

# Statistics of camera-based single-particle tracking

Andrew J. Berglund\*

*Center for Nanoscale Science and Technology,  
National Institute of Standards and Technology,  
Gaithersburg, MD 20899 USA*

Camera-based single-particle tracking enables quantitative determination of transport properties and provides nanoscale information about material characteristics such as viscosity and elasticity. However, static localization noise and the blurring of a particle's position over camera integration times introduce artifacts into measurement results even for a particle executing simple diffusion. Common data analysis methods based on the mean-square displacement do not properly account for these effects. In this paper, we analyze the statistics of tracking data for freely-diffusing particles in realistic experimental scenarios. We derive a convenient and asymptotically optimal maximum likelihood estimator for the diffusion coefficient and for the magnitude of localization noise together with the corresponding Fisher information, which bounds the performance of all unbiased estimators. We find that the effect of varying the illumination profile during the camera integration time is quantified by a motion blur coefficient,  $R$ . We also find that a double-pulse illumination sequence maximizes the information content in some common experimental scenarios. Our results provide a rigorous theoretical framework and practical experimental recipe for achieving optimal performance in camera-based single-particle tracking.

## I. INTRODUCTION

Since its inception over two decades ago [1, 2], camera-based single-particle tracking has become a common tool in many scientific areas, including membrane biophysics [3], colloid physics [4], and microrheology of complex fluids [5–8]. In a typical experiment, the trajectory of an individual particle is imaged through a microscope and recorded with a digital camera at a frame rate ranging from a few Hz up to a few hundred Hz. The position of the particle is extracted from a sequence of images [4] routinely, yet remarkably, with nanometer accuracy [9–11]. Transport properties, particle size and shape, and material properties such as viscosities or frequency-dependent elastic moduli can be determined from measured velocities and diffusion coefficients, which are extracted from raw data - a sequence of tracked positions  $\{X_k\}$ . The quantitative utility of this widespread method relies critically on the quality of the statistical inference that relates observed trajectories to underlying diffusion coefficients. However, the estimation of diffusion coefficients from realistic particle-tracking data is a subtle task, with several difficulties recognized only recently. In this paper, we develop a rigorous but practical statistical framework for overcoming these difficulties.

Estimators of the diffusion coefficient from single-particle tracking data have traditionally relied on the mean-square displacement (MSD), a measure of the fluctuations in a particle's position over specified time intervals. However, unavoidable experimental realities of camera-based single-particle tracking, such as finite-resolution imaging and motion blurring due to camera

integration times, together with technical difficulties arising from correlations between MSD values calculated at different time intervals, obscure the statistics of the MSD. These difficulties commonly result in experimental artifacts, inefficient use of data, and possibly incorrect measurement results. In this paper, we resolve these difficulties by presenting a complete statistical description of camera-based single-particle tracking for particles undergoing pure diffusion. In Section II, we briefly review the utility and difficulties of using the MSD for single-particle tracking data analysis. In Section III, we calculate the full distribution of single-particle tracking measurement results in the presence of motion blur and Gaussian-distributed localization noise. Despite its fundamental importance for data analysis, this distribution has not previously been presented. In Sections IV and V it is shown that for large data sets, this distribution is approximately diagonal in the frequency (Fourier) basis so that the power spectrum of the displacements offers significantly simplified statistics while also having a straightforward interpretation. With this approximate result, we compute an asymptotically optimal maximum likelihood estimator (MLE) of  $D$  and  $\sigma$  together with the corresponding Fisher information matrix, which bounds the variance of any unbiased estimator of  $D$  and  $\sigma$  including those based on the MSD. Finally, in Section VI we show through numerical simulation that the MLE significantly outperforms the MSD as a data analysis tool. Possible extensions of these results to anomalous diffusion and microrheology of linear viscoelastic materials are briefly discussed in section VII. In Appendix A, we update previous literature results on the statistics of the MSD to include the effects of motion blur and localization noise. In Appendix B, we give the technical details of an important mathematical approximation.

---

\*Electronic address: [andrew.berglund@nist.gov](mailto:andrew.berglund@nist.gov)

## II. THE MEAN-SQUARE DISPLACEMENT

In this section, we review the status of the most common data analysis tool in single-particle tracking, the mean-square displacement (MSD). Consider the analysis of one-dimensional motion in the  $x$ -direction from a data set  $\{X_k\}$ ,  $1 \leq k \leq N+1$ . In a fundamental study [12], Qian et al. analyzed the statistics of the MSD, defined as

$$\rho_n = \frac{1}{N-n+1} \sum_{k=1}^{N-n+1} (X_{k+n} - X_k)^2 \quad (1)$$

with  $1 \leq n \leq N$ . They calculated both the mean and variance of  $\rho_n$ , showing that for a particle with diffusion coefficient  $D$  sampled at time intervals  $\Delta t$ , the expected value of the MSD is

$$\langle \rho_n \rangle = 2Dn\Delta t, \quad (2)$$

where  $\langle \cdot \rangle$  represents ensemble averaging over repeated realizations of the process. Based on this equation,  $\rho_n$  has a simple visual interpretation: for a particle undergoing simple diffusion, the expectation value of a plot of  $\rho_n$  vs.  $n$  should be a straight line with slope  $2D\Delta t$ . Conversely, a non-linear relation between (the expected value of)  $\rho_n$  and  $n$  is taken as an indication of non-Brownian, anomalous diffusion. The statistical analysis of Ref. [12] and the intuitive visual interpretation of  $\rho_n$  have led to widespread adoption of the MSD for data analysis in single-particle tracking [3].

There are two significant difficulties with data analysis based on the MSD. First, the transformation of the raw data  $\{X_k\}$  into  $\{\rho_n\}$  results in a complicated, highly correlated data set. Qian et al. noted this difficulty, and quantified it through calculation of the covariance  $C_{mn} = \langle \rho_m \rho_n \rangle - \langle \rho_m \rangle \langle \rho_n \rangle$ .  $C_{nn}$ , the variance in  $\rho_n$ , is strongly dependent on  $n$ , which shows that all data points in a plot of  $\rho_n$  vs.  $n$  cannot be assigned equal statistical weight. More troubling, when  $m \neq n$ , the covariance  $C_{mn}$  is non-zero because  $\rho_n$  and  $\rho_m$  are derived from the same underlying data set  $\{X_k\}$  [12]. As a result, the values of  $\rho_n$  from a single experiment are not symmetrically distributed about the line given in Eq. 2. Following up on these observations, Saxton found that even for pure, unobstructed diffusion, numerical estimates of  $D$  obtained by fitting a line to  $\rho_n$  vs.  $n$  depend on subjective choices such as the range of  $n$  over which to perform the fit. Seemingly satisfactory results were obtained for both pure and obstructed diffusion using statistical weight coefficients proportional to  $C_{nn}$  [13]. The off-diagonal correlations  $C_{mn}$  were not considered.

The second difficulty with the MSD arises because the statistics of  $\rho_n$  considered above only apply when the data set  $\{X_k\}$  represents the *true* trajectory of the particle. This is not the case in practical scenarios, and consequent difficulties with data analysis based on the mean-square displacement have been recognized more recently, and only partially. Martin et al. recognized that static

localization noise - the random error in the measurement of an immobilized particle's position - alters the MSD [14]. They found (in the present notation) that

$$\langle \rho_n \rangle = 2Dn\Delta t + 2\sigma^2, \quad (3)$$

where  $\sigma$  is the static localization error, the standard deviation in measured positions of an immobile particle [10, 11]. A log-log plot of  $\rho_n$  vs.  $n$ , intended to ferret out deviations from linear scaling, will therefore exhibit a reduced slope at small  $n$  and may be incorrectly attributed to anomalous “sub-diffusion.” In the context of particle-tracking microrheology, Savin and Doyle [15] recognized the additional effect of motion blur (“dynamic error”), wherein a particle's *average* position over the camera frame interval  $\Delta t$  is measured, rather than its instantaneous position as tacitly assumed in Eqs. 2 and 3. They found, for averaging of the position over the full frame, that

$$\langle \rho_n \rangle = 2Dn\Delta t + 2\sigma^2 - \frac{2}{3}D\Delta t. \quad (4)$$

Thus, even if static errors are absent ( $\sigma = 0$ ), dynamic errors alter the MSD. Motion blur due to full-frame averaging was also considered in Ref. [16] in the context of optical-tweezer calibration, in Ref. [17] as it relates specifically to diffusion coefficient estimation, and in Ref. [18] to determine the static error  $\sigma$  in a three-dimensional tracking geometry where particle immobilization was impractical.

Based on results given below, we find that each of these formulas is a specific case of

$$\langle \rho_n \rangle = 2Dn\Delta t + 2\sigma^2 - 4DR\Delta t \quad , \quad 0 \leq R \leq \frac{1}{4} \quad (5)$$

where the “motion blur coefficient”  $R$  (discussed in detail below) characterizes the illumination profile, or equivalently the shutter state, during the camera integration time. Thus we confirm that Eq. 2 applies when there is no measurement error ( $\sigma = 0$ ) and no motion blur ( $R = 0$ ), Eq. 3 applies when there is no motion blur, and Eq. 4 applies when the motion blur is due to full-frame averaging ( $R = 1/6$ ).

Summarizing the present status of the MSD, we find that consensus has been reached regarding the value of  $\langle \rho_n \rangle$ , but the full distribution of  $\{\rho_n\}$  has not been found even when localization noise and motion blur are neglected. Nevertheless, one can use the knowledge of  $\langle \rho_n \rangle$  to form an unbiased estimator of the diffusion coefficient, by choosing a particular  $n$  (usually  $n = 1$ ), plugging the observed value of  $\rho_n$  into Eq. (5), and solving for  $D$ . However, such an estimator “wastes” data in the sense that there exist other estimators that can perform just as well but with less data. On the other hand, we can estimate  $D$  with the more complicated line-fitting procedure described in Ref. [13], weighted by the variances  $C_{nn}$  while presumably a better estimate would also incorporate  $C_{mn}$ . However, this fitting procedure yields

such a complicated function of the data that questions of bias and efficiency become largely intractable. Furthermore, this procedure cannot be correctly implemented until the analysis of Ref. [12] is updated to include localization noise and motion blur. In Appendix A, we use the results of Section III below to accomplish this task by calculating  $C_{mn}$  for the case of non-zero  $R$  and  $\sigma$ .

### III. DISTRIBUTION OF MEASUREMENT RESULTS

In view of the difficulties described in Sect. II, a new data analysis method is desired that, like the MSD, has a simple interpretation but a firmer statistical footing. We begin by establishing a model of single-particle tracking and using it to derive the distribution of measurement results.

Suppose that a particle moving in one dimension by pure Brownian motion with diffusion coefficient  $D$  is imaged by a camera with frame integration time  $\Delta t$ . During the frame time  $\Delta t$ , the camera shutter may be opened or closed for a variable interval or, equivalently, the illuminating intensity may be varied during the frame. The observed position of the particle is then the average of its position weighted by a “shutter function”  $s(t)$ , a non-negative function whose integral over the frame is unity. Finally, each individual frame is corrupted by additive measurement noise  $\varepsilon$ . For the  $k$ th frame ending at time  $t = k\Delta t$ , the observed position  $X_k$  is given by

$$X_k = \int_{(k-1)\Delta t}^{k\Delta t} s[t' - (k-1)\Delta t] X_{true}(t') dt' + \varepsilon_k, \quad (6)$$

where  $X_{true}(t')$  is the true position of the particle at time  $t'$  and  $\varepsilon_k$  is the value of the additive localization noise in frame  $k$ . Next, we assume that  $\varepsilon_k$  can be approximated as zero-mean Gaussian measurement noise with  $\langle \varepsilon_j \varepsilon_k \rangle = \sigma^2 \delta_{jk}$ . Note that  $\sigma$  is the measurement resolution for a static particle, which can be as small as a few nanometers under realistic experimental conditions [9–11].

From Eq. 6 and the Brownian motion property [19]

$$\langle X_{true}(t') X_{true}(t'') \rangle = X_{true}(0)^2 + 2D \min(t', t''),$$

we find that the measured *displacements*  $\Delta_k = X_{k+1} - X_k$  are distributed according to a multivariate Gaussian distribution with

$$\begin{aligned} \langle \Delta_k \rangle &= 0 \\ \langle \Delta_j \Delta_k \rangle &= \begin{cases} 2D\Delta t - 2(2DR\Delta t - \sigma^2) & , \quad j = k \\ 2DR\Delta t - \sigma^2 & , \quad j = k \pm 1 \\ 0 & , \quad \text{otherwise} \end{cases} \end{aligned} \quad (7)$$

$$(8)$$

where the *motion blur coefficient*  $R$  summarizes the effect

of motion blur:

$$R = \frac{1}{\Delta t} \left[ \int_0^{\Delta t} dt' \int_0^{\Delta t} dt'' s(t') s(t'') \min(t', t'') - \int_0^{\Delta t} t' s(t') dt' \right] \quad (9)$$

$$= \frac{1}{\Delta t} \int_0^{\Delta t} S(t) [1 - S(t)] dt. \quad (10)$$

Here,

$$S(t) = \int_0^t s(t') dt' \quad (11)$$

is the fraction of the total illumination occurring before time  $t$  (within the frame), a nondecreasing function ranging from  $S(0) = 0$  to  $S(\Delta t) = 1$ . The integrand in Eq. 10 lies between 0 and 1/4, so we find

$$0 \leq R \leq 1/4,$$

a result that holds for any arbitrary variation of the illumination intensity.  $R$  is zero if and only if the shutter function  $s(t)$  consists of a single vanishingly narrow peak (a delta function) at any point within the frame, in which case there is no motion blur. For uniform illumination,  $s(t) = 1/\Delta t$  and  $R = 1/6$ . Interestingly, these are not the limiting cases for  $R$ ; rather, a double-pulse sequence with a narrow pulse at the start and end of the frame gives the maximum  $R = 1/4$ .

The above considerations lead us to an important observation: although a free Brownian particle moves with uncorrelated displacements, Eq. 8 shows that motion blur and static localization noise induce correlations between *observed* displacements. Localization errors  $\sigma$  induce a negative correlation, which is understood by noting that  $\Delta_{k-1} = X_k - X_{k-1}$  and  $\Delta_k = X_{k+1} - X_k$  depend on the same noise value  $\varepsilon_k$  (cf. Eq. 6) incorporated with opposite sign. Motion blur induces a positive correlation, which is a familiar effect when averaging over the frame is thought of as a low-pass filter acting on the underlying motion [15]. Because of this correlation, measurement results are not independently distributed, and the appropriate framework for statistical inference is time series analysis [20, 21].

### IV. APPROXIMATION OF THE LIKELIHOOD FUNCTION

We can use the distribution of measurement results,  $\Delta_k$ , to construct estimators of the diffusion coefficient  $D$  and static localization noise  $\sigma$ . We choose the maximum-likelihood estimator since it has a simple form and asymptotically optimal properties in many situations [22]. Let  $\mathbf{\Delta}$  be the  $N$ -component column vector of observed displacements  $\Delta_k$ , and let  $\mathbf{\Sigma}$  be the  $N \times N$  covariance matrix defined by Eq. 8. Denoting the nonzero

elements of  $\Sigma$  by

$$\alpha = 2D\Delta t - 2(2DR\Delta t - \sigma^2) \quad (12)$$

$$\beta = 2DR\Delta t - \sigma^2, \quad (13)$$

we can write the (log) likelihood function of the data  $\Delta$  as (dropping an irrelevant constant term)

$$L(\Delta) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \Delta^T \Sigma^{-1} \Delta. \quad (14)$$

From Eq. 14, one can already implement a maximum-likelihood estimator by plugging in the observed displacements  $\Delta$  and numerically maximizing  $L(\Delta)$  considered as a function of  $D$  and  $\sigma$ . However, this procedure requires calculating the determinant and inverse of the  $N \times N$  covariance matrix  $\Sigma$  at each value of  $D$  and  $\sigma$  during the numerical search. For measurements consisting of hundreds or thousands of points, this becomes a numerical inconvenience or impossibility. Furthermore, to facilitate calculation of quantities such as the Fisher information, a simpler closed-form expression for the likelihood function is desired. Closed-form expressions for the determinant and inverse of  $\Sigma$  are available [23], but the results are not particularly simple. Fortunately, a standard approximation leads to significant simplification of the problem. In this approximation [24], we replace  $\Sigma$  with a more convenient matrix defined by setting the corner elements  $\Sigma_{1,N}$  and  $\Sigma_{N,1}$  equal to  $\beta$ . The rows of this new matrix, denoted  $\Sigma_c$ , are cyclic permutations (it is ‘‘circulant’’), which immediately implies that it is diagonal in the frequency (Fourier) domain - a very convenient property indeed. As  $N$  becomes large,  $\Sigma_c$  becomes a very good approximation to  $\Sigma$ ; further details and technical justification are given in Appendix B. Under this approximation, the frequency components

$$\tilde{\Delta}_k = \frac{1}{\sqrt{N}} \sum_{j=1}^N \Delta_j e^{\frac{2\pi i}{N}(j-1)(k-1)} \quad (15)$$

are (complex) Gaussian distributed with

$$\langle \tilde{\Delta}_j \tilde{\Delta}_k^* \rangle = \psi_k \delta_{jk} \quad (16)$$

$$\psi_k = \alpha + 2\beta \cos\left(\frac{2\pi}{N}(k-1)\right). \quad (17)$$

$\psi_k$  are the eigenvalues of  $\Sigma_c$  while  $\tilde{\Delta}_k$  is the  $k$ th component of the discrete Fourier transform of  $\Delta$ ; in other words,  $\Sigma_c$  is diagonal in the discrete Fourier transform (frequency) basis for all values of  $\alpha$  and  $\beta$ , or equivalently  $D$  and  $\sigma$ . We can now write the likelihood function in the frequency domain as

$$L(\Delta) \approx L_p(\Delta) = -\frac{1}{2} \sum_{k=1}^N \left( \log \psi_k + \frac{1}{N} \frac{|\tilde{\Delta}_k|^2}{\psi_k} \right). \quad (18)$$

The choice of subscript ‘‘p’’ is explained in Appendix B. This equation is convenient for numerical maximization,

since it is a simple weighted sum over the power spectrum [or periodogram [20]] of the data  $|\tilde{\Delta}_k|^2$ , found from a discrete Fourier transform of the observed displacements  $\Delta$ . Equation 18 is the main result of this section; it facilitates simple calculation of the maximum likelihood estimate of  $D$  and  $\sigma$  and is simple enough to accommodate thorough analysis.

## V. FISHER INFORMATION IN $D$ AND $\sigma$

From the simple form of  $L_p(\Delta)$  provided by Eq. 18, we can calculate the Fisher information matrix  $\mathbf{I}_{D\sigma}$  in the parameters  $D$  and  $\sigma$ .  $\mathbf{I}_{D\sigma}$  is a measure of the information about  $D$  and  $\sigma$  contained in a sample, and its inverse  $\mathbf{I}_{D\sigma}^{-1}$  gives the Cramer-Rao bound on the covariance matrix of all unbiased estimators [22]. Although the MLE is not necessarily an unbiased estimator, under certain conditions (satisfied here [22]), it asymptotically approaches the Cramer-Rao bound as the number of data points  $N$  becomes large; that is, the MLE of one or several parameters becomes asymptotically normal (multivariate Gaussian distributed) with covariance matrix given by  $\mathbf{I}_{D\sigma}^{-1}$ . In this way, the Fisher information matrix characterizes the optimal performance of all unbiased estimators and also characterizes the asymptotic behavior of the particular estimator derived above, the MLE. In this section, we calculate the Fisher information in  $D$  and  $\sigma$  and use it to find the best possible performance in estimates of  $D$  and  $\sigma$  for realistic particle tracking experiments.

We begin by calculating the Fisher information matrix  $\mathbf{I}_{\alpha\beta}$  in the more convenient parameters  $\alpha$  and  $\beta$ . A simple transformation converts these into the desired result for  $D$  and  $\sigma$ . For a zero-mean multivariate Gaussian distribution, the  $(j, k)$  matrix element of the Fisher information between two parameters,  $\theta_1$  and  $\theta_2$ , is given by

$$(\mathbf{I}_{\theta_1\theta_2})_{jk} = \frac{1}{2} \sum_{l=1}^N \frac{1}{\psi_l^2} \frac{\partial \psi_l}{\partial \theta_j} \frac{\partial \psi_l}{\partial \theta_k} \quad (19)$$

where  $\psi_l$  is the  $l$ th eigenvalue of the covariance matrix. When  $N$  is large, we can approximate the sum over  $l$  as an integral (see Appendix B) to find the matrix elements [25]

$$\begin{aligned} (\mathbf{I}_{\alpha\beta})_{11} &= \frac{N}{2} \frac{1}{2\pi} \int_0^{2\pi} \frac{du}{(\alpha + 2\beta \cos u)^2} = \frac{N}{2} \frac{\alpha}{(\alpha^2 - 4\beta^2)^{3/2}} \\ (\mathbf{I}_{\alpha\beta})_{12} &= \frac{N}{2} \frac{1}{2\pi} \int_0^{2\pi} \frac{2 \cos u \, du}{(\alpha + 2\beta \cos u)^2} = \frac{N}{2} \frac{2\beta}{(\alpha^2 - 4\beta^2)^{3/2}} \\ (\mathbf{I}_{\alpha\beta})_{22} &= \frac{N}{2} \frac{1}{2\pi} \int_0^{2\pi} \frac{4 \cos^2 u \, du}{(\alpha + 2\beta \cos u)^2} \\ &= \frac{N}{2} \left( \frac{1}{\beta^2} + \frac{8\alpha - \alpha^3/\beta^2}{(\alpha^2 - 4\beta^2)^{3/2}} \right). \end{aligned} \quad (20)$$

	$\mathbf{I}_{D\sigma}$	$\mathbf{V}_{D\sigma}$
$R = 0$	$T \begin{pmatrix} \frac{1}{2D^2\Delta t} & \frac{\sigma}{D^2\Delta t^2} \\ \cdot & \frac{3\sigma^2}{D^2\Delta t^3} \end{pmatrix}$	$\frac{1}{T} \begin{pmatrix} 6D^2\Delta t & -\frac{2D^2\Delta t^2}{\sigma} \\ \cdot & \frac{D^2\Delta t^3}{\sigma^2} \end{pmatrix}$
$R = \frac{1}{6}$	$T \begin{pmatrix} \frac{1}{2D^2\Delta t} & \frac{2.2\sigma}{D^2\Delta t^2} \\ \cdot & \frac{18\sigma^2}{D^2\Delta t^3} \end{pmatrix}$	$\frac{1}{T} \begin{pmatrix} 4.3D^2\Delta t & -\frac{0.53D^2\Delta t^2}{\sigma} \\ \cdot & \frac{0.12D^2\Delta t^3}{\sigma^2} \end{pmatrix}$
$R = \frac{1}{4}$	$T \begin{pmatrix} \frac{1}{2D^2\Delta t} & \frac{1}{\sqrt{2D^3\Delta t^3}} \\ \cdot & \frac{2}{\sqrt{D\sigma^2\Delta t^3}} \end{pmatrix}$	$\frac{1}{T} \begin{pmatrix} 2D^2\Delta t & -D\sigma\Delta t \\ \cdot & \sqrt{\frac{D\sigma^2\Delta t^3}{2}} \end{pmatrix}$

TABLE I: Table of Fisher information matrices  $\mathbf{I}_{D\sigma}$  and corresponding covariance matrices  $\mathbf{V}_{D\sigma} = \mathbf{I}_{D\sigma}^{-1}$  in the small noise ( $\epsilon \gg 1$ ) limit.

The Fisher matrix  $\mathbf{I}_{\alpha\beta}$  is converted into  $\mathbf{I}_{D\sigma}$  by the formula [22]

$$\mathbf{I}_{D\sigma} = \mathbf{J}^T \mathbf{I}_{\alpha\beta} \mathbf{J}, \quad \mathbf{J} = \begin{pmatrix} \frac{\partial \alpha}{\partial D} & \frac{\partial \alpha}{\partial \sigma} \\ \frac{\partial \beta}{\partial D} & \frac{\partial \beta}{\partial \sigma} \end{pmatrix}. \quad (21)$$

When  $N$  is large, Equations 20-21 give the Fisher information for any values of diffusion coefficient  $D$  and static localization noise  $\sigma$ , while accounting for motion blur through  $R$ . When the number of samples becomes large enough, the distribution of a joint maximum-likelihood estimate of  $D$  and  $\sigma$  is Gaussian with  $2 \times 2$  covariance matrix  $\mathbf{V}_{D\sigma} = \mathbf{I}_{D\sigma}^{-1}$ . If  $\sigma$  is known and  $D$  is estimated, the asymptotic variance of the MLE is given by  $1/(\mathbf{I}_{D\sigma})_{11}$ ; similarly, if  $D$  is known and  $\sigma$  is estimated, the asymptotic variance of the MLE is  $1/(\mathbf{I}_{D\sigma})_{22}$ . These formulas cover the general cases, but they are rather complicated, so we will break them down into two categories depending on whether the static localization noise  $\sigma$  is large or small compared to the characteristic diffusion length within a frame.

#### A. Large localization noise $\sigma \gg \sqrt{D\Delta t}/2$

First, consider the case where the localization noise is large compared to the diffusion length during the camera frame,  $\sigma \gg \sqrt{D\Delta t}/2$ . In this case, we can approximate the Fisher information matrix and its inverse by expanding in the small parameter  $\epsilon = \sqrt{D\Delta t}/(2\sigma^2)$ . To facilitate comparison of different frame intervals  $\Delta t$ , we suppose that the total measurement time is  $T$  so the number of data points is  $N \approx T/\Delta t$ . Keeping the lowest-order term for each matrix element, we find (using  $\cdot$  to denote

symmetric off-diagonal terms)

$$\begin{aligned} \mathbf{I}_{D\sigma} &\approx \frac{T}{\Delta t} \begin{pmatrix} \frac{\epsilon}{4D^2} & \frac{\epsilon^2}{\sqrt{2D^3\Delta t}} \\ \cdot & \frac{4\epsilon^2}{D\Delta t} \end{pmatrix} \\ &= T \begin{pmatrix} \frac{1}{4\sqrt{2D^3\sigma^2\Delta t}} & \frac{1}{2\sigma^2\sqrt{2D\Delta t}} \\ \cdot & \frac{2}{\sigma^2\Delta t} \end{pmatrix} \end{aligned} \quad (22)$$

and

$$\begin{aligned} \mathbf{V}_{D\sigma} &\approx \frac{\Delta t}{T} \begin{pmatrix} \frac{4D^2}{\epsilon} & -\frac{1}{\epsilon}\sqrt{\frac{D^3\Delta t}{2}} \\ \cdot & \frac{D\Delta t}{4\epsilon^2} \end{pmatrix} \\ &= \frac{1}{T} \begin{pmatrix} 4\sqrt{2D^3\sigma^2\Delta t} & -D\sigma\Delta t \\ \cdot & \frac{\sigma^2\Delta t}{2} \end{pmatrix}. \end{aligned} \quad (23)$$

Note the important feature that the diagonal elements of  $\mathbf{I}_{D\sigma}$  and  $\mathbf{V}_{D\sigma}$  are reciprocals of each other; this indicates that in the large noise limit, the variance of an estimate of  $D$  when  $\sigma$  is known asymptotically approaches the variance of  $D$  when  $\sigma$  is *unknown* but estimated jointly with  $D$ . A similar result holds for estimation of  $\sigma$  when  $D$  is known. In other words, whenever the localization noise is large,  $\sigma \gg \sqrt{D\Delta t}/2$ , with sufficiently long measurement time,  $D$  can be estimated equally well whether or not the value of the localization noise  $\sigma$  is known prior to the experiment. As expected, motion blur has no contribution in this large-noise case.

#### B. Small localization noise $\sigma \ll \sqrt{D\Delta t}/2$

In the case of small localization noise, we expand in the small parameter  $1/\epsilon = \sqrt{2\sigma^2}/(D\Delta t)$ . In this case, motion blur is a significant factor, and we will consider its effects for  $R = 0$  (short-pulse excitation),  $R = 1/6$

	Large noise ( $\epsilon \ll 1$ )	Small noise ( $\epsilon \gg 1$ )		
		$R = 0$	$R = 1/6$	$R = 1/4$
$\frac{\langle \delta D^2 \rangle}{D^2}$ ( $\sigma$ known)	$\frac{4}{T} \sqrt{\frac{2\sigma^2 \Delta t}{D}}$	$\frac{2\Delta t}{T}$	$\frac{2\Delta t}{T}$	$\frac{2\Delta t}{T}$
$\frac{\langle \delta D^2 \rangle}{D^2}$ ( $\sigma$ unknown)	$\frac{4}{T} \sqrt{\frac{2\sigma^2 \Delta t}{D}}$	$\frac{6\Delta t}{T}$	$\frac{4.3\Delta t}{T}$	$\frac{2\Delta t}{T}$
$\frac{\langle \delta \sigma^2 \rangle}{\sigma^2}$ ( $D$ known)	$\frac{\Delta t}{2T}$	$0.33 \frac{D^2 \Delta t^3}{\sigma^4 T}$	$0.06 \frac{D^2 \Delta t^3}{\sigma^4 T}$	$\frac{\Delta t}{T} \sqrt{\frac{D\Delta t}{2\sigma^2}}$
$\frac{\langle \delta \sigma^2 \rangle}{\sigma^2}$ ( $D$ unknown)	$\frac{\Delta t}{2T}$	$\frac{D^2 \Delta t^3}{\sigma^4 T}$	$0.12 \frac{D^2 \Delta t^3}{\sigma^4 T}$	$\frac{\Delta t}{T} \sqrt{\frac{D\Delta t}{2\sigma^2}}$

TABLE II: Table of squared noise-to-signal ratios for a range of experimental scenarios. As an example of how to read the table, consider the first row:  $\langle \delta D^2 \rangle = (\mathbf{V}_{D\sigma})_{11}$  is the variance in an estimate of diffusion coefficient  $D$ , and ( $\sigma$  known) refers to the case that the localization noise  $\sigma$  is known prior to the experiment through a separate measurement. The squared noise-to-signal ratio,  $\langle \delta D^2 \rangle / D^2$  is then recorded for large and small localization noise and, in the latter case, several cases of varying motion blur.

(full-frame averaging) and  $R = 1/4$  (double-pulse excitation). As above, we take  $N \approx T/\Delta t$ . The analysis is complicated by the fact that the lowest order term in  $1/\epsilon$  may be dominated by a higher order (in  $1/\epsilon$ ) quantity multiplied by a large or even divergent  $R$ -dependent term near  $R = 1/4$ . Mathematically, we can avoid divergence problems by plugging in values for  $R$  prior to expanding in  $1/\epsilon$ , but we must keep in mind that the results may only hold in practical cases where  $(R - 1/4)$  is *very* small - smaller even than high-order terms in  $1/\epsilon$ . Of course, the exact expressions of Eqs. 20-21 hold asymptotically in all cases and are simple to evaluate numerically. Results for this small localization noise case are quoted in Table I.

### C. Summary table of results

In Table II, we summarize the asymptotically optimal noise-to-signal ratio in estimating  $D$  and  $\sigma$ , by recording appropriate values from matrices  $\mathbf{I}_{D\sigma}$  and  $\mathbf{V}_{D\sigma}$ . From the table, several useful facts can be gleaned. First, in the common experimental scenario where  $D$  is estimated and  $\sigma$  is small and known prior to the experiment, motion blur has no effect on the asymptotic efficiency of estimating  $D$ . Note that motion blur still affects the statistics of the measurement through  $R$ , but the maximum-likelihood estimator of Eq. 18 fully accounts for the altered statistics, making the final estimate immune to motion blur artifacts. Second, for a typical resolution-calibration experiment with an immobile particle  $D$ , the variance in estimating  $\sigma$  tends to  $\sigma^2/(2N)$  where  $N = T/\Delta t$  is the number of data points. Finally, a surprising result can be seen in the second row of the table: when the diffusion coefficient is estimated and the localization noise is *unknown* (that is, both  $D$  and  $\sigma$  are free parameters in the maximum-likelihood search), the variance in  $D$  and  $\sigma$  are each improved by engineering the experiment to exhibit *maximum* motion-blur,  $R = 1/4$ . For this case, an estimate of  $D$  can in principle be just as accurate as in case where  $\sigma$  is known. It is important to

keep in mind that in many experimental cases, the measurement noise  $\sigma$  may not be independent of the frame time  $\Delta t$  or the motion blur parameter  $R$ . For example, achieving  $R = 0$  or  $R = 1/4$  requires very short pulsing of the excitation intensity. If the localization noise  $\sigma$  is photon-limited, moving to shorter pulses will increase the noise  $\sigma$ , and intermediate values of  $R$  and  $\sigma$  may be necessary to achieve the minimum variance in estimating  $D$ . These cases will not fall neatly into the summary table, but can be treated with the exact results in Eq. 20-21.

## VI. NUMERICAL SIMULATIONS

To investigate the performance of the MLE, we numerically simulated a one-dimensional tracking experiment on a freely-diffusing particle, including the effects of motion blur and localization noise. For each simulation, we generated an  $(N + 1)$ -element set of positions  $\{X_k\}$ , specifying  $D$ ,  $\Delta t$ ,  $\sigma$ , and the shutter function  $s(t)$ , from which a motion blur coefficient  $R$  was calculated according to equation (10). From each data set, we estimated  $D$  and  $\sigma$  using two estimators, the MLE derived above and a variance-weighted fit to the MSD. For the latter case, we calculated  $\rho_n$  from  $\{X_k\}$  according to Eq. (1) and estimated  $D$  and  $\sigma$  by least-squares fitting to Eq. (5). We accounted for the non-uniform statistical weight of each  $\rho_n$  by weighting each point  $\rho_n$  by  $1/C_{nn}$  as calculated in Appendix A. This weighted-MSD estimator roughly corresponds to the procedures of Refs. [12] and [13], updated to incorporate localization noise and motion blur. As in those references, the off-diagonal correlations  $C_{mn}$  were not considered.

In Figure 1 we show the effect of the localization noise on estimates of  $D$ . Similarly, in Figure 2, we show the effect of varied diffusion coefficients on estimates of  $\sigma$ . In the figures, we show the 10th and 90th percentile in estimating one of the parameters, and compare this to the same percentiles for an optimal estimator that achieves the Cramer-Rao bound (shaded regions in both figures). The MLE clearly outperforms the estimator based on

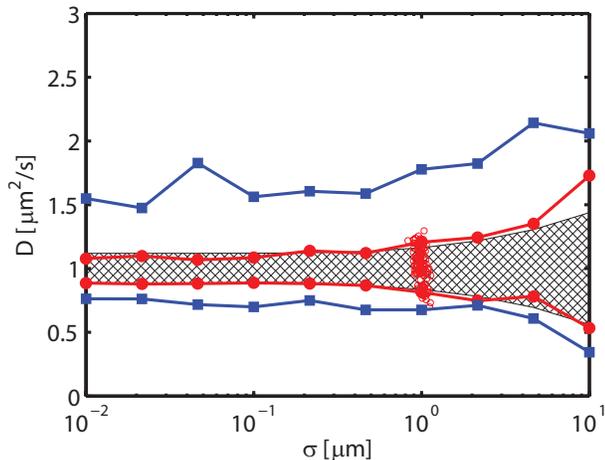


FIG. 1: (Color online) 10th and 90th percentiles in estimating  $D$ . For each simulation,  $D = 1 \mu\text{m}^2/\text{s}$ ,  $\Delta t = 1 \text{ s}$ ,  $N = 500$ , and the shutter function  $s(t)$  was constant over the frame time ( $R = 1/6$ ). For each value of measurement noise  $\sigma$ , 100 individual simulations were performed and the values of  $D$  and  $\sigma$  were estimated using the MLE and MSD. For each estimator, the 10th and 90th percentile of  $D$  estimates is shown (MLE: filled circles, red online; MSD: filled squares, blue online); in other words, at each value of  $\sigma$ , the region between the upper and lower point contains 80 % of the  $D$  estimates. The same percentiles for an optimal unbiased estimator, with covariance matrix equal to  $\mathbf{I}_{D\sigma}^{-1}$ , are also shown (hatched region). A single set of MLE estimator results with  $\sigma = 1 \mu\text{m}$  is shown (open circles, red online). The MLE achieves nearly optimal performance, and significantly outperforms the MSD.

the mean-square displacement, approximating an optimal unbiased estimator in many cases. Note that, because the MLE is normal (Gaussian) and unbiased only in the asymptotic limit  $N \rightarrow \infty$ , some cases in the figure show biased performance with variances below the Cramer-Rao bound.

Finally, in Figure 3, we show the effect of varying the shutter function over the full range of motion blur coefficients  $0 \leq R \leq 1/4$ , in the moderately low noise regime corresponding to  $1/\epsilon \approx 0.14$ . Estimates of  $D$  are slightly improved by increasing  $R$  (the standard deviation decreases by about 10 %) while the sensitivity to  $\sigma$  is greatly improved with increasing  $R$ . In experimental scenarios, this fact may be used to increase sensitivity of calibration experiments on moving particles by tailoring the temporal profile of the illumination intensity. Anecdotally, we find that the maximum likelihood estimator is significantly easier to implement in computer code and executes much faster than the MSD estimator, primarily due to the difficulty of calculating new weight factors  $1/C_{nn}$  at each point in the numerical search.

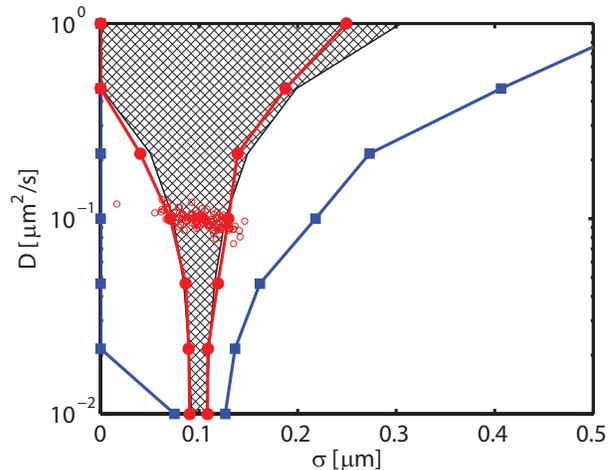


FIG. 2: (Color online) 10th and 90th percentiles in estimating  $\sigma$ . For each simulation,  $\sigma = 0.1 \mu\text{m}$ ,  $\Delta t = 1 \text{ s}$ ,  $N = 500$ . The shutter function  $s(t)$  was constant over the frame time ( $R = 1/6$ ). For each value of  $D$ , 100 individual simulations were performed and the values of  $D$  and  $\sigma$  were estimated using the MLE and MSD. For each estimator, the 10th and 90th percentile of  $D$  estimates is shown (MLE: filled circles, red online; MSD: filled squares, blue); in other words, at each value of  $D$ , the region between the left and right point contains 80 % of the  $\sigma$  estimates. The same percentiles for an optimal unbiased estimator, with covariance matrix equal to  $\mathbf{I}_{D\sigma}^{-1}$ , are also shown (hatched region). A single set of MLE estimator results with  $D = 0.1 \mu\text{m}^2/\text{s}$  is shown (open circles, red online). The MLE achieves nearly optimal performance, and significantly outperforms the MSD.

## VII. CONCLUSIONS

We have derived the full statistics of single particle tracking measurements on freely diffusing particles, properly accounting for the effects of localization noise and motion blur. We defined a motion blur coefficient  $R$  that fully accounts for the latter effect. We derived a computationally simple maximum-likelihood estimator and also derived information-theoretic limits on the measurement sensitivities of separate or joint measurements of the diffusion coefficient  $D$  and localization noise  $\sigma$ . We showed by numerical simulation that the MLE approaches optimality on data sets consisting of a few hundred points, and that the MSD is significantly sub-optimal. In Appendix A, we give updated expressions for the mean, variance, and covariance values of the mean-square displacement when measurement noise and motion blur are considered.

In future work, a linear viscoelastic response can be included in the analysis in order to facilitate proper statistical estimation in particle-tracking microrheology or linear materials [5–8]. For linear materials, the dynamics of the displacement increments will be more complicated but remain multivariate Gaussian distributed, so that results from time series spectral analysis will likely

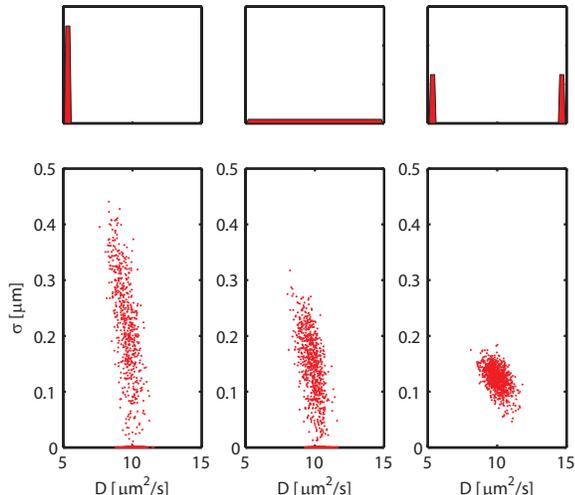


FIG. 3: (Color online) Comparison of maximum-likelihood estimates for different motion blur coefficients  $R$ . For each case,  $N = 1000$  data points were simulated with  $D = 10 \mu\text{m}^2/\text{s}$ ,  $\Delta t = 0.1 \text{ s}$ , and  $\sigma = 0.1 \mu\text{m}$ . Three different shutter functions were used (top row) giving, from left to right,  $R = 0.01$ ,  $R = 0.17$ , and  $R = 0.24$ . A scatter plot of the resulting maximum-likelihood estimates are in the  $D - \sigma$  plane (bottom row). Increasing  $R$  has little effect on  $D$  estimates, but significantly improves the estimate of  $\sigma$ .

be similarly useful in that context. In cases of anomalous diffusion, the displacement statistics will no longer be Gaussian distributed, but observation of a smoothed power spectrum of displacements may nevertheless provide a visual symptom similar to the deviations from linearity of the MSD. Such an effect can be investigated through numerical simulation.

Software for calculating the Fisher information matrix, the covariance of the MSD, and implementing the maximum-likelihood estimator is available from the author on request.

### Acknowledgments

The author gratefully acknowledges helpful discussions with A. Rukhin of the NIST Statistical Engineering Division as well as G. Gallatin, J. A. Liddle, and P. Mathai of the NIST Center for Nanoscale Science and Technology.

### APPENDIX A: STATISTICS OF THE MEAN-SQUARE DISPLACEMENT

In this appendix, we give the mean and covariance  $C_{mn} = \langle \rho_n \rho_m \rangle - \langle \rho_n \rangle \langle \rho_m \rangle$  of the MSD incorporating static localization noise and motion blur. In the main text, we calculated the Gaussian statistics of  $\Delta$ . We can

write  $\rho_n$  as a quadratic form on  $\Delta$ :

$$\rho_n = \Delta^T \mathbf{A}_n \Delta$$

where  $\mathbf{A}_n$  is a symmetric  $N \times N$  matrix found from examining Eq. 1. A standard theorem [26] states that  $\rho_n$  and  $\rho_m$  are independently distributed if and only if  $\mathbf{A}_n \Sigma \mathbf{A}_m = 0$ . This condition is not satisfied in the present case, owing to the strong correlation of  $\rho_n$  and  $\rho_m$ .

To calculate the mean and covariance  $C_{mn}$ , we use the fact that [26]

$$C_{mn} = 2\text{Tr}(\mathbf{A}_m \Sigma \mathbf{A}_n \Sigma)$$

where  $\text{Tr}$  denotes the trace. After a lengthy calculation, one finds that

$$\langle \rho_n \rangle = \text{Tr}(\mathbf{A}_n \Sigma) = n(\alpha + 2\beta) - 2\beta, \quad (\text{A1})$$

which is Eq. 5, and for  $m \geq n$

$$C_{mn} = \frac{2}{K_m K_n} [\alpha^2 Z_{m,n}^{0,0} + 4\alpha\beta Z_{m,n}^{0,1} + 2\beta^2 (Z_{m,n}^{1,1} + Z_{m,n}^{1,-1})] \quad (\text{A2})$$

where

$$Z_{m,n}^{a,b} = \sum_{s=s_-}^{s_+} \{ [K_m + \min(0, s, m-n-s)] \times [n + \min(0, s+a, m-n-s-a)] \times [n + \min(0, s-b, m-n-s+b)] \} \quad (\text{A3})$$

with

$$s_- = \max(m-N, 1-n-a, 1-n+b) \\ s_+ = \min(N-n, m-1-a, m-1+b)$$

$K_m = N + m - 1$  and  $K_n = N + n - 1$ . The use of  $\min$  and  $\max$  functions conveniently summarizes many contingencies that depend on the values of  $N$ ,  $m$ ,  $n$ ,  $a$ , and  $b$ . Once a set has been chosen, the summation can be explicitly computed. The variance and covariance results of Ref. [12], neglecting the effects of static localization errors and motion blur ( $\beta = 0$ ) are recovered from the  $Z_{m,n}^{0,0}$  term.

### APPENDIX B: APPROXIMATION OF THE LIKELIHOOD FUNCTION

In this appendix, a discussion of the approximate likelihood function is presented.

There are two ways to justify the approximate form of the likelihood function in Eq. 18. First, as discussed briefly in the main text, we can view Eq. 18 as arising

from the approximation of  $\Sigma$  by  $\Sigma_c$ . However, it is generally not the case that one can simply alter a few elements of a matrix and expect to retain similar values of, for example, its eigenvalues, determinant, or inverse. Some mathematical justification is therefore required: the ‘‘circulant approximation’’ used here is a well-studied technique in applied mathematics - the standard reference is [24]. Application of results in Ref. [24] to the present context reveal that  $\Sigma \sim \Sigma_c$  and  $\Sigma^{-1} \sim \Sigma_c^{-1}$  as  $N$  grows to infinity, where ‘‘ $\sim$ ’’ denotes the matrix weak norm. This indicates a certain ‘‘average’’ equivalence of these matrices, but does not imply that their individual entries are close together. Further justification for the approximation comes from that observation that the eigenvalues of  $\Sigma$  and  $\Sigma_c$  are *asymptotically absolutely equally distributed*, which means that in the limit  $N \rightarrow \infty$ , the eigenvalues (and all functions of them, including the determinant) are equal [24].

While these results suggest a good matrix approximation, they do not strictly imply that  $L(\Delta) \approx L_p(\Delta)$ . To see this, Eq. 18 should be viewed a second way, as an application of Whittle’s approximation [21, 27] in which the likelihood function is approximated by its ‘‘principal part,’’ given here by:

$$L_p(\Delta) = \frac{N}{2} \left[ \frac{1}{2\pi} \int_0^{2\pi} \log 2\pi f(\lambda) d\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\lambda, \Delta)}{f(\lambda)} d\lambda \right], \quad (\text{B1})$$

where  $f(\lambda) = (1/2\pi)(\alpha + 2\beta \cos \lambda)$  is the spectral density of the (exact) covariance matrix  $\Sigma$  and

$$I_N(\lambda, \Delta) = \frac{1}{2\pi N} \left| \sum_{k=1}^N \Delta_k e^{-i\lambda k} \right|^2.$$

When  $N$  is large, we can approximate the integral over  $\lambda$  as a Riemann sum over the convenient grid of points  $\lambda_s = 2\pi s/N$ ,  $s = 0 \cdots N - 1$ , in which case Eq. B1 becomes Eq. 18 after some manipulation of indices. Whittle’s approximation is well-characterized [21], in particular it is known that the approximate likelihood converges to the true likelihood in probability and is asymptotically normally distributed according to the inverse of the Fisher information matrix given by Eqs. 21.

- 
- [1] H. Geerts, M. De Brabander, R. Nuydens, S. Geuens, M. Moeremans, J. De Mey, and P. Hollenbeck, *Biophys. J.* **52**, 775 (1987).
  - [2] J. Gelles, B. Schnapp, and M. Sheetz, *Nature* **351**, 450 (1988).
  - [3] M. J. Saxton and K. Jacobson, *Annu. Rev. Biophys. Biomol. Struct.* **26**, 373 (1997).
  - [4] J. Crocker and D. Grier, *J. Colloid Interface Sci.* **179**, 298 (1996).
  - [5] T. G. Mason, K. Ganesan, J. H. Van Zanten, D. Wirtz, and S. C. Kuo, *Physical Review Letters* **79**, 3282 (1997).
  - [6] T. Waigh, *Rep. Prog. Phys.* **68**, 685 (2005).
  - [7] D. Weihs, T. Mason, and M. Teitell, *Biophys. J.* **91**, 4296 (2006).
  - [8] D. Wirtz, *Annual Review of Biophysics* **38**, 301 (2009).
  - [9] M. Cheezum, W. Walker, and W. Guilford, *Biophys. J.* **81**, 2378 (2001).
  - [10] R. Thompson, D. Larson, and W. Webb, *Biophys. J.* **82**, 2775 (2002).
  - [11] R. Ober, S. Ram, and E. Ward, *Biophys. J.* **86**, 1185 (2004).
  - [12] H. Qian, M. P. Sheetz, and E. L. Elson, *Biophys. J.* **60**, 910 (1991).
  - [13] M. Saxton, *Biophys. J.* **72**, 1744 (1997).
  - [14] D. Martin, M. Forstner, and J. Käs, *Biophys. J.* **83**, 2109 (2002).
  - [15] T. Savin and P. S. Doyle, *Biophys. J.* **88**, 623 (2005).
  - [16] W. Wong and K. Halvorsen, *Optics Express* **14**, 12517 (2006).
  - [17] D. Montiel, H. Cang, and H. Yang, *J. Phys. Chem. B* **110**, 19763 (2006).
  - [18] M. D. McMahon, A. J. Berglund, P. Carmichael, J. J. McClelland, and J. A. Little, *ACS Nano* **3**, 609 (2009).
  - [19] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer-Verlag, 1985), 2nd ed.
  - [20] P. Brockwell and R. Davis, *Time series: theory and methods* (Springer Verlag, 2009).
  - [21] K. Dzhaparidze, *Parameter estimation and hypothesis testing in spectral analysis of stationary time series* (Springer-Verlag, 1986).
  - [22] S. Zacks, *The Theory of Statistical Inference* (John Wiley & Sons, 1971).
  - [23] C. F. Fischer and R. A. Usmani, *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **6**, 127 (1969).
  - [24] R. M. Gray, *Toeplitz And Circulant Matrices: A Review (Foundations and Trends(R) in Communications and Information Theory)* (Now Publishers Inc., Hanover, MA, USA, 2006), ISBN 1933019239.
  - [25] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic press, San Diego, 1980), 6th ed.
  - [26] S. Searle, *Linear models* (Wiley New York, 1971).
  - [27] P. Whittle, *Arkiv för matematik* **2**, 423 (1953).