partial\hat{D}}\right)^2\frac{1}{\alpha_i}\right\} + \frac{(o_{i+1} - \mu_i)^2}{\alpha_i^2}\left[\left(\frac{\partial\alpha_i}{\partial\hat{D}}\right)^2\frac{1}{\alpha_i} - \frac{1}{2}\frac{\partial^2\alpha_i}{\partial\hat{D}^2}\right]\\
&\quad + \frac{1}{\alpha_i}\left\{(o_{i+1} - \mu_i)\left[\frac{2}{\alpha_i}\frac{\partial\alpha_i}{\partial\hat{D}}\frac{\partial\mu_i}{\partial\hat{D}} - \frac{\partial^2\mu_i}{\partial\hat{D}^2}\right] + \left(\frac{\partial\mu_i}{\partial\hat{D}}\right)^2\right\},
\end{aligned}$$

where the derivatives of α_i and μ_i can be solved recursively given that the initial first and second derivatives are

$$\frac{\partial^2\mu_1}{\partial\hat{D}^2} = 0, \quad \frac{\partial^2\alpha_1}{\partial\hat{D}^2} = 0, \quad \frac{\partial\mu_1}{\partial\hat{D}} = 0, \quad \frac{\partial\alpha_1}{\partial\hat{D}} = \frac{\partial\varepsilon_1}{\partial\hat{D}} + \frac{\partial\varepsilon_2}{\partial\hat{D}} + \frac{\partial\omega_1}{\partial\hat{D}}.$$

The recursions on the second derivatives are

$$\begin{aligned}
\frac{\partial^2\mu_i}{\partial\hat{D}^2} &= \frac{\varepsilon_i}{\alpha_{i-1}}\frac{\partial^2\mu_{i-1}}{\partial\hat{D}^2} + \frac{2}{\alpha_{i-1}}\frac{\partial\mu_{i-1}}{\partial\hat{D}}\left[\frac{\partial\varepsilon_i}{\partial\hat{D}} - \frac{\varepsilon_i}{\alpha_{i-1}}\frac{\partial\alpha_{i-1}}{\partial\hat{D}}\right] + \frac{(o_i - \mu_{i-1})}{\alpha_{i-1}^2}\left[2\frac{\partial\varepsilon_i}{\partial\hat{D}}\frac{\partial\alpha_{i-1}}{\partial\hat{D}} - 2\frac{\varepsilon_i}{\alpha_{i-1}}\left(\frac{\partial\alpha_{i-1}}{\partial\hat{D}}\right)^2 + \varepsilon_i\frac{\partial^2\alpha_{i-1}}{\partial\hat{D}^2}\right],\\
\frac{\partial^2\alpha_i}{\partial\hat{D}^2} &= \left(\frac{\varepsilon_i}{\alpha_{i-1}}\right)^2\frac{\partial^2\alpha_{i-1}}{\partial\hat{D}^2} + \frac{4\varepsilon_i}{\alpha_{i-1}^2}\frac{\partial\varepsilon_i}{\partial\hat{D}}\frac{\partial\alpha_{i-1}}{\partial\hat{D}} - \frac{2\varepsilon_i^2}{\alpha_{i-1}^3}\left(\frac{\partial\alpha_{i-1}}{\partial\hat{D}}\right)^2 - \frac{2}{\alpha_{i-1}}\left(\frac{\partial\varepsilon_i}{\partial\hat{D}}\right)^2,
\end{aligned}$$

and the recursions on the first derivatives are

$$\frac{\partial\mu_i}{\partial\hat{D}} = \frac{\varepsilon_i}{\alpha_{i-1}}\frac{\partial\mu_{i-1}}{\partial\hat{D}} + \frac{(o_i - \mu_{i-1})}{\alpha_{i-1}}\left(\frac{\varepsilon_i}{\alpha_{i-1}}\frac{\partial\alpha_{i-1}}{\partial\hat{D}} - \frac{\partial\varepsilon_i}{\partial\hat{D}}\right), \quad \frac{\partial\alpha_i}{\partial\hat{D}} = \frac{\partial\omega_i}{\partial\hat{D}} + \frac{\partial\varepsilon_{i+1}}{\partial\hat{D}} + \frac{\partial\varepsilon_i}{\partial\hat{D}} + \frac{\varepsilon_i}{\alpha_{i-1}}\left(\frac{\varepsilon_i}{\alpha_{i-1}}\frac{\partial\alpha_{i-1}}{\partial\hat{D}} - 2\frac{\partial\varepsilon_i}{\partial\hat{D}}\right).$$

APPENDIX F: AN EXPECTED INFORMATION CALCULATION

The expected information, conditional on v_i, can be represented as

$$\begin{aligned}\mathcal{I}(\hat{D}) &= \left\langle \frac{\partial^2}{\partial \hat{D}^2} \sum_{i=1}^{N-1} \frac{1}{2} \ln(2\pi\alpha_i) + \frac{(o_{i+1}-\mu_i)^2}{2\alpha_i} \right\rangle \\ &= \sum_{i=1}^{N-1} \frac{1}{2\alpha_i} \left[\frac{\partial^2 \alpha_i}{\partial \hat{D}^2} - \left(\frac{\partial \alpha_i}{\partial \hat{D}} \right)^2 \frac{1}{\alpha_i} \right] + \frac{\langle (o_{i+1}-\mu_i)^2 \rangle}{\alpha_i^2} \left[\left(\frac{\partial \alpha_i}{\partial \hat{D}} \right)^2 \frac{1}{\alpha_i} - \frac{1}{2} \frac{\partial^2 \alpha_i}{\partial \hat{D}^2} \right] \\ &\quad + \frac{1}{\alpha_i} \left\{ 2 \left\langle (o_{i+1}-\mu_i) \frac{\partial \mu_i}{\partial \hat{D}} \right\rangle \frac{\partial \alpha_i}{\partial \hat{D}} \frac{1}{\alpha_i} - \left\langle (o_{i+1}-\mu_i) \frac{\partial^2 \mu_i}{\partial \hat{D}^2} \right\rangle + \left\langle \left(\frac{\partial \mu_i}{\partial \hat{D}} \right)^2 \right\rangle \right\}.\end{aligned}$$

The Fisher information can be calculated recursively by evaluating each of the expectation values. The representation of the recursive variables as separated Gaussians leads to the following relations:

$$\langle (o_{i+1}-\mu_i)^2 \rangle = \alpha_i, \quad \langle (o_{i+1}-\mu_i)(o_i - \mu_{i-1}) \rangle = 0.$$

The remaining nonzero expectation value has the initial condition

$$\left\langle \left(\frac{\partial^2 \mu_1}{\partial \hat{D}^2} \right)^2 \right\rangle = 0,$$

and it can be solved recursively,

$$\left\langle \left(\frac{\partial \mu_i}{\partial \hat{D}} \right)^2 \right\rangle = \frac{1}{\alpha_{i-1}} \left(\frac{\partial \alpha_{i-1}}{\partial \hat{D}} \frac{\varepsilon_i}{\alpha_{i-1}} - \frac{\partial \varepsilon_i}{\partial \hat{D}} \right)^2 + \left(\frac{\varepsilon_i}{\alpha_{i-1}} \right)^2 \left\langle \left(\frac{\partial \mu_{i-1}}{\partial \hat{D}} \right)^2 \right\rangle.$$

Then, $\mathcal{I}(\hat{D})$ reduces to

$$\mathcal{I}(\hat{D}) = \sum_{i=1}^{N-1} \frac{1}{2\alpha_i^2} \left(\frac{\partial \alpha_i}{\partial \hat{D}} \right)^2 + \frac{1}{\alpha_i} \left\langle \left(\frac{\partial \mu_i}{\partial \hat{D}} \right)^2 \right\rangle. \tag{F1}$$

[1] M. J. Saxton and K. Jacobson, Annu. Rev. Biophys. Biomolec. Struct. **26**, 373 (1997).
[2] M. J. Saxton, Biophys. J. **99**, 1490 (2010).
[3] A. Sanamrad, F. Persson, E. G. Lundius, D. Fange, A. H. Gynnå, and J. Elf, Proc. Natl. Acad. Sci. USA **111**, 11413 (2014).
[4] C. P. Calderon, M. A. Thompson, J. M. Casolari, R. C. Paffenroth, and W. Moerner, J. Phys. Chem. B **117**, 15701 (2013).
[5] N. Monnier, S.-M. Guo, M. Mori, J. He, P. Lénárt, and M. Bathe, Biophys. J. **103**, 616 (2012).
[6] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser, Nat. Meth. **5**, 695 (2008).
[7] A. Sergé, N. Bertaux, H. Rigneault, and D. Marguet, Nat. Meth. **5**, 687 (2008).
[8] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter *et al.*, Nat. Meth. **11**, 281 (2014).
[9] A. D. Mont, C. P. Calderon, and A. B. Poore, in *SPIE Defense+ Security* (International Society for Optics and Photonics, Baltimore, MD, 2014).
[10] J. W. Yoon, A. Bruckbauer, W. J. Fitzgerald, and D. Klenerman, Biophys. J. **94**, 4932 (2008).
[11] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, Nat. Meth. **7**, 373 (2010).
[12] Y. Chen, J. Vela, H. Htoon, J. L. Casson, D. J. Werder, D. A. Bussian, V. I. Klimov, and J. A. Hollingsworth, J. Am. Chem. Soc. **130**, 5026 (2008).
[13] H. Qian, M. P. Sheetz, and E. L. Elson, Biophys. J. **60**, 910 (1991).
[14] D. S. Martin, M. B. Forstner, and J. A. Käs, Biophys. J. **83**, 2109 (2002).
[15] T. Savin and P. S. Doyle, Biophys. J. **88**, 623 (2005).
[16] X. Michalet, Phys. Rev. E **82**, 041914 (2010).
[17] X. Michalet and A. J. Berglund, Phys. Rev. E **85**, 061916 (2012).
[18] D. Montiel, H. Cang, and H. Yang, J. Phys. Chem. B **110**, 19763 (2006).
[19] A. J. Berglund, Phys. Rev. E **82**, 011917 (2010).
[20] B. Shuang, C. P. Byers, L. Kisley, L.-Y. Wang, J. Zhao, H. Morimura, S. Link, and C. F. Landes, Langmuir: ACS J. Surf. Colloids **29**, 228 (2013).
[21] C. L. Vestergaard, P. C. Blainey, and H. Flyvbjerg, Phys. Rev. E **89**, 022726 (2014).
[22] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[23] M. E. El-Mikkawy, Appl. Math. Comput. **150**, 669 (2004).
[24] Y. Pawitan, *In All Likelihood: Statistical Modeling and Inference using Likelihood* (Oxford University Press, Oxford, 2001).
[25] B. Efron and D. V. Hinkley, Biometrika **65**, 457 (1978).
[26] P. K. Relich, M. J. Olah, P. J. Cutler, and K. A. Lidke, See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.93.042401 for software to estimate diffusion and generate the figures in this paper.

[27] S. S. Wilks, Ann. Math. Stat. **9**, 60 (1938).
[28] B. R. Frieden, *Physics from Fisher Information: A Unification* (Cambridge University Press, Cambridge, 1998).
[29] A. Parsian and N. Nematollahi, J. Statist. Planning Infer. **52**, 77 (1996).
[30] K. V. Mardia and R. Marshall, Biometrika **71**, 135 (1984).
[31] S. M. Ross *et al.*, *Stochastic Processes* (Wiley, New York, 1996), Vol. 2.
[32] L. Isserlis, Biometrika **12**, 134 (1918).