

## Optimal estimation of diffusion coefficients from single-particle trajectories

**Vestergaard, Christian L.; Blainey, Paul C.; Flyvbjerg, Henrik**

*Published in:*

Physical Review E. Statistical, Nonlinear, and Soft Matter Physics

*Link to article, DOI:*

[10.1103/PhysRevE.89.022726](https://doi.org/10.1103/PhysRevE.89.022726)

*Publication date:*

2014

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Vestergaard, C. L., Blainey, P. C., & Flyvbjerg, H. (2014). Optimal estimation of diffusion coefficients from single-particle trajectories. *Physical Review E. Statistical, Nonlinear, and Soft Matter Physics*, 89(2), 022726. DOI: 10.1103/PhysRevE.89.022726

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# Optimal estimation of diffusion coefficients from single-particle trajectories

Christian L. Vestergaard,<sup>1,\*</sup> Paul C. Blainey,<sup>2,†</sup> and Henrik Flyvbjerg<sup>1,‡</sup>

<sup>1</sup>*Department of Micro- and Nanotechnology, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark*

<sup>2</sup>*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 4 January 2013; revised manuscript received 27 November 2013; published 28 February 2014)

How does one optimally determine the diffusion coefficient of a diffusing particle from a single-time-lapse recorded trajectory of the particle? We answer this question with an explicit, unbiased, and practically optimal *covariance-based estimator* (CVE). This estimator is regression-free and is far superior to commonly used methods based on measured mean squared displacements. In experimentally relevant parameter ranges, it also outperforms the analytically intractable and computationally more demanding *maximum likelihood estimator* (MLE). For the case of diffusion on a flexible and fluctuating substrate, the CVE is biased by substrate motion. However, given some long time series and a substrate under some tension, an extended MLE can separate particle diffusion on the substrate from substrate motion in the laboratory frame. This provides benchmarks that allow removal of bias caused by substrate fluctuations in CVE. The resulting unbiased CVE is optimal also for short time series on a fluctuating substrate. We have applied our estimators to human 8-oxoguanine DNA glycolase proteins diffusing on flow-stretched DNA, a fluctuating substrate, and found that diffusion coefficients are severely overestimated if substrate fluctuations are not accounted for.

DOI: [10.1103/PhysRevE.89.022726](https://doi.org/10.1103/PhysRevE.89.022726)

PACS number(s): 87.16.A—, 87.16.dj, 87.80.Nj, 87.16.Wd

## I. INTRODUCTION

Diffusion is ubiquitous in biology, and many cellular processes rely on diffusion as a passive means of transport. Quantitative knowledge of the diffusion coefficient is paramount for the precise understanding of these processes. Recent developments in fluorescent labels have made it possible to track diffusion of single molecules, e.g., proteins on biopolymers such as DNA [1–3] or microtubules [4–6], on surfaces [7], in lipid membranes [8–10], and inside cells [11–13] with time-lapse photography. Data mostly consist of relatively short time series with considerable experimental localization error. This makes it a challenge to determine diffusion coefficients. This challenge is even higher when individuality of diffusion coefficients is a concern, since one then cannot average over multiple trajectories of different molecules to reduce statistical error.

The standard approach relies on Einstein's classical result for the *mean squared displacement* (MSD) of a particle undergoing free diffusion. It estimates the diffusion coefficient by fitting a straight line to experimental values of the MSD [14]. This gives the right value on average, but that does not mean that it is a good way to estimate the diffusion coefficient. It is accurate (=unbiased), but when the MSD is calculated from a single trajectory (or a few), its precision depends on the number of MSD values fitted to [15], and for good *signal-to-noise ratio* (SNR) the precision actually *decreases* when *more* points are used in the fit [8,16]. This result is counterintuitive if one's intuition is based only on experience with uncorrelated data. It is a fact, nevertheless. We demonstrate and explain it below.

The *optimized least-squares fit* (OLSF [17]) improves on the MSD estimator by including the optimal number of points in the fit [15]. *Generalized least squares* (GLS) fully accounts for correlations between experimental estimates of MSDs. These estimators have been proven optimal when linearly dependent on the parameters to be estimated [18,19]. These proofs are not valid here, however, because both these estimators depend nonlinearly on the parameters of interest—the diffusion coefficient  $D$  and the variance  $\sigma^2$  of localization errors—and we show below that OLSF and GLS are suboptimal.

The complicated dependence of the MSDs on data makes it very difficult to derive a *maximum likelihood estimator* (MLE) based on the MSDs. However, a MLE of  $D$  and  $\sigma^2$  was recently derived based on the much simpler statistics of the single-time-lapse displacements [19,20]. This estimator is known to be optimal, asymptotically, in the limit of infinitely long time series. There it is unbiased and as precise as is possible: Its variance is as small as the Cramér-Rao bound [18].

For some systems, however, it is difficult or impossible to obtain long time series. Most experiments with individual biological molecules are limited by fluorophore lifetimes; proteins diffuse out of the field of view in confocal microscopy; proteins that diffuse on biopolymers detach. Consequently, data predominantly consist of short time series, for which optimality of MLE is far from guaranteed. In this range we find that a simple *covariance-based estimator* (CVE) is better, since its variance practically reaches the Cramér-Rao bound and it is unbiased, whereas the more complicated MLE, OLSF, and GLS are biased. This CVE is an explicit function of data, i.e., it is regression-free and is thus orders of magnitude faster than maximum-likelihood and least-squares estimators.

For diffusion on many cellular structures, on DNA, and in lipid membranes, the recorded movement contains a contribution from thermal motion of the substrate. If the time scale of these thermal fluctuations is much shorter than the time lapse of the recording, they will contribute to the movement as a random, uncorrelated error on positions, i.e., a white noise

\*cvestergaard@gmail.com

†Current address: MIT Department of Biological Engineering and Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA; pblainey@broadinstitute.org

‡henrik.flyvbjerg@nanotech.dtu.dk

on positions. Hence, it can be absorbed in the total variance,  $\sigma^2$ , of localization errors, which receives also a contribution from true errors on localization due to photonic shot noise in images of the tracked object.

However, substrate motion needs to be modeled to some degree beyond white noise if the longest time scale of substrate motion is comparable to the time lapse or longer. In a taut region of a substrate, the amplitudes of thermal motion are so small that they can be modeled with a linear model. We take this as our definition of “taut.” On a finite substrate, the spectrum of this thermal dynamics is discrete. When the spectrum of characteristic times is well separated compared to the time lapse—which holds for smaller, tauter substrates—only one or a few slowest modes of substrate motion are resolved by our time-lapse recording, while all higher modes contribute as a white-noise error on positions. We derive a MLE that explicitly accounts for substrate fluctuations in this manner and is optimal for long time series. We also derive an unbiased CVE for diffusion on a fluctuating substrate which can be used to obtain unbiased and optimal estimates of diffusion coefficients for short time series, where the MLE fails.

We estimate diffusion coefficients of fluorescently marked human 8-oxoguanine DNA glycolase (hOGG1) repair proteins on DNA from time-lapse measurements. The data have previously been analyzed using MSD-based methods [21]. We measure diffusion coefficients in the range  $0.1\text{--}0.5\ \mu\text{m}^2/\text{s}$ . We show that the DNA fluctuations induce a bias in the estimates of diffusion coefficients of up to  $0.25\ \mu\text{m}^2/\text{s}$ ; i.e., we may overestimate diffusion coefficients by several times their actual values if the fluctuations are not taken into account. The increased resolution that our method offers enables us to see a clear negative correlation between a protein’s residence time on the DNA molecule and its diffusion coefficient. This correlation was hidden when the data were analyzed with MSD-based methods. The negative correlation is explained by a two-state kinetics of hOGG1’s diffusion on  $\lambda$  DNA [22].

Section II gives protocols for how to analyze a time series of a single particle diffusing on a substrate or in a medium which is fluctuating or at rest. The protocols are kept short, with details given in ensuing sections. Thus, Sec. II is a road map to practical application of this paper. Not all applications need all of the paper, so the road map provides short-cuts.

Section III reviews the statistics of time series of a freely diffusing particle recorded in the presence of localization error and motion blur and discusses why common MSD-based estimators are suboptimal. We describe a rigorous statistical test of whether a recorded trajectory describes free diffusion. This test is based on the experimentally determined periodogram of the recorded time series, which is compared to the theoretical spectrum of a freely diffusing particle. We derive the unbiased CVE of the diffusion coefficient, for both the case in which the variance of localization errors also must be estimated from the same time series and the case where the variance of localization errors is known *a priori* or has been determined independently. We test the performance of the CVE on Monte Carlo generated data and compare it to other near-optimal estimators based on the MSDs, to the MLE, and to the ultimate limit on the precision of any unbiased estimator, the Cramér-Rao bound.

Section IV develops a linear model for the motion of a fluctuating substrate, in particular the motion of a taut unstretchable polymer, here DNA. We derive the statistics of diffusion on such a fluctuating substrate. Section V builds on the theory of Sec. IV to extend MLE and CVE to diffusion on a fluctuating substrate.

Section VI specializes these estimators to the case of diffusion on a taut, unstretchable polymer. In this case the transversal motion of the substrate is directly accessible experimentally. We test the estimators on Monte Carlo generated data, compare their performance, and compare their precision to the Cramér-Rao bound.

Section VII applies the estimators developed in Sec. VI to time-lapse recorded data of hOGG1 repair proteins diffusing on fluctuating, flow-stretched  $\lambda$  DNA and shows that our theory for diffusion on a fluctuating substrate accurately describes the data. Section VIII concludes.

Appendix A and Supplemental Materials (Tables I, II, and III) summarize our acronyms and notation [23]. Many technical details have been relegated to nine, mostly short, appendixes. Appendix B discusses application of the GLS estimator to our estimation task. Appendix C gives the statistical properties of covariances and of the CVE for the case of free diffusion. Appendix D shows how application of discrete sine transformation or discrete Fourier transformation decorrelates data. Appendix E describes the MLE for the case of free diffusion. Appendix F describes the effect of finite time-lapse recording and motion blur on a power spectrum. Appendix G details estimators to be used on a fluctuating substrate. Appendix H details the variances of these estimators. Appendix I details Monte Carlo simulations of data used to test estimators. Appendix J describes how to apply two statistical tests of agreement between theory and data: Pearson’s  $\chi^2$  goodness-of-fit test and the  $\chi^2$  test for variance.

## II. HOW TO ANALYZE A TIME SERIES OF POSITIONS OF A SINGLE PARTICLE THAT DIFFUSES IN A MEDIUM WHICH IS AT REST OR FLUCTUATES

This section gives a “road map” to the practical use of this article, since few readers will need all its sections to analyze a given problem. We give a protocol for how to analyze a time-lapse recorded time series of positions of a particle diffusing on a substrate or in a medium that is at rest (Sec. II A) or fluctuates (Sec. II B). In Sec. II C we give a protocol for how to analyze a time series in the specific case of a particle diffusing on a taut, unstretchable, and fluctuating polymer, such as DNA. In this case one is helped by the fact that information about the transversal motion of the substrate is directly observable in the time series.

### A. Diffusion on a substrate or in a medium at rest

A time series  $(x_n)_{n=0}^N$  of measured positions of a particle undergoing free diffusion in a medium or on a substrate at rest should be analyzed as follows.

- (i) Calculate the set of one-time-lapse displacements  $(\Delta x_n)_{n=1}^N$  from the time series  $(x_n)_{n=0}^N$  as  $\Delta x_n = x_n - x_{n-1}$ .
- (ii) Estimate the particle’s diffusion coefficient  $D$  (and variance  $\sigma^2$  of the localization error if it is not known *a priori*)

from  $(\Delta x_n)_{n=1}^N$  using the CVE as described in Sec. III C. The values of  $D$  and  $\sigma^2$  determine the SNR; see Eq. (6) below. If the SNR is smaller than one, and a small bias is of no concern, then  $D$  (and  $\sigma^2$ ) is more efficiently estimated using MLE as described in Sec. III D.

(iii) Test whether the measured data  $(\Delta x_n)_{n=1}^N$  agree with the initial hypothesis that the particle undergoes free diffusion in a medium or on a substrate at rest by comparing the periodogram of  $(\Delta x_n)_{n=1}^N$  to its expected value, the spectrum, as described in Sec. III B.

If the data and the theory do not agree, an alternative hypothesis should be considered, e.g., that the particle does not undergo free diffusion, or that the substrate or medium fluctuates. In the latter case, see the next section, Sec. II B.

### B. Diffusion on a fluctuating substrate

An experimental time series of laboratory coordinates of a particle diffusing on a taut, fluctuating substrate can be analyzed as follows.

(i) If it is reasonable to assume that substrate fluctuations are so fast that all they do is contribute to localization errors as a white noise, then follow the procedure described in Sec. II A to test if this hypothesis is true. If it is, use estimates of diffusion coefficients as obtained following this procedure. If the hypothesis is rejected by the data, proceed to step (ii).

(ii) From the time series  $(x_n)_{n=0}^N$ , calculate the set of one-time-lapse displacements  $(\Delta x_n)_{n=1}^N$  as  $\Delta x_n = x_n - x_{n-1}$ , and from  $(\Delta x_n)_{n=1}^N$  calculate the set of periodogram values  $(\hat{P}_{\Delta x, f})_{f=\Delta f}^{f_{\text{Nyq}}}$  as described in Sec. V A.

(iii) Estimate diffusion coefficient  $D$ , variance  $\sigma^2$  of the localization error, and parameters  $\phi$  characterizing the substrate's motion using MLE as described in Sec. V A. If the substrate's motion has been characterized independently, either by direct measurement or by averaging multiple ML estimates as described in Sec. V B, use the unbiased CVE given in Sec. V B to estimate  $D$  and  $\sigma^2$ .

(iv) Compare the measured periodogram  $(\hat{P}_{\Delta x, f})_{f=\Delta f}^{f_{\text{Nyq}}}$  to its expected values, the power spectrum, as described in Sec. V C, to test whether data support the hypothesis that the particle undergoes free diffusion on the fluctuating substrate.

### C. Diffusion on taut, fluctuating DNA

An experimental time series of laboratory coordinates of a particle diffusing on a taut polymer, such as DNA, reveals the DNA's transversal motion at the location of the particle. It does this directly, uncontaminated by the particle's diffusive motion, because the particle only diffuses in the DNA's longitudinal direction; we qualify this statement below. Let  $x$  and  $y$  denote the laboratory coordinates in the longitudinal and the transversal direction, respectively, of the particle on the taut DNA strand. Then our protocol for how to analyze such a time series  $(x_n, y_n)_{n=0}^N$  is the following.

(i) If the DNA is (or might be) so taut that the longest relaxation time of its fluctuations is much shorter than the time lapse  $\Delta t$  of measurements, then the protocol in Sec. II A above should be used to analyze the  $x$  coordinate of the time series. If the DNA indeed is this taut, that is confirmed by this protocol.

(ii) If the DNA is not that taut, then calculate  $(\Delta x_n)_{n=1}^N$ ,  $(\hat{P}_{\Delta x, f})_{f=\Delta f}^{f_{\text{Nyq}}}$ , and  $(\hat{P}_{y, f})_{f=\Delta f}^{f_{\text{Nyq}}}$  from  $(x_n, y_n)_{n=0}^N$  as described in Sec. VI.

(iii) If the DNA's motion has not already been characterized by independent measurements, it must be characterized using information in the time series  $(x_n, y_n)_{n=0}^N$ . To this end, use the MLE given in Sec. VI A on time series with  $N \geq 50$ . This MLE optimally estimates the diffusion coefficient  $D$  of the particle, the variance  $\sigma^2$  of the localization error, and parameters  $\phi$  that characterize the DNA's motion.

(iv) If the parameters describing the DNA's motion have been characterized independently by direct measurement or by averaging multiple ML estimates as described in Sec. V B, then estimate  $D$  and  $\sigma^2$  optimally by using the bias-subtracted CVE given in Sec. V B.

(v) Test the hypothesis that the recorded particle undergoes free diffusion along the fluctuating DNA molecule by comparing the measured periodograms,  $(\hat{P}_{\Delta x, f})_{f=\Delta f}^{f_{\text{Nyq}}}$  and  $(\hat{P}_{y, f})_{f=\Delta f}^{f_{\text{Nyq}}}$ , to their theoretically expected values, the power spectra, as described in Sec. VI C.

## III. ESTIMATION OF THE DIFFUSION COEFFICIENT OF PARTICLE DIFFUSING ON SUBSTRATE AT REST

We here review the statistics of time-lapse recorded laboratory coordinates of a particle undergoing free diffusion in a medium (or on a substrate) at rest in the laboratory. We review the statistics of single-time-lapse displacements of the particle. We also review the statistics of the particle's experimentally determined MSDs, and we discuss why common estimators based on the MSDs are suboptimal (Sec. III A). We give a rigorous procedure for testing the hypothesis that the particle undergoes free diffusion in a medium (or on a substrate) at rest (Sec. III B). We derive a CVE that is based on the measured covariances of single-time-lapse displacements. CVE is simple, explicit, unbiased, and practically optimal as an estimator of the particle's diffusion coefficient  $D$  and of the variance  $\sigma^2$  of the localization errors on positions (Sec. III C). We give a fast algorithm for the MLE of  $D$  and  $\sigma^2$ . We also give an expression for the Cramér-Rao bound, which limits the precision of any unbiased estimator of  $D$  and  $\sigma^2$  (Sec. III D). Finally, we compare the performance of estimators of diffusion coefficients to each other and to the Cramér-Rao bound. This is done using synthetic data generated by Monte Carlo simulations (Sec. III E).

### A. Statistics of recorded time series

#### 1. Single-time-lapse displacements

Consider a particle diffusing in  $d$  dimensions. Its trajectory is recorded using time-lapse photography. The particle's positions are determined in each frame using a localization algorithm. For a fluorescent particle, e.g., this algorithm would be a super-resolution microscopy method [24]. The result is a time series of the particle positions,  $\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_N$ , measured with constant time lapse  $\Delta t$ . From this time series, we form the time series of single-time-lapse displacements  $\Delta \mathbf{r}_1, \Delta \mathbf{r}_2, \dots, \Delta \mathbf{r}_N$  defined by  $\Delta \mathbf{r}_n = \mathbf{r}_n - \mathbf{r}_{n-1}$ . The series



of displacements is stationary, as opposed to the series of positions. Being stationary ensures that averages formed from the series will converge to their expected values as  $N \rightarrow \infty$ .

Each Cartesian coordinate of a particle diffusing in an isotropic and homogeneous medium is independent. With no loss of generality, we therefore only consider diffusion in one dimension below.

Let the function  $\zeta(t)$  describe the state of the camera shutter during a time lapse  $\Delta t$ .  $\zeta(t) = 0$  means closed shutter, while  $\zeta(t) > 0$  means open shutter. The scale of  $\zeta(t)$  is fixed by the normalization condition  $\int_0^{\Delta t} \zeta(t) dt = 1$ . The measured  $x$  coordinate of the particle's position at time  $t_n = n\Delta t$  is then given by [20]

$$x_n = \int_{t_n - \Delta t}^{t_n} x_{\text{true}}(t_n - t) \zeta(t) dt + \sigma \xi_n. \quad (1)$$

Here  $x_{\text{true}}$  is the true position of the particle, and the time integral describes the motion blur that results from finite exposure time. The second term describes localization errors associated with the time-averaged position given by the first term. The stochastic variable  $\xi_n$  is a normalized, discrete Gaussian white noise: It has zero mean, unit variance, zero autocovariance. The real, positive parameter  $\sigma$  is the standard deviation of the localization error. This error is the sum of all localization errors in effect, including substrate motion, if relevant and of such high frequency that it contributes in this place.

From Eq. (1) the covariance matrix of the measured displacements  $\Delta x_1, \dots, \Delta x_N$  of the diffusing particle is found to be [20]

$$\langle (\Delta x_n)^2 \rangle = 2D\Delta t + 2(\sigma^2 - 2DR\Delta t), \quad (2)$$

$$\langle \Delta x_n \Delta x_{n+1} \rangle = -(\sigma^2 - 2DR\Delta t), \quad (3)$$

$$\langle \Delta x_m \Delta x_n \rangle = 0, \quad \text{for } |n - m| > 1. \quad (4)$$

Here  $\langle \dots \rangle$  denotes the expected value, and the parameter  $R$  is the motion blur coefficient defined by

$$R = \frac{1}{\Delta t} \int_0^{\Delta t} S(t)[1 - S(t)] dt, \quad (5)$$

where  $S(t) = \int_0^t \zeta(t') dt'$  [20]. The motion blur coefficient can take values in the interval  $[0, 1/4]$ . In practice the camera shutter is usually kept open for the full duration of the time lapse, which makes  $R = 1/6$ . This is done in order to maximize the number of photons recorded in an image, thereby minimizing the localization error [1,2,4–13,21,25,26].

Diffusion is a scale-free process, but the root-mean-squared displacement taking place during a time lapse,  $\sqrt{2D\Delta t}$ , defines a length, the so-called *diffusion length* associated with the time lapse. This length is the amplitude of our signal, the signal being the displacement recorded for each time lapse. The standard error on this signal is  $\sqrt{2}\sigma$ , where  $\sigma$  is the standard deviation of the localization error as defined above. For given length  $N$  of a time series and motion blur coefficient  $R$ , the

performance of an estimator is then characterized by the SNR,

$$\text{SNR} \equiv \frac{\sqrt{2D\Delta t}}{\sqrt{2}\sigma} = \frac{\sqrt{D\Delta t}}{\sigma}. \quad (6)$$

Typically,  $1 < \text{SNR} < 20$  [4,5,26–28]. In the following we compare estimators by comparing how they perform as function of  $N$  and SNR. We also compare the variances of estimators to the ultimate lower bound on the variance of any unbiased estimator, the Cramér-Rao bound [18]. These variances are only marginally affected by the value of the motion blur coefficient  $R$  [25]. So we only discuss the performance of estimators for the case of maximal exposure time [ $\zeta(t) = 1/\Delta t$ ,  $R = 1/6$ ] in the following, but note that their performances are similar for other values of  $R$ .

## 2. Mean squared displacements

Einstein argued in 1905 that the mean length of time-lapse recorded displacements of a particle suspended in a static liquid are not proportional to the elapsed time, and hence one cannot deduce the particle's mean velocity from it. Instead, MSD is proportional to time, and the constant of proportionality gives the particle's diffusion coefficient,  $D$  [29]. Adding the effects of localization error and motion blur, we find that the expected value of the measured squared displacement of a diffusing particle is

$$\langle d(t)^2 \rangle = \langle [x(t) - x(0)]^2 \rangle = 2Dt + 2(\sigma^2 - 2DR\Delta t). \quad (7)$$

Since 1905, diffusion coefficients have been determined from trajectories of individual particles with Eq. (7) or variants thereof [14,15]. This is OK when data are so rich that precision is not an issue. When precision *is* an issue, the MSD is a poor-to-miserable estimator. Its precision depends strongly on the extent to which one accounts for the fact that its values,  $\overline{d_n^2}$ , at different times  $t_n = n\Delta t$  are highly correlated [Fig. 1(a)] [15], when they all are estimated from the same (or a few) time series of positions  $x_0, x_1, \dots, x_N$  using

$$\overline{d_n^2} = \frac{1}{N - n + 1} \sum_{i=0}^{N-n} (x_{i+n} - x_i)^2. \quad (8)$$

This fact is not common knowledge [1,2,4–14,21,26,30]; neither is the fact that for a good SNR an *ordinary least squares* (OLS) or *weighted least squares* (WLS) fit of a straight line to  $\overline{d_n^2}$  plotted against  $n$  for  $n = 1, 2, \dots, n_{\text{max}} \leq N$  yields an estimate for  $D$  that becomes *worse* when *more* points ( $n, \overline{d_n^2}$ ) are included in the fit [Fig. 1(b)] [8,15]. One might intuit that more data points supply more information. Such intuition is based on experience with statistically independent data points. The values of the MSD are not statistically independent, however. They are so strongly correlated that when more values are included in a fit, the added noise may exceed the added signal.

This counterintuitive result can be understood by looking at the extreme case where the localization error is zero. In this case the least-squares estimator based on the single data point  $(1, \overline{d_1^2})$ ,  $\hat{D} = \overline{d_1^2}/(2\Delta t)$ , is equal to the MLE of  $D$  and is optimal; it is unbiased and its precision reaches the Cramér-Rao bound. It is optimal because  $\overline{d_1^2}$  is a *sufficient statistic* for  $D$ ; all information available in the time series about the value

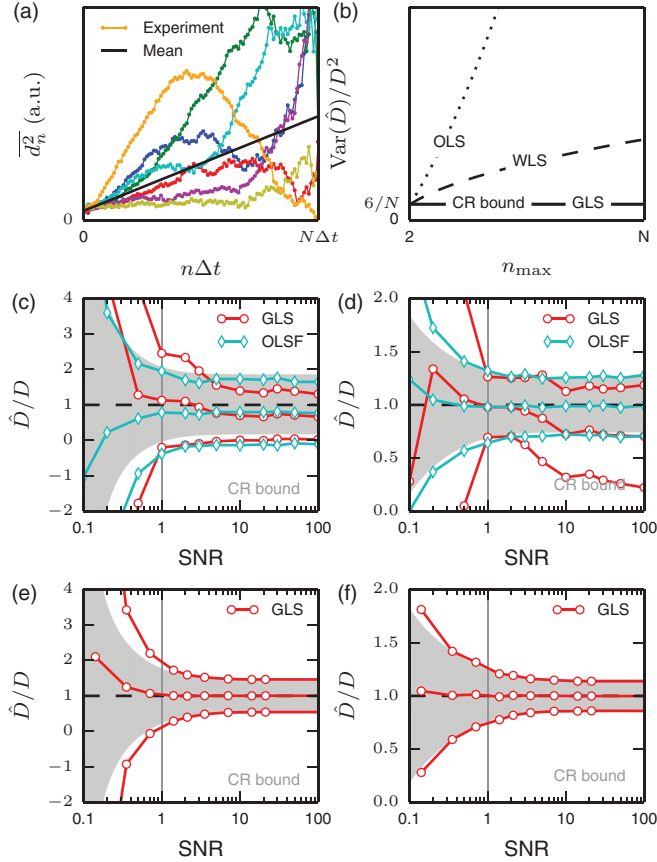


FIG. 1. (Color online) Performance of MSD-based estimators. (a) Experimental MSDs,  $\overline{d_1^2}, \overline{d_2^2}, \dots, \overline{d_N^2}$ , calculated from simulated Brownian motion trajectories. Each color represents one trajectory. The straight black line shows their expected values  $\langle d(t_n)^2 \rangle$ . The values of  $\overline{d_n^2}$  for different  $n$  values are highly correlated because they are estimated from the same trajectory. The variance of  $\overline{d_n^2}$  increases with  $n$ . (b) Variances of MSD-based estimates of the diffusion coefficient  $D$  as a function of the number of MSD points used in the fit. Variances of ordinary (OLS), weighted (WLS), and generalized (GLS) least-squares fits are compared to their information-theoretical lower limit, the Cramér-Rao bound (CR bound), for SNR = 10. (c)–(f) Mean plus/minus standard error of optimized least-squares fit (OLSF) [15] and GLS estimates  $\hat{D}$  in units of the true value  $D$ . Edge of shaded gray region is the Cramér-Rao bound. (c) For unknown noise amplitude and  $N = 10$ , the OLSF and GLS estimators are biased and do not reach the Cramér-Rao bound in practice. (d) For unknown noise amplitude and  $N = 100$ , the OLSF almost reaches the Cramér-Rao bound and is practically unbiased, while the GLS estimator is significantly biased and does not reach the Cramér-Rao bound. (e), (f) For known noise amplitude and (e)  $N = 10$  or (f)  $N = 100$ , the GLS estimator is practically optimal.

of  $D$  is contained in  $\overline{d_1^2}$ . Thus, including more points gives *no* new information about  $D$ . If we furthermore treat these points as independent of the first point  $(1, \overline{d_1^2})$  of the measured MSD (as in WLS) or, even worse, also give them the same weight as given to this first point  $(1, \overline{d_1^2})$  (OLS), we actually *decrease* the precision of our estimate of  $D$  by including more points.

We extend the example to the case where the localization error amplitude is unknown but small (it needs not be very

small; for SNR > 2 the arguments below hold). Here the set  $(\overline{d_1^2}, \overline{d_2^2})$  is *almost* a sufficient statistic for  $D$  (see also the discussion of the relationship between the CVE and the MSDs below). Thus, additional points  $(n, \overline{d_n^2})$  contain almost no new information about  $D$ , and if we include them in a less than optimal way, i.e., using OLS or WLS, then the precision of our estimate of  $D$  decreases when more points are included.

This underscores the danger of being naive about statistics: In situations where Eq. (7) is very true, common estimators based on Eq. (8) nevertheless give poor estimates for  $D$ ; though accurate, they are not precise. Thus, an excellent particle-tracking experiment may appear mediocre if common practice is followed and the analysis is based naively on MSD.

A recent method aims to find the optimal number of MSD points to include in an unweighted OLSF [15,19]. Alternatively, one may include the full covariance matrix of the  $\overline{d_n^2}$  and fit to all  $\overline{d_n^2}$  in a GLS fit (Appendix B). Both of these estimators depend nonlinearly on the estimated parameters. Consequently, theoretical results for their optimality, which were derived for linear dependence on parameters [15,18], do not apply here [Figs. 1(c)–1(f)].

## B. Testing whether a recorded trajectory describes free diffusion

Equation (4) is as important as Eqs. (2) and (3), since it states that the signature of free diffusion is  $\overline{\Delta x_n \Delta x_m} \approx 0$  for  $|n - m| > 1$ . Specifically, in order for a time series to be consistent with free diffusion, these covariance estimates must scatter about zero [Figs. 2(a) and 2(d)] with variances that depend on  $D$  and  $\sigma^2$  as (Appendix C)

$$\text{var}(\overline{\Delta x_n \Delta x_m}) = \frac{\alpha^2 + 4\alpha\beta + 6\beta^2}{N - |n - m|} - \frac{2\beta^2}{(N - |n - m|)^2}. \quad (9)$$

Here  $\alpha = 2D\Delta t$  and  $\beta = \sigma^2 - 2D\Delta t R$  are, respectively, the squared diffusion length corresponding to the time lapse and the variance of the localization error at finite exposure time. The estimated covariances are correlated and they are Gaussian distributed only in the limit of long time series. This makes them unsuitable for statistical testing of whether a given time series of displacements is consistent with free diffusion or not, just like the MSDs are. Instead, one should compare the periodogram based on the *discrete sine transform* [DST, Eq. (11) below] of the measured displacements with its expected values (the power spectrum) for the case of a freely diffusing particle. Figures 2(b) and 2(e) show such periodograms and their expected values. The comparison is made easy by the fact that periodogram values corresponding to different frequencies are statistically independent of each other.

### 1. Spectrum of displacements for finite $N$

For quantitative statistical testing of whether a recorded trajectory describes free diffusion, the MSDs and the covariance function of the single-time-lapse displacements are both impractical due to their complicated distributions and high inherent correlations. Instead, one should compare the periodogram  $\check{P}_k = 2(\widetilde{\Delta x_k})^2 / [(N + 1)\Delta t]$  to its expected

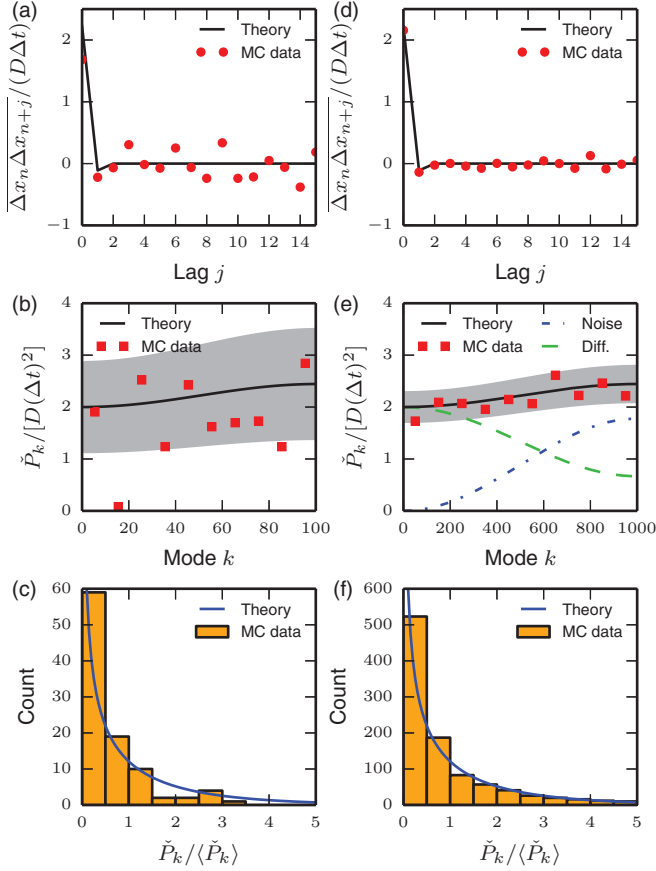


FIG. 2. (Color online) Statistics of single-time-lapse displacements of a particle diffusing in a medium or on a substrate at rest. (a)–(c) The statistics of a time series of length  $N = 100$ . (d)–(f) The same statistics for  $N = 1000$ . (a),(d) Covariance of single-time-lapse displacements calculated from a Monte Carlo (MC) simulated time series shown in units of  $D\Delta t$  and compared to their expected values. The theoretical covariance shows an isolated large, positive value at zero lag, the signature of free diffusion. A small negative value at unit lag is the signature of localization errors; motion blur makes it less negative and may even change its sign. Consequently, the value at unit lag may be difficult to resolve on a background of stochastic noise. (b),(e) Periodogram of single-time-lapse Monte Carlo generated displacements in units of  $D(\Delta t)^2$  compared to their expected values, the power spectrum. The power spectrum is the sum of a term due to localization errors (Noise) and a term due to diffusive motion of the particle (Diff.). Shown values are block averages over 10 (b) and 100 (e) periodogram values each [31]. This averaging facilitates comparison by eye with the expected values, the theoretical curve. The gray area marks the 68% confidence interval (CI) for the blocked values, which is equivalent to the mean plus/minus one standard deviation for Gaussian distributed data. On average, 2/3 of the points should fall in the gray area. The theoretical curve is not a fit to the data, but the ultimate truth, which is known in Monte Carlo simulations. For real experimental data, fitting is necessary before comparing and should be done to with CVE or MLE as described in the text. Block averages are shown since the raw periodogram values have a SNR of  $1/\sqrt{2}$ , which makes visual comparison impractical. (c),(f) The normalized periodogram values  $\tilde{\epsilon}_k = \tilde{P}_k / \langle \tilde{P}_k \rangle$  follow a  $\Gamma$  distribution with shape and scale parameters of  $1/2$  and  $2$ , respectively. In (a),(b),(d),(e) SNR = 1.5 and  $R = 1/6$  (maximally open shutter).

value,

$$P_k = \langle \tilde{P}_k \rangle = 2D(\Delta t)^2 + 2[\sigma^2 \Delta t - 2DR(\Delta t)^2] \left( 1 - \cos \frac{\pi k}{N+1} \right), \quad (10)$$

which is called  $\psi_k(D, \sigma^2)$  in [20]. Here  $\tilde{\Delta x}_k$  is the DST of  $(\Delta x_1, \Delta x_2, \dots, \Delta x_N)$ ,

$$\tilde{\Delta x}_k = \Delta t \sum_{n=1}^N \sin \left( \frac{\pi kn}{N+1} \right) \Delta x_n, \quad (11)$$

where the multiplicative term  $\Delta t$  ensures that the DST has the same physical dimensions as its continuum analog, the *Fourier sine transform*. The DST of  $(\Delta x_n)_{n=1}^N$  is efficiently calculated using a *fast sine transform* such as MATLAB or SciPy's `dst` [32] or as described in [33,34]. The  $N$  values  $(\tilde{P}_k)_{k=1}^N$  are statistically independent (Appendix D) and are thus suitable for statistical testing, e.g., by simple visual inspection, by comparing the measured periodogram to its expected values [Figs. 2(b) and 2(e)], and by comparing the distribution of normalized values  $\tilde{\epsilon}_k = \tilde{P}_k / \langle \tilde{P}_k \rangle$  to a  $\Gamma$  distribution with shape and scale parameter values, respectively,  $1/2$  and  $2$  [Figs. 2(c) and 2(f)]. The latter comparison can be made quantitative with Pearson's  $\chi^2$  goodness-of-fit test (Appendix J), taking into account that two parameters,  $D$  and  $\sigma^2$ , were fitted (one parameter if  $\sigma^2$  was determined independently).

Comparison between the experimentally measured periodogram and the theoretically predicted power spectrum is difficult for time series of short or intermediate length due to poor statistics of these [Fig. 2(b)]. Since longer time series [Fig. 2(e)] may not be available, comparison can be facilitated by averaging over multiple time series of particles that show the same diffusion coefficients and localization errors. Averaging over a set of ten such time series of length  $N = 100$  gives a dispersion which is the same (to order  $1/N$ ) as that of a single time series of 1000 points [Fig. 2(e)].

The statistical test presented here is based on independent statistics and assumes that we have exact knowledge of their distribution based on the theoretical results for the measured time series [Eq. (2)–(4)]; i.e., it is a *parametric test*. It is thus more reliable and more powerful than previously proposed tests based on nonparametric estimates of the autocorrelations and their variances or marginal distributions of correlated single-time-lapse displacements [35].

## 2. Spectrum of displacements in the large- $N$ limit

For long time series,  $\tilde{P}_k$ , the periodogram obtained with the DST, approaches the periodogram  $\hat{P}_{\Delta x, f} = |\hat{\Delta x}_f|^2 \Delta f$  obtained with the discrete Fourier transform (DFT). Here  $\Delta f = 1/(N\Delta t)$ , and  $\hat{\Delta x}_f$  is the DFT of  $(\Delta x_1, \Delta x_2, \dots, \Delta x_N)$ ,

$$\hat{\Delta x}_f = \Delta t \sum_{n=1}^N e^{-i2\pi f t_n} \Delta x_n, \quad (12)$$

where  $f = f_k = k\Delta f$ . It is efficiently calculated using *fast Fourier transformation* (FFT) such as MATLAB or SciPy's `fft` [32] or as in [33,34]. In this case one can compare the periodogram  $\hat{P}_{\Delta x, f}$  to its expected value, the

power spectrum,

$$P_{\Delta x, f}(D, \sigma^2) = 2D(\Delta t)^2 + 2[\sigma^2 \Delta t - 2DR(\Delta t)^2][1 - \cos(2\pi f \Delta t)]. \quad (13)$$

The normalized periodogram values  $\hat{\epsilon}_{\Delta x, f} = \hat{P}_{\Delta x, f} / P_{\Delta x, f}(D, \sigma^2)$  should be exponentially distributed on the non-negative real numbers with expected value equal to one. In this case also, the comparison is made quantitative with Pearson's  $\chi^2$  goodness-of-fit test (Appendix J), taking into account that two parameters,  $D$  and  $\sigma^2$ , were fitted (one parameter if  $\sigma^2$  was determined independently).

### C. The covariance-based estimator

Equations (2) and (3) tell us how to construct simple, unbiased estimators for  $D$  and  $\sigma^2$  from a single recorded trajectory. We replace the expected values in Eqs. (2) and (3) with unbiased estimators of these and solve for  $D$  and  $\sigma^2$ . This gives unbiased CVEs of  $D$  and  $\sigma^2$ ,

$$\hat{D} = \frac{(\overline{\Delta x_n})^2}{2\Delta t} + \frac{\overline{\Delta x_n \Delta x_{n+1}}}{\Delta t}, \quad (14)$$

$$\hat{\sigma}^2 = \overline{R(\Delta x_n)^2} + (2R - 1)\overline{\Delta x_n \Delta x_{n+1}}. \quad (15)$$

where  $\overline{\cdots}$  denotes averages over the time series  $\Delta x_1, \dots, \Delta x_N$ . For a particle diffusing in  $d$  dimensions, estimates are obtained by averaging over the estimates obtained from the  $d$  one-dimensional time series of individual coordinates. In this manner the standard errors on estimates are reduced by a factor  $1/\sqrt{d}$ .

If the value of the parameter  $\sigma^2$  is known *a priori*, or if it has been estimated in advance, as described in Sec. III C 3, then the CVE of the diffusion coefficient reduces to

$$\hat{D} = \frac{(\overline{\Delta x_n})^2 - 2\sigma^2}{2(1 - 2R)\Delta t}. \quad (16)$$

The CVEs of  $D$  [Eqs. (14) and (16)] are guaranteed to be unbiased and are practically optimal, as long as the SNR is larger than one (Sec. III E).

#### 1. Variance of the CVE

Equation (14) results in the following variance of its estimate for  $D$  (Appendix C), to second order in  $1/N$ ,

$$\text{var}(\hat{D}) = D^2 \left[ \frac{6 + 4\varepsilon + 2\varepsilon^2}{N} + \frac{4(1 + \varepsilon)^2}{N^2} \right], \quad (17)$$

where  $\varepsilon = \sigma^2 / (D\Delta t) - 2R$ .

Equation (16), on the other hand, results in the following variance of its estimate for  $\hat{D}$ ,

$$\text{var}(\hat{D}) = \frac{D^2(2 + 4\varepsilon + 3\varepsilon^2)}{N(1 - 2R)^2} + \frac{\text{var}(\hat{\sigma}^2)}{(1 - 2R)^2(\Delta t)^2}. \quad (18)$$

Here the second term describes the contribution from a stochastic error on our known value for  $\sigma^2$ . This contribution is proportional to the variance  $\text{var}(\hat{\sigma}^2)$  of that error, when, as assumed here, the latter error is uncorrelated with the error on  $(\Delta x_n)^2$ .

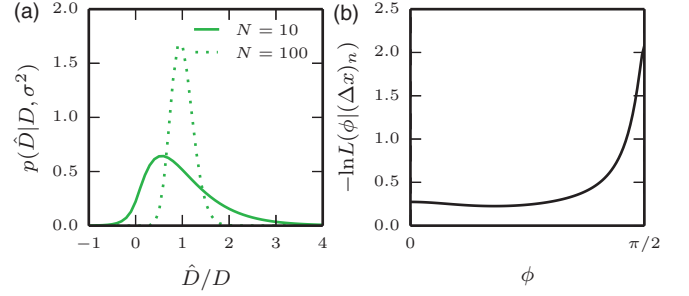


FIG. 3. (Color online) Distribution of the CVE of the diffusion coefficient  $D$  (a) and minus-log-likelihood landscape (b) for diffusion on a substrate at rest. (a) Distribution of CVEs  $\hat{D}$ , here measured in units of the true value  $D$ . The distribution approaches a Gaussian as  $N$  is increased. (b) Minus-log-likelihood as a function of the unitless variable  $\phi$  defined as  $\phi = \text{arccot}(\sqrt{D\Delta t}/\sigma)$  (Appendix E) for a Monte Carlo generated time series of length  $N = 100$ . SNR = 2,  $R = 1/6$  (maximally open shutter), and  $\sigma^2$  is not known *a priori*.

For importance weighting of estimates—e.g., when calculating the weighted mean of estimates from time series of different lengths—the length  $N$  of a time series should be used as weight when possible, since it is known exactly. The inverse variance of an estimate should be avoided as bias due to correlations between the estimated parameters and the estimated variances; see [36] for an example of such correlations and resulting bias of estimate.

#### 2. Higher moments and probability density of the CVE

Higher moments of the distribution  $p(\hat{D}|D, \sigma^2)$  of estimates  $\hat{D}$  [37] resulting from the CVE for given true values  $D$  and  $\sigma^2$  are found by differentiating its characteristic function  $\tilde{p}(\omega|D, \sigma^2)$ —see Appendix C, Eq. (C10)—with respect to  $\omega$  and setting  $\omega$  to zero,

$$\langle \hat{D}^k \rangle^{\text{CVE}} = i^k \frac{\partial^k \tilde{p}(\omega|D, \sigma^2)}{\partial \omega^k} \Big|_{\omega=0}. \quad (19)$$

The distribution itself is equal to the inverse Fourier transform of the characteristic function,

$$p(\hat{D}|D, \sigma^2) = \int_{-\infty}^{\infty} \frac{e^{i\hat{D}\omega}}{2\pi} \tilde{p}(\omega|D, \sigma^2) d\omega. \quad (20)$$

Thus, the distribution of estimates by CVE, and hence exact confidence intervals for estimates, can be found by numerical Fourier transformation of  $\tilde{p}$  [Fig. 3(a)]. This can be done effectively using a FFT algorithm with corrections for end contributions as described in [33,34].

#### 3. Independent determination of the variance of localization errors

The position of a diffusing fluorescent particle is estimated by fitting to its measured point-spread function (PSF). When this is done as described in [24], one can estimate the variance  $\sigma^2$  of localization errors directly from the fitting procedure.

If it is not possible to estimate  $\sigma^2$  directly from the fitting procedure, an alternative approach may be used: If estimates of  $\sigma^2$  from many time series are indistinguishable up to stochastic



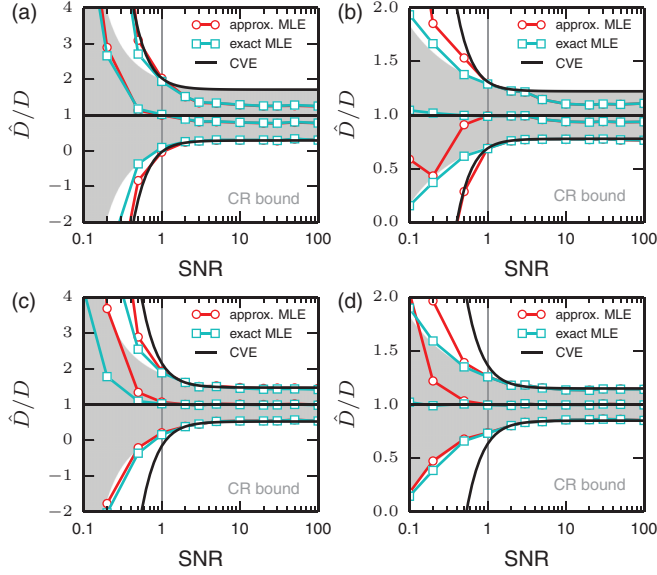


FIG. 4. (Color online) Quality of various estimators for the diffusion coefficient. Mean plus/minus standard error of estimators given in units of the true value  $D$  as function of SNR. Estimates  $\hat{D}$  obtained with approximate MLE, MLE, and CVE applied to 1000 Monte Carlo generated time series of positions of a freely diffusing particle recorded with motion blur and noise on positions. Edges of the shaded gray regions represent the information-theoretical lower limit on the standard error, the Cramér-Rao bound. (a),(c) Time series of length  $N = 10$ . (b),(d) Time series of length  $N = 100$ . All simulations were performed with shutter time equal to time lapse ( $R = 1/6$ ). (a),(b) For unknown amplitude of localization errors, the MLEs reach and even surpass the Cramér-Rao bound for high SNR, where they are biased. The CVE is unbiased and attains the Cramér-Rao bound for  $\text{SNR} > 1$ . (c),(d) For known amplitude of localization errors, both MLEs and CVE are unbiased and attain the Cramér-Rao bound for  $\text{SNR} > 1$ .

errors, then these estimates can be averaged to obtain a more precise estimate of the noise amplitude. Diffusion coefficients can then be estimated again, using the average noise variance estimate as a fixed parameter, since this averaged quantity covaries little with the individual time series. In this manner, essentially all information in a time series is used to estimate its diffusion coefficient  $D$ . This reduces the standard error on estimates of the diffusion coefficient by a factor of up to 1.8 in the limit of high SNR and the absence of motion blur ( $R = 0$ ). The standard error of the CVE is reduced by a factor  $\approx 1.5$  for maximally open shutter ( $R = 1/6$ ) [compare Figs. 4(c) and 4(d) to Figs. 4(a) and 4(b)].

#### 4. Relationship between the CVE and the MSDs

The MSD in Eq. (7) suggests how to construct a maximally simple MSD-based estimator of the diffusion coefficient  $D$  from  $\overline{d_1^2}$  and  $\overline{d_2^2}$ :  $\hat{D}_{\text{msd}} = (\overline{d_2^2} - \overline{d_1^2})/(2\Delta t)$ . On average, this maximally simple MSD-based estimator is exactly equivalent to the CVE of  $D$ : They have the same expected value,  $D$ . This follows from the relations  $\langle d(2\Delta t)^2 \rangle = 2\langle (\Delta x_n)^2 \rangle +$

$2\langle \Delta x_n \Delta x_{n+1} \rangle$  and  $\langle d(\Delta t)^2 \rangle = \langle (\Delta x_n)^2 \rangle$ . Thus,

$$\langle \hat{D}_{\text{msd}} \rangle = \frac{\langle (\Delta x_n)^2 \rangle + 2\langle \Delta x_n \Delta x_{n+1} \rangle}{2\Delta t} = \langle \hat{D}_{\text{cve}} \rangle. \quad (21)$$

CVE is the more precise of the two, however. The estimates of  $\langle d(n\Delta t)^2 \rangle$  and  $\langle \Delta x_m \Delta x_n \rangle$ ,  $\overline{d_n^2}$  and  $\overline{\Delta x_n \Delta x_m}$ , respectively, do not use the information present in the displacements  $(\Delta x_1, \Delta x_2, \dots, \Delta x_N)$  in the same way.  $\overline{d_n^2}$  places less weight on the end displacements  $\Delta x_1$  and  $\Delta x_N$ , while CVE weights all displacements equally (Appendix C). This makes CVE of  $D$  more precise than the MSD-based method.

#### D. Maximum likelihood estimation and the Cramér-Rao bound

Using the power spectral density in Eq. (10), one can easily construct a MLE of  $D$  and  $\sigma^2$  using the periodogram based on the DST of  $(\Delta x_1, \Delta x_2, \dots, \Delta x_N)$  [Eq. (11)] [19]. The MLE is usually found by maximizing the logarithm of the likelihood function  $\mathcal{L}(\theta)$  with respect to the two parameters  $\theta = (D, \sigma^2)$  as detailed in [20]. Appendix E gives a fast algorithm for ML estimation of  $(D, \sigma^2)$  which uses the scale-invariance of free diffusion to reduce the dimensionality of the optimization problem from two to one dimensions. Besides increasing the speed of the MLE algorithm, the reduction from two to one dimensions enables us to inspect the likelihood landscape visually as a function of a single parameter and thus ensure that the algorithm has converged to the global minimum [Fig. 3(b)].

##### 1. Variance and the Cramér-Rao bound

The variance of the MLE is to first order in  $1/N$  given by the Cramér-Rao bound  $\mathcal{I}(\theta)^{-1}$ , where  $\mathcal{I}(\theta)$  is the Fisher information matrix [18]. Its entries are

$$\mathcal{I}(\theta)_{ij} = -\left\langle \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j} \right\rangle = \sum_{k=1}^N \frac{1}{P_k(\theta)^2} \frac{\partial P_k(\theta)}{\partial \theta_i} \frac{\partial P_k(\theta)}{\partial \theta_j}, \quad (22)$$

with

$$\frac{\partial P_k(\theta)}{\partial D} = 2(\Delta t)^2 \left[ 1 - 2R \left( 1 - \cos \frac{\pi k}{N+1} \right) \right] \quad (23)$$

and

$$\frac{\partial P_k(\theta)}{\partial \sigma^2} = 2\Delta t \left( 1 - \cos \frac{\pi k}{N+1} \right). \quad (24)$$

The Cramér-Rao bound sets a lower limit on the variance of any unbiased estimator [18]. An unbiased estimator that reaches the Cramér-Rao bound is thus considered optimal.

For the case in which the amplitude  $\sigma$  of the localization error has been estimated independently, and only  $D$  is estimated from the time series, the Fisher information matrix is a  $1 \times 1$  matrix, i.e., a scalar. This scalar equals

$$\mathcal{I}(D) = \sum_{k=1}^N \left[ \frac{1}{P_k(D, \sigma^2)} \frac{\partial P_k(D, \sigma^2)}{\partial D} \right]^2. \quad (25)$$

The Cramér-Rao bound on the variance of unbiased estimators of  $D$  is given by  $1/\mathcal{I}(D)$ . Since MLE is known to saturate the Cramér-Rao bound asymptotically as  $N \rightarrow \infty$ , this bound,  $1/\mathcal{I}(\hat{D})$ , serves as an estimate (accurate to order  $1/N$ ) of the variance of the MLE of  $D$ .

### E. Monte Carlo simulations

We tested the CVE on Monte Carlo generated time series. We compared its performance to two near-optimal MSD-based estimators—OLSF [15] and the GLS estimator (Appendix B)—and to the “gold standard,” MLE (Sec. III D and Appendix E) [19], as well as to an approximate MLE [20]. The precision of each estimator was compared to the Cramér-Rao bound, the ultimate bound on the precision of any unbiased estimator (Sec. III D).

#### 1. How to simulate

Monte Carlo simulation was done as follows. The DST diagonalizes the covariance matrix given in Eqs. (2)–(4) (Appendix D). Thus, it transforms the  $N$  displacements  $\Delta x_n$  into  $N$  independent, normally distributed “transformed” displacements  $\Delta x_k$  with mean zero and variances given by Eq. (10). Such transformed displacements were generated from Gaussian random numbers and transformed back, using the inverse DST given by a matrix  $\mathbf{M}$  with entries

$$M_{i,j} = \frac{2}{(N+1)\Delta t} \sin\left(\frac{\pi i j}{N+1}\right). \quad (26)$$

This gives synthetic data sets with the statistics of time series of time-lapse recorded freely diffusing particles [Eqs. (2)–(4)]. We tested our estimators on these data sets.

#### 2. Simulation results

As shown in Sec. III A 2, the MSD-based estimators are suboptimal. The CVE and the MLEs practically reach the Cramér-Rao bound, and the MLEs even surpass it for high SNR [Figs. 4(a) and 4(b)]; this is possible because the MLEs are biased [Figs. 4(a) and 4(b)], which means that the total error of the MLE can be, and is here, smaller than that of any unbiased estimator. This extra precision comes at a cost, however: a systematic error in the estimate, a *bias* (Appendix E). This bias complicates statistical analysis of estimates from multiple time series, since averages and other statistics do not converge to their true values.

The CVE, in contrast, is unbiased by construction, and, as Fig. 4 shows, it practically reaches the Cramér-Rao bound as long as the SNR is larger than one. In experiments, typically  $1 < \text{SNR} < 20$ . There, CVE is the best estimator of diffusion coefficients.

Monte Carlo simulations confirm that a more precise estimate of  $D$  can be obtained by using *a priori* knowledge of  $\sigma^2$  [Figs. 4(c) and 4(d)]. We furthermore see that the MLE is practically unbiased when  $\sigma^2$  is known *a priori* and that both the MLE and the CVE reach the Cramér-Rao lower bound for  $\text{SNR} > 1$  [Figs. 4(c) and 4(d)].

#### 3. Conclusion

In conclusion, we find that the CVE is to be preferred in practice since (i) it is unbiased and practically optimal in experimentally relevant situations. This is not the case for the MLEs, the OLSF, and the GLS estimators, which are biased. (ii) The CVE is given by a simple analytical expression; it is thus regression-free and is orders of magnitude faster than the MLEs, the OLSF, and the GLS estimator. (iii) The exact

variance and distribution of the CVE are known. For the MLE, this is only the case asymptotically as  $N \rightarrow \infty$ .

## IV. STATISTICS OF DIFFUSION ON A FLUCTUATING SUBSTRATE

In this section we derive the statistics of recorded time series of a particle diffusing on a fluctuating, taut substrate. A fixed point on such a substrate moves in the laboratory frame. We first derive the statistics of this motion (Sec. IV A). We then derive the covariance of single-time-lapse displacements of a particle diffusing on the substrate (Sec. IV B). Next, we calculate the power spectrum of the single-time-lapse displacements (Sec. IV C). Finally, we consider the specific case of a particle diffusing on a taut, incompressible fiber or polymer, such as DNA, and derive the statistics (covariances and power spectra) of this motion (Sec. IV D).

### A. Statistics of a fluctuating substrate

Consider a fluctuating elastic substrate. The substrate is  $d$ -dimensional and of finite extent ( $L_x, L_y, L_z$ ) but much larger than the characteristic diffusion length,  $\sqrt{2Dt_N}$ , of a particle diffusing in or on it. We assume that the substrate’s motion is massively overdamped, i.e., inertia plays no role. Let  $\mathbf{s}$  denote the coordinates of a physical point in the substrate in the reference frame of the substrate. Let  $\mathbf{r}(\mathbf{s}, t) = [x(\mathbf{s}, t), y(\mathbf{s}, t), z(\mathbf{s}, t)]^T$  denote the coordinates of this point in the reference frame of the laboratory at time  $t$ . Thermal fluctuations or other causes drive the substrate out of mechanical equilibrium. We assume that its excursions from equilibrium are so small that the restoring forces are well described as Hookean. Then the dynamics of the substrate is linear and a superposition of independent spatial eigenmodes.

From the assumptions above and the equipartition theorem follows that the amplitude of each spatial eigenmode has the dynamics of an overdamped harmonic oscillator at finite temperature with the same temperature applying to all eigenmodes. Thus, without knowing the specific form of these spatial eigenmodes, we know the character of the dynamics of a given physical point of the substrate: Its distance to its equilibrium position evolves as the sum of spatial eigenmodes evaluated at that point of the substrate. Since the spectrum of relaxation times of these eigenmodes is discrete for a finite, taut substrate, and faster modes have smaller amplitudes, two cases are so simple that we can account for the motion of the substrate without knowing its dynamics in detail.

(i) If the longest relaxation time of the eigenmodes is somewhat shorter than the time lapse  $\Delta t$ , substrate motion appears uncorrelated in our recordings, which means it contributes as a white-noise localization error. The CVE derived in Sec. III can then be used to obtain optimal estimates of the diffusion coefficients of particles diffusing in or on the substrate.

(ii) If the longest relaxation time of the eigenmodes is longer than the time lapse  $\Delta t$ , but all other relaxation times are somewhat shorter than  $\Delta t$ —which is not at all unrealistic; see examples of spectra of relaxation times in [22]—then substrate motion appears in our recordings as a white-noise localization error plus the thermal motion of one overdamped harmonic oscillator. The parameter values characterizing this

overdamped harmonic oscillator may vary across the substrate, but locally it may be determined experimentally.

We consider the motion in the laboratory of a point  $\mathbf{s}$  of the substrate. Let  $x(\mathbf{s}, t)$  denote the first laboratory-coordinate of this point. In case (ii) above, the autocovariance of this coordinate is

$$\rho_x(t - t' | \mathbf{s}) = \text{cov}(x(\mathbf{s}, t), x(\mathbf{s}, t')) = K_x(\mathbf{s}) \tau_x(\mathbf{s}) e^{-|t-t'|/\tau_x(\mathbf{s})}, \quad (27)$$

where  $K_x$  describes the amplitude of the motion and  $\tau_x(\mathbf{s})$  its correlation time. Here  $\tau_x(\mathbf{s}) = (2\pi f_{c,x})^{-1}$ , where  $f_{c,x}$  is the so-called *corner frequency* of the *Lorentzian* power spectrum of this motion. Both  $K_x$  and  $f_{c,x}$  depend on  $\mathbf{s}$  in general.

Now let  $P_f^{(\text{sub}, \text{cont})}$  denote the power spectrum of a continuously recorded variable that describes the substrate's motion in the laboratory, and let an additional subscript indicate which variable. Then the power spectrum of  $x(\mathbf{s}, t)$  is

$$\begin{aligned} P_{x,f}^{(\text{sub}, \text{cont})} &= \int_0^{t_N} \int_0^{t_N} e^{-i2\pi f(t-t')} \rho_x(t - t' | \mathbf{s}) dt' dt \\ &= \frac{K_x(\mathbf{s})}{2\pi^2 [f_{c,x}(\mathbf{s})^2 + f^2]} \\ &\quad + \frac{K_x(\mathbf{s})(1 - e^{-2\pi f_{c,x}(\mathbf{s})t_N}) [f_{c,x}(\mathbf{s})^2 - f^2]}{4\pi^3 f_{c,x}(\mathbf{s}) t_N [f_{c,x}(\mathbf{s})^2 + f^2]}, \end{aligned} \quad (28)$$

where  $f = f_k = k\Delta f$  with  $k$  integer. Here, the last term describes the effect of finite time of measurement,  $t_N = N\Delta t$ , on the power spectrum. It is included only to provide a quantitative indicator of how long one must measure to avoid its contribution. For  $2\pi f_{c,x} t_N = t_N/\tau_x(\mathbf{s}) \gg 1$ , which is always the case for situations of practical interest, this second term in Eq. (28) can be ignored. For two- and three-dimensional substrates, similar expressions result for each dimension, provided that a separation of spatial variables is possible.

### B. Covariance of displacements of a particle diffusing on a fluctuating substrate

Observed from the laboratory frame of reference, the trajectory of a particle moving on a fluctuating substrate can be described by the particle's trajectory on the substrate,  $\mathbf{s}(t)$ , and the consecutive positions in the laboratory of the points visited on the substrate at the time they are visited, i.e.,

$$\mathbf{r}(t) = \mathbf{r}(\mathbf{s}(t), t). \quad (29)$$

As in the previous section (Sec. III), we look at the particle's single-time-lapse displacements,  $\Delta \mathbf{r}_n = \mathbf{r}_n - \mathbf{r}_{n-1}$ . In practice, the camera's shutter is kept open during the whole time lapse to maximize the number of photons recorded [1, 2, 4–13, 21, 26] and thus maximize the information content of recorded time series [25]. The recorded positions are consequently motion blurred, and a single-time-lapse displacement in, e.g., the  $x$  direction, is modeled as

$$\begin{aligned} \Delta x_n &= \frac{1}{\Delta t} \int_{t_{n-1}}^{t_n} x(\mathbf{s}(t), t) dt + \sigma \xi_{x,n} \\ &\quad - \frac{1}{\Delta t} \int_{t_{n-2}}^{t_{n-1}} x(\mathbf{s}(t), t) dt + \sigma \xi_{x,n-1}, \end{aligned} \quad (30)$$

where  $(\xi_{x,n})_{n=1}^N$  is a normalized Gaussian white noise, i.e., independent  $N(0, 1)$  variables.

We use that the dimensions of the substrate are large compared to the diffusion length of the particle and Taylor expand around the average position,  $\bar{\mathbf{s}}$ , of the particle on the substrate. Neglecting terms of order  $\sqrt{2D\Delta t}/L_x \ll 1$  or higher, we thus have that

$$\Delta x_n \simeq \frac{\partial x_n(\bar{\mathbf{s}})}{\partial s_x} \Delta s_{x,n} + \Delta x_n(\bar{\mathbf{s}}) + \sigma \Delta \xi_{x,n}, \quad (31)$$

where  $\Delta s_{x,n} = \frac{1}{\Delta t} \int_{t_{n-1}}^{t_n} [s_x(t) - s_x(t - \Delta t)] dt$  is the diffusive displacement of the particle along the  $x$  axis of the reference frame of the substrate, and  $\Delta x_n(\bar{\mathbf{s}}) = \frac{1}{\Delta t} \int_{t_{n-1}}^{t_n} [x(\bar{\mathbf{s}}, t) - x(\bar{\mathbf{s}}, t - \Delta t)] dt$  is the displacement in reference frame of the laboratory of the point  $\bar{\mathbf{s}}$  on the substrate during one time lapse.  $\partial x_n(\bar{\mathbf{s}})/\partial s_x = \frac{1}{\Delta t} \int_{t_{n-1}}^{t_n} \partial x(\bar{\mathbf{s}}, t)/\partial s_x dt$  describes the degree of stretching of the substrate during the  $n$ th time lapse.

Since the substrate motion is uncorrelated with the particle's diffusion, the autocovariance of  $\Delta x_n$  is just the sum of the autocovariances of the three terms in Eq. (31). We furthermore use that  $\partial x/\partial s_x$  and  $\Delta s_{x,n}$  are uncorrelated except for terms of order  $2D\Delta t/L_x^2$ . We then get

$$\langle (\Delta x_n^2) \rangle = 4\zeta(\bar{\mathbf{s}})^2 D\Delta t/3 + 2\sigma^2 + C_{\Delta x}(0|\bar{\mathbf{s}}), \quad (32)$$

$$\langle \Delta x_n \Delta x_{n\pm 1} \rangle = -[\sigma^2 - 2\zeta(\bar{\mathbf{s}})^2 D\Delta t] + C_{\Delta x}(1|\bar{\mathbf{s}}), \quad (33)$$

$$\langle \Delta x_m \Delta x_n \rangle = C_{\Delta x}(|n - m||\bar{\mathbf{s}}|), \text{ for } |n - m| > 1. \quad (34)$$

Here  $\zeta_x(\bar{\mathbf{s}}) = \sqrt{(\partial x(\bar{\mathbf{s}})/\partial s_x)^2}$  is the mean degree of stretching of the substrate. It is equal to one for three-dimensional and flat substrates, and smaller than one for substrates that tend to bend and twist.  $C_{\Delta x}(|n - m||\bar{\mathbf{s}}|) = \text{cov}(\Delta x_n(\bar{\mathbf{s}}), \Delta x_m(\bar{\mathbf{s}}))$  is the autocovariance of the displacements of the  $x$  coordinate in the laboratory of the point  $\bar{\mathbf{s}}$  on the substrate in the presence of motion blur. It is equal to  $C_{\Delta x}(j|\bar{\mathbf{s}}) = 2C_x(j|\bar{\mathbf{s}}) - C_x(j - 1|\bar{\mathbf{s}}) - C_x(j + 1|\bar{\mathbf{s}})$ , where  $C_x$  is the autocovariance of the  $x$  coordinate in the laboratory of the point  $\bar{\mathbf{s}}$  on the substrate in the presence of motion blur. It is given by

$$\begin{aligned} C_x(j\Delta t|\bar{\mathbf{s}}) &= \frac{1}{(\Delta t)^2} \int_0^{\Delta t} \int_0^{\Delta t} \rho_x(j\Delta t - t + t'|\bar{\mathbf{s}}) dt' dt \\ &= \begin{cases} \frac{2[c-1+2\pi f_{c,x}(\bar{\mathbf{s}})\Delta t]}{[2\pi f_{c,x}(\bar{\mathbf{s}})]^3(\Delta t)^2} K_x(\bar{\mathbf{s}}) & \text{for } j = 0, \\ \frac{(1-c)^2}{[2\pi f_{c,x}(\bar{\mathbf{s}})]^3(\Delta t)^2} K_x(\bar{\mathbf{s}}) c^{|j-1|} & \text{for } j \neq 0, \end{cases} \end{aligned} \quad (35)$$

where  $c = \exp[-2\pi f_{c,x}(\bar{\mathbf{s}})\Delta t]$ .

The covariance [Eqs. (32)–(34)] of single-time-lapse displacements from a Monte Carlo generated time series is shown in Figs. 5(a) and 5(d). Notice the nonzero value of the covariance at unit lag and the negative values with exponentially decreasing magnitude at higher lags, characteristic of diffusion on a fluctuating substrate, e.g., DNA [Fig. 6(a)] and biological membranes [38, Table 4].

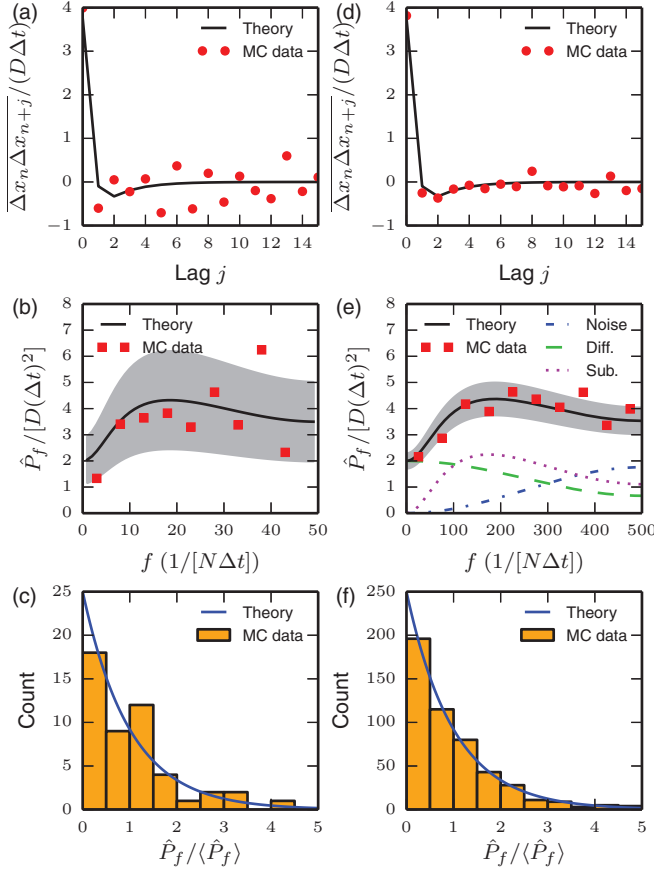


FIG. 5. (Color online) Statistics of single-time-lapse displacements of a particle diffusing on a fluctuating substrate. Panels (a)–(c) show the statistics of a time series of length  $N = 100$ , while (d)–(f) show statistics for  $N = 1000$ . (a),(d) Covariance of single-time-lapse displacements in units of  $D\Delta t$  calculated from a Monte Carlo (MC) simulated time series and compared to their expected values. The covariance shows an isolated positive value at zero lag, the sum of positive contributions due to diffusion, localization error, and substrate fluctuations. The value at unit lag is a sum of both positive and negative contributions. At higher lag, the covariance is negative and decreases exponentially in magnitude, a signature of substrate motion. (b),(e) Periodogram of single-time-lapse Monte Carlo generated displacements in units of  $D(\Delta t)^2$  compared to their expected values, the power spectrum. Shown values are block averages over 10 (b) or 100 (e) periodogram values each [31]. Block averages are shown since the raw periodogram values have a SNR of one, which makes visual comparison impractical. The gray area marks the 68% confidence interval (CI) for the block-averaged values. The power spectrum is composed of three terms: a term due to localization error (Noise), a diffusive (Diff.) term, and a term due to substrate (Sub.) motion. (c),(f) The normalized periodogram values  $\epsilon_f = \hat{P}_f / P_f(\theta)$  follow an exponential distribution with unit expected value. In (a),(b),(d),(e)  $R = 1/6$  (maximally open shutter), and  $D = 0.31 \mu\text{m}^2 \text{s}^{-1}$ ,  $\sigma^2 = 1500 \text{nm}^2$  (SNR = 1.5),  $2\pi f_{c,x} = 50 \text{Hz}$ , and  $K_x = 0.37 \mu\text{m}^2 \text{s}^{-1}$ , corresponding to typical values observed experimentally.

### C. Power spectrum of displacements of a particle diffusing on a fluctuating substrate

It is useful to know the power spectrum of measured displacements of a particle diffusing on a fluctuating substrate.

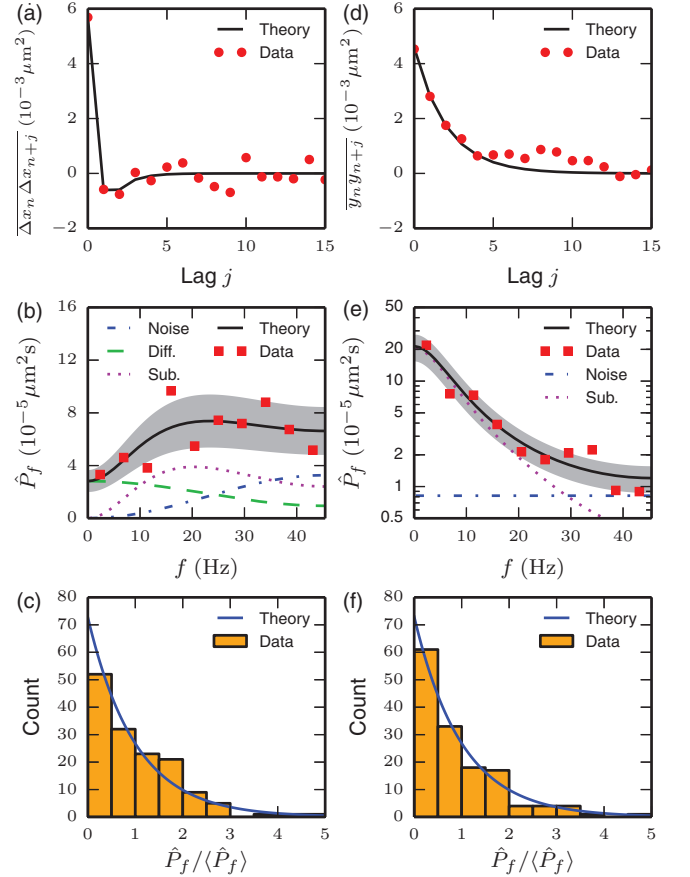


FIG. 6. (Color online) Statistics of time-lapse measured time series of a hOGG1 protein diffusing on flow-stretched  $\lambda$  DNA. The time series consist of 293 points measured with maximally open shutter and time lapse  $\Delta t = 11 \text{ms}$ . In (a),(b),(d),(e) black lines are fits of the phenomenological theory (Sec. IV D) for diffusion on DNA to data. (a) Covariance of single-time-lapse displacements of the protein along the DNA's longitudinal direction. Compare this to Figs. 5(a) and 5(d). (b) Periodogram of the particle's single-time-lapse displacements along the DNA's longitudinal direction. Compare to Figs. 5(b) and 5(e). (c) Distribution of normalized periodogram values  $\epsilon_{\Delta x,f} = \hat{P}_{\Delta x,f} / P_{\Delta x,f}(\theta)$  compared to a theoretically expected exponential distribution. Compare this panel to Figs. 5(c) and 5(f). (d) Covariance of the protein's measured transversal positions. Compare this to Figs. 7(a) and 7(d). (e) Periodogram of the particle's transversal positions. Compare to Figs. 7(b) and 7(e). (f) Distribution of normalized periodogram values  $\epsilon_{y,f} = \hat{P}_{y,f} / P_{y,f}(\theta)$  and theoretically expected exponential distribution with unit mean. Compare this panel to Figs. 7(c) and 7(f). Shown values for data in (b) and (e) are block averages, each over 29 periodogram values, and the gray area marks the 68% confidence interval (CI) for the blocked values. The estimated parameter values are  $\hat{D} = 0.12 \mu\text{m}^2 \text{s}^{-1}$ ,  $\sigma^2 = 750 \text{nm}^2$ ,  $2\pi \hat{f}_c = 44 \text{Hz}$ ,  $\hat{K}_x = 0.33 \mu\text{m}^2 \text{s}^{-1}$ , and  $\hat{K}_y = 0.20 \mu\text{m}^2 \text{s}^{-1}$ .

As in Sec. III, we use it to test whether the recorded motion of a particle is consistent with diffusion on a fluctuating substrate (Sec. V C). We do this by comparing the measured periodogram to the power spectrum [Figs. 5(b) and 5(e) and 6(b)] and by using that the periodogram values are exponentially distributed about their expected values, the power spectral



values. We also use this power spectrum to construct a MLE for the diffusion coefficient  $D$ , the variance  $\sigma^2$  of the localization error, and parameters characterizing substrate fluctuations, all from a single measured time series (Sec. V A).

As for the covariance of displacements, the power spectrum is the sum of three terms, the three individual power spectra of the three independent terms in Eq. (31). We derive expressions for each of them separately.

### 1. Power spectrum of the diffusive and localization error terms

Let  $P_f$  denote the power spectrum of a time series, and let an additional subscript on  $P_f$  indicate which time series. Thus, the power spectrum of  $\Delta s_{x,1}, \dots, \Delta s_{x,N}$  is

$$P_{\Delta s_{x,f}} = 2D(\Delta t)^2\{1 - [1 - \cos(2\pi f \Delta t)]/3\}, \quad (36)$$

in consequence of the white-noise property of diffusive displacements and the low-pass filtering effect of motion blur. Likewise, the power spectrum of the localization error  $\sigma \xi_{x,n}$  is equal to  $\sigma^2 \Delta t$ . The DFT of  $\Delta \xi_{x,n} = \xi_{x,n} - \xi_{x,n-1}$  is equal to  $(1 - e^{i2\pi f \Delta t})\xi_{x,f}$  to order  $1/N$ . Using that, we find that the power spectrum of  $\sigma \Delta \xi_{x,n}$  is equal to

$$P_{\sigma \Delta \xi_{x,f}} = 2\sigma^2 \Delta t [1 - \cos(2\pi f \Delta t)]. \quad (37)$$

### 2. Power spectrum of substrate fluctuations

Since a time series is recorded at a finite frame rate  $f_{\text{sample}} = 1/\Delta t$ , its periodogram is *aliased* [31, Appendix F].

$$P_{\Delta x,f} = 2\zeta(\bar{s})^2 D \Delta t^2 + [2\sigma^2 \Delta t - 2\zeta(\bar{s})^2 D \Delta t^2 / 3][1 - \cos(2\pi f \Delta t)] + \frac{2K_x(\bar{s})}{(2\pi f_{c,x})^3 \Delta t c} [1 - \cos(2\pi f \Delta t)] \left[ c^2 - 1 + 4\pi f_{c,x} \Delta t c + \frac{(1+c)(1-c)^3}{1+c^2-2c \cos(2\pi f \Delta t)} \right]. \quad (40)$$

Examples of the power spectrum and periodogram of time series of a particle diffusing on a fluctuating substrate are shown in Figs. 5(b) and 5(e).

### 4. Distribution of the periodogram

The displacements measured in the laboratory frame for the particle diffusing on the substrate are stationary processes, since they are the sum of three time series, each of which is a stationary process; two of them, localization errors and the displacements themselves in the reference frame of the substrate, even consist of jointly Gaussian distributed terms. Consequently, both the real and the imaginary parts of the Fourier transform of displacements consist of independently Gaussian distributed terms (Appendix D). Since the real and imaginary parts of the Fourier transform at a given frequency have the same variance, periodogram values are exponentially distributed about their expected values, being the sum of two squares of Gaussian values with same variance [Figs. 5(c) and 5(f)]. The periodogram values corresponding to different frequencies are furthermore independent, except for terms of order  $1/N^2$  or higher (Appendix D).

### D. Statistics of diffusion on fluctuating, stretched DNA

We now consider a particle diffusing on an incompressible, taut, fluctuating, flexible or semiflexible fiber or polymer

Moreover, the camera's finite shutter time acts as a low-pass filter. Consequently, the periodogram should not be compared to the power spectrum based on a continuous-time theory,  $P_f^{(\text{sub,cont})}$ , but on an aliased and low-pass filtered version of that power spectrum (Appendix F),

$$P_{x,f}^{(\text{sub})} = \frac{K_x(\bar{s})}{(2\pi f_{c,x})^3 \Delta t c} \left[ c^2 - 1 + 4\pi f_{c,x} \Delta t c + \frac{(1+c)(1-c)^3}{1+c^2-2c \cos(2\pi \Delta t f)} \right], \quad (38)$$

with  $f_{c,x} = f_{c,x}(\bar{s})$  and  $c = \exp(-2\pi f_{c,x} \Delta t)$ .

As above, we use that

$$P_{\Delta x,f} = 2[1 - \cos(2\pi f \Delta t)]P_{x,f} \quad (39)$$

up to neglected contributions from the two ends of the series, contributions of order  $N^{-2}$ . Insertion of Eq. (38) into Eq. (39) yields the contribution from substrate fluctuations to the measured power spectrum.

### 3. Total power spectrum

Equations (36)–(39) give us the power spectrum of a particle diffusing on a fluctuating substrate recorded in the presence of motion blur and localization error,

such as DNA. We designate by  $L$  the contour length of this polymer. For a one-dimensional substrate of this type, let  $s$  denote the coordinate of a physical point on the substrate as measured along the contour of the substrate, e.g., the base-pair number for DNA. Let  $\mathbf{r}(s,t) = [x(s,t), y(s,t), z(s,t)]^T$  denote the coordinates of this point in the laboratory reference frame at a given time  $t$ . Thermal fluctuations and/or other forces drive the substrate out of equilibrium and a point on the substrate feels a restoring force. Since the substrate is taut, fluctuations are small and the restoring force is locally well approximated as Hookean, i.e., proportional to the distance to equilibrium. As Fig. 6(e) shows, this approximation is excellent for our purpose. This is confirmed by theoretical investigation of substrates stretched by various means [22]. The results presented in this section are easily generalized to include multiple modes (Appendix G).

#### 1. Fluctuations of a taut polymer in solution

We consider an inextensible semiflexible or flexible fiber or polymer (DNA is inextensible at the stretching forces used in measurements of diffusion on DNA [39]). Since the fiber's length is fixed, its longitudinal motion is determined by its motion in the two transversal directions by the condition  $\mathbf{r}' = 1$ , where the prime denotes the derivative with respect to  $s$ . Furthermore, since the fiber is taut, we have by Taylor

expansion

$$x' \simeq 1 - \frac{1}{2}[(y')^2 + (z')^2]. \quad (41)$$

The fiber's motion in the transversal directions is locally that of a harmonic oscillator, because we have assumed that we only resolve the spatial eigenmode with the slowest relaxation time. We do not know the functional form of this eigenmode, unless we have a dynamical theory for the substrate motion. However, we manage without by observing that whatever the functional form of that eigenmode is, its only role in describing the motion of a fixed point or narrow region on the substrate is to specify the amplitude of this motion. Its dynamics is that of an overdamped harmonic oscillator. So we manage by measuring that unknown amplitude experimentally at each location of interest. Its associated relaxation time is also determined experimentally. We then have, e.g., in the  $y$  direction, that

$$\rho_y(t - t' | s) = K_y(s) \tau_y(s) e^{-|t-t'|/\tau_y(s)}, \quad (42)$$

where  $K_y(s)$  parametrizes the amplitude of  $\rho_y$  at  $s$ , and  $\tau_y(s)$  is its characteristic time. By virtue of the Wiener-Khinchin theorem,  $\tau_y(s) = (2\pi f_{c,y})^{-1}$ , where  $f_{c,y}$  is the corner frequency of the Lorentzian power spectrum of the  $y$  coordinate of substrate fluctuation at  $s$ .

The autocovariance of the longitudinal motion is found from Eq. (41) by using that  $y'$  and  $z'$  are Gaussian distributed,

$$\rho_x(t | s) = \frac{1}{2} \left[ \int_0^s \rho_{y'}(t | s') ds' \right]^2 + \frac{1}{2} \left[ \int_0^s \rho_{z'}(t | s') ds' \right]^2, \quad (43)$$

where  $\rho_{y'}$  and  $\rho_{z'}$  are the autocovariances of  $y'$  and  $z'$  and are proportional to  $\rho_y$  and  $\rho_z$ , respectively, with proportionality constants that only depend on  $s$ . Usually the motion in the two transversal directions contribute equally to the longitudinal motion (for DNA stretched by optical tweezers or a plug flow this is exact, for DNA stretched by a shear flow this is approximate [22]). In this case  $f_{c,y} = f_{c,z} = f_c$ , and hence  $\tau_y(s) = \tau_z(s) = \tau(s)$ , so Eq. (43) reduces to

$$\rho_x(t | s) = \frac{1}{2} K_x(s) \tau(s) e^{-2|t|/\tau(s)}. \quad (44)$$

The exact forms of  $K_y$  and  $K_x$  depend on the substrate and how it is stretched. They are, in general, unknown, but can be calculated explicitly for some specific substrates and methods of stretching [22].

## 2. Covariances of displacements of a particle diffusing on taut DNA

*a. Transversal motion.* The measured  $y$  position of a particle diffusing on a fluctuating polymer is given by

$$y_n = \frac{1}{\Delta t} \int_0^{\Delta t} y_{\text{true}}(t_n - t') dt' + \sigma \xi_n^{(y)}. \quad (45)$$

Since localization errors are uncorrelated with substrate motion, and since terms of order  $2D\Delta t/L^2 \ll 1$  or higher can be ignored, the autocovariance of  $y$  is given by

$$\langle (y_n - \langle y \rangle)(y_m - \langle y \rangle) \rangle = \sigma^2 + \mathcal{C}_y(0 | \bar{s}), \quad (46)$$

$$\langle (y_m - \langle y \rangle)(y_n - \langle y \rangle) \rangle = \mathcal{C}_y(|n - m| \Delta t | \bar{s}), \quad \text{for } |n - m| \geq 1. \quad (47)$$

Here  $\mathcal{C}_y$  is the autocovariance of the  $y$  coordinate of the substrate's motion,

$$\mathcal{C}_y(j \Delta t | \bar{s}) = \begin{cases} \frac{2(c-1+2\pi f_c \Delta t)}{(2\pi f_c)^3 (\Delta t)^2} K_y(\bar{s}) & \text{for } j = 0, \\ \frac{(1-c)^2}{(2\pi f_c)^3 (\Delta t)^2} K_y(\bar{s}) c^{|j-1|} & \text{for } j \neq 0, \end{cases} \quad (48)$$

with  $c = \exp(-2\pi f_c \Delta t)$ . Examples of Monte Carlo simulations and experimental measurements of the covariance of the transversal positions of a particle diffusing on fluctuating DNA are shown in Figs. 7(a) and 7(d) and 6(d).

*b. Longitudinal motion.* The covariance of the measured longitudinal displacements of a particle diffusing on a polymer is described by Eqs. (32)–(34) [Figs. 5(a) and 5(d) and 6(a)], except that the autocovariance of the substrate's position,  $\mathcal{C}_x$ , is given by

$$\mathcal{C}_x(j \Delta t | \bar{s}) = \begin{cases} \frac{2(c^2-1+4\pi f_c \Delta t)}{(4\pi f_c)^3 (\Delta t)^2} K_x(\bar{s}) & \text{for } j = 0, \\ \frac{(1-c^2)^2}{(4\pi f_c)^3 (\Delta t)^2 c^2} K_x(\bar{s}) c^{2|j|} & \text{for } j \neq 0. \end{cases} \quad (49)$$

The *degree of stretching or tortuosity*,  $\zeta(\bar{s})$ , is constant along a DNA molecule or another polymer stretched by applying a force to its ends. It is approximately constant and close to one for a tethered polymer in a strong hydrodynamic flow, except near the polymer's free end [22,40–44].

## 3. Power spectra of displacements of a particle diffusing on taut DNA

*a. Transversal motion.* The power spectrum of the transversal motion is calculated from Eq. (45) in the same manner as Eq. (40) was derived from Eq. (31). This gives

$$P_{y,f} = \sigma^2 \Delta t + \frac{K_y(\bar{s})}{(2\pi f_c)^3 \Delta t c} \left[ c^2 - 1 + 4\pi f_c \Delta t c + \frac{(1+c)(1-c)^3}{1+c^2-2c \cos(2\pi f \Delta t)} \right]. \quad (50)$$

See Figs. 7(b) and 7(e) and 6(e) for examples of the power spectrum  $P_{y,f}$ .

*b. Longitudinal motion.* The power spectrum of the particle's displacements in the longitudinal direction of the substrate is derived in the same manner as above and is given by

$$P_{\Delta x,f} = 2\zeta(\bar{s})^2 D \Delta t^2 + [2\sigma^2 \Delta t - 2\zeta(\bar{s})^2 D \Delta t^2 / 3] [1 - \cos(2\pi f \Delta t)] + \frac{2K_x(\bar{s})}{(4\pi f_c)^3 \Delta t c^2} [1 - \cos(2\pi f \Delta t)] \left[ c^4 - 1 + 8\pi f_c \Delta t c^2 + \frac{(1+c^2)(1-c^2)^3}{1+c^4-2c^2 \cos(2\pi f \Delta t)} \right]. \quad (51)$$

Figures 5(b), 5(e), and 6(b) show examples of what  $P_{\Delta x,f}$  may look like.

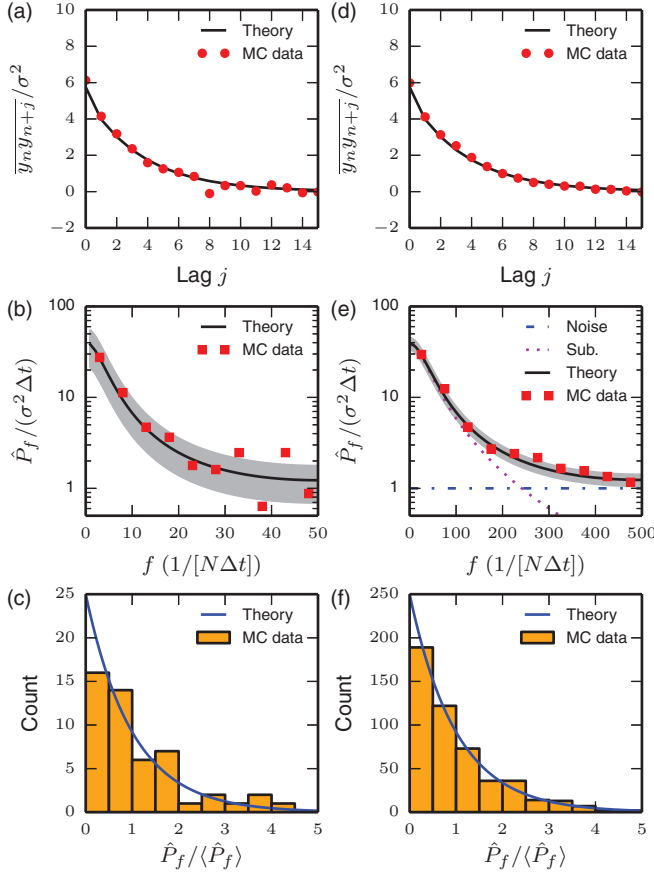


FIG. 7. (Color online) Statistics of the time-lapse measured transversal position of a particle diffusing on a taut but flexible, fluctuating fiber, such as DNA. (a),(d) Covariance of the measured positions in units of  $\sigma^2$  calculated from a Monte Carlo (MC) simulated time series and compared to their expected values. The covariance shows an positive value at zero lag, the sum of contributions from localization error and substrate fluctuations, and an exponential decrease for higher lags, the signature of substrate fluctuations. (b),(e) Periodogram of Monte Carlo generated positions in units of  $\sigma^2 \Delta t$  compared to their expected values, the power spectrum. Shown values are block averages over 10 (b) or 100 (e) periodogram values each [31]. The gray area marks the 68% confidence interval (CI) for the block-averaged values. The power spectrum contains an additive constant term due to uncorrelated localization errors on positions (Noise) and a Lorentzian additive term from the substrate's motion (Sub.). (c),(f) The normalized periodogram values  $\epsilon_f = \hat{P}_f / P_f(\theta)$  are exponentially distributed on the positive real axis with unit expected value. In (a),(b),(d),(e)  $R = 1/6$  (maximally open shutter),  $\sigma^2 = 1500 \text{ nm}^2$ ,  $2\pi f_c = 25 \text{ Hz}$ , and  $K_y = 0.20 \mu\text{m}^2 \text{ s}^{-1}$ .

*c. Distribution of the periodogram.* The measured transversal positions of the particle on the substrate is the sum of two Gaussian distributed terms, so the transversal position is Gaussian itself. The real and imaginary parts of the Fourier transforms of these displacements are thus also Gaussian distributed. As a result, their squared moduli are exponentially distributed, as demonstrated in Figs. 7(c) and 7(f) and 6(f)]. These exponentially distributed periodogram values are approximately independent (Appendix D).

Less is needed, however, to arrive at exponentially distributed periodogram values: The longitudinal substrate fluctuations are not Gaussian distributed, since longitudinal motion is a nonlinear function of the transversal motion. This means that  $\Delta x_n$  is not Gaussian. However, since its Fourier transform is a weighted sum over all the measured  $\Delta x_n$ , it is approximately Gaussian by virtue of the central limit theorem. Furthermore, the measured Fourier transform is a sum of the Fourier transforms of the three independent terms in Eq. (31), of which the two others are Gaussian. The periodogram values of the measured longitudinal displacements are thus, in practice, exponentially distributed [Figs. 5(c) and 5(f) and 6(c)] and independent (to order  $1/N^2$ , Appendix D).

## V. ESTIMATION OF DIFFUSION COEFFICIENT OF A PARTICLE DIFFUSING ON A FLUCTUATING SUBSTRATE

Based on the statistics of diffusion on a taut, fluctuating substrate (Sec. IV), we here derive two estimators of the diffusion coefficient of a particle diffusing on such a substrate from a time series of its measured positions. When the substrate's motion cannot be characterized independently, e.g., by direct measurement, and the measured time series is long, the MLE derived in Sec. V A optimally estimates the particle's diffusion coefficient  $D$ , the variance  $\sigma^2$  of localization error on positions, and parameters  $\phi$  characterizing the substrate's motion. For the case where substrate motion has been characterized independently, we derive in Sec. V B an unbiased CVE which optimally estimates  $D$  along with  $\sigma^2$  even from a short time series. We finally give a procedure for rigorous statistical testing of whether a recorded time series describes free diffusion on a fluctuating substrate in Sec. V C.

### A. Maximum likelihood estimation

We construct a MLE of the diffusion coefficient  $D$ , the variance  $\sigma^2$  of localization errors, and parameters of substrate motion  $\phi$  in the point  $\bar{s}$  on the substrate. This MLE uses the periodogram of the measured time series as input. If one has recorded  $N$  displacements  $\Delta \mathbf{r}_0, \Delta \mathbf{r}_1, \dots, \Delta \mathbf{r}_N$  of a particle diffusing on an elastic,  $d$ -dimensional substrate, one can calculate  $d$  periodograms, one for each coordinate of the time series. Since the motion along the coordinate axes are independent, estimates can be obtained for each coordinate individually and averaged afterwards, as in the case of free diffusion (Sec. III).

In the  $x$  direction, the periodogram values are

$$\hat{P}_{\Delta x, f} = |\hat{\Delta x}_f|^2 \Delta f, \quad (52)$$

where  $f \in \{\Delta f, 2\Delta f, \dots, f_{\text{Nyq}}\}$ , with  $\Delta f = 1/(N\Delta t)$  and the Nyquist frequency  $f_{\text{Nyq}} = 1/(2\Delta t) = f_{\text{sample}}/2$  for  $N$  even and  $f_{\text{Nyq}} = (N-1)/(2N\Delta t)$  for  $N$  odd. The DFT  $\hat{\Delta x}_f$  of the displacements  $(\Delta x_n)_{n=1}^N$  is calculated efficiently with, e.g., MATLAB or SciPy's fft [32].

Given the measured periodogram  $(\hat{P}_{\Delta x})_f$ , the MLE is the set of parameter values  $(\hat{D}, \hat{\sigma}^2, \hat{\phi})$  which maximizes the log-likelihood function

$$\ln \mathcal{L}(D, \sigma^2, \phi | (\hat{P}_{\Delta x})_f) = \sum_{f=\Delta f}^{f_{\text{Nyq}}} \left( \ln P_{\Delta x, f} + \frac{\hat{P}_{\Delta x, f}}{P_{\Delta x, f}} \right). \quad (53)$$

The power spectral values  $P_{\Delta x, f}$  are given by Eq. (40). If more than one mode is needed to properly describe the substrate's fluctuations, or if the camera shutter is not kept open for the full duration of the time lapse, see Appendix G. It gives expressions for the power spectrum  $P_{\Delta x, f}$  for a multi-mode theory and for a camera shutter time that is shorter than the time lapse.

If the parameters describing substrate motion have been determined independently, they can be used as fixed parameters in the MLE algorithm. This is done by plugging them into Eq. (53) and maximizing  $\ln \mathcal{L}$  with respect to the remaining parameters only. In this case, however, a simpler, unbiased CVE which also reaches the Cramér-Rao bound, should be used instead (Sec. VB).

### 1. Variance of the MLE

We define  $\theta = (D, \sigma^2, \phi)^T$ . The covariance matrix of the MLE is, to first order in  $1/N$ , equal to the inverse of the Fisher information matrix  $\mathcal{I}$ , the elements of which are

$$\mathcal{I}_{jk} = \sum_{f=\Delta f}^{f_{\text{Nyq}}} (P_{\Delta x, f})^{-2} \frac{\partial P_{\Delta x, f}}{\partial \theta_j} \frac{\partial P_{\Delta x, f}}{\partial \theta_k}. \quad (54)$$

The variance of the MLE is estimated by inserting estimated values of  $\theta$  in Eq. (54).

If parameters characterizing substrate fluctuations have been estimated independently and used in the MLE, the variances of its estimates of  $D$  and  $\sigma^2$  are as given in Appendix H.

### B. Covariance-based estimation

If independent estimates  $(\hat{K}_x, \hat{f}_{c,x})$  of parameters characterizing substrate motion have been obtained, an unbiased CVE of the diffusion coefficient can be constructed by subtracting the bias caused by substrate fluctuations from the CVE that was designed for free diffusion [Eq. (14)]. For a particle with mean position  $\bar{s}$  on the substrate, the bias due to substrate fluctuations is

$$b_D[K_x(\bar{s}), f_{c,x}(\bar{s})] = \left[ \frac{1-c}{2\pi f_{c,x}(\bar{s})\Delta t} \right]^3 K_x(\bar{s}), \quad (55)$$

with  $c = \exp[-2\pi f_{c,x}(\bar{s})\Delta t]$ . An unbiased CVE of  $D$  is then

$$\hat{D} = \left[ \frac{\Delta x_n^2}{2\Delta t} + \frac{\overline{\Delta x_n \Delta x_{n+1}}}{\Delta t} - b_D(\hat{K}_x, \hat{f}_{c,x}) \right] / \zeta(\bar{s})^2, \quad (56)$$

where eventual incomplete stretching of the substrate has also been taken into account by dividing by  $\zeta(\bar{s})^2$ . In the same manner we can derive an unbiased CVE for  $D$  in the case when also the localization error has been characterized independently. It is given in Appendix G. In practice, it is not more precise than the CVE given by Eq. (56) (Sec. VID). So one can in all cases use Eq. (56).

The CVE [Eq. (56)] requires prior knowledge of the parameters  $K_x(\bar{s})$ ,  $f_{c,x}(\bar{s})$ , and  $\zeta(\bar{s})$  that characterize the substrate and its fluctuations.  $\zeta(\bar{s})$  is often known *a priori*. For a three-dimensional homogeneous substrate, e.g., it is equal to one. For a polymer stretched by constant forces applied to its ends, it is constant and equal to the overall degree of stretching while  $K_x(\bar{s})$  and  $f_{c,x}(\bar{s})$  are measured experimentally.

If these parameters cannot be determined by direct measurement of the substrate motion, estimates can be obtained by averaging over MLEs (Sec. VIA) obtained from long time series of particles with a mean position close to  $\bar{s}$ . If we have recorded  $M$  time series that are long enough for reliable estimation of  $\phi(\bar{s}) = [K_x(\bar{s}), f_{c,x}(\bar{s})]$ , we calculate the weighted average of  $\hat{K}_x$  and  $\hat{f}_{c,x}$  as [45]

$$\bar{\hat{\phi}}_i = \frac{\sum_{m=1}^M N_m \hat{\phi}_{i,m}}{\sum_{m=1}^M N_m}, \quad (57)$$

where  $N_m$  is the length of the  $m$ th time series. The variance of the weighted average is estimated by

$$\text{var}(\bar{\hat{\phi}}_i) = \frac{\sum_{m=1}^M N_m (\hat{\phi}_{i,m} - \bar{\hat{\phi}}_i)^2}{(M-1) \sum_{m=1}^M N_m}. \quad (58)$$

These estimates can then be used in Eq. (56) to obtain practically optimal and unbiased estimates of the diffusion coefficient [Figs. 8(c) and 8(d)].

One may consider alternative approaches to reduce or eliminate the bias of the original CVE without using ML estimated parameter values, e.g., by adding higher autocovariances to reduce the bias or by estimating substrate fluctuation parameters directly, using higher autocovariances. Both approaches considerably increase the variance of estimates of the diffusion coefficient, however, compared to the CVE defined above, and hence should not be used.

### 1. Variance of the CVE

The variance of the unbiased CVE is to order  $1/N^2$  (Appendix H)

$$\begin{aligned} \frac{\text{var}(\hat{D})}{D^2} &= V_0(D, \sigma^2, \phi) + \frac{2V_1(D, \sigma^2, \phi)}{\zeta(\bar{s})^2 D} \\ &+ \frac{V_2(\phi) + \text{var}[b_D(\hat{\phi})]}{[\zeta(\bar{s})^2 D]^2} + \frac{\text{var}[\zeta(\bar{s})^2]}{\zeta(\bar{s})^4}, \end{aligned} \quad (59)$$

where we have introduced three functions,

$$V_0(D, \sigma^2, \phi) = \frac{6 + 4\varepsilon + 2\varepsilon^2}{N} + \frac{4(1 + \varepsilon)^2}{N^2}, \quad (60)$$

$$\begin{aligned} V_1(D, \sigma^2, \phi) &= \frac{(3 + \varepsilon)\mathcal{C}_0 + 4\mathcal{C}_1 + 2\mathcal{C}_2 - \varepsilon\mathcal{C}_3}{N(\Delta t)^2} \\ &+ \frac{2(1 + \varepsilon)\mathcal{C}_0 + 2\varepsilon\mathcal{C}_2 + \varepsilon\mathcal{C}_3}{N^2(\Delta t)^2}, \end{aligned} \quad (61)$$

$$\begin{aligned} V_2(\phi) &= \frac{1}{(\Delta t)^2} \sum_{j=1}^{N-1} \left[ \frac{N-j}{N^2} \mathcal{C}_j^2 + \frac{4(N-j)}{N(N-1)} \mathcal{C}_{j-1} \mathcal{C}_j \right. \\ &+ \left. \frac{2(N-j-1)}{(N-1)^2} (\mathcal{C}_{j-1} \mathcal{C}_{j+1} + \mathcal{C}_j^2) \right] \\ &+ \frac{3\mathcal{C}_0^2/2 + \mathcal{C}_1^2}{N(\Delta t)^2} + \frac{\mathcal{C}_0^2 + \mathcal{C}_1^2}{N^2(\Delta t)^2}, \end{aligned} \quad (62)$$

using  $\varepsilon = [\sigma^2 - \zeta(\bar{s})^2 D \Delta t / 3] / [\zeta(\bar{s})^2 D \Delta t]$  and  $\mathcal{C}_j \equiv \mathcal{C}_{\Delta x}(j \Delta t | \bar{s})$ , with  $\mathcal{C}_{\Delta x}(j \Delta t | \bar{s}) = 2\mathcal{C}_x(j \Delta t | \bar{s}) - \mathcal{C}_x(|j-1| \Delta t | \bar{s}) - \mathcal{C}_x(|j+1| \Delta t | \bar{s})$  and  $\mathcal{C}_x(j \Delta t | \bar{s})$  given in Eq. (35). A



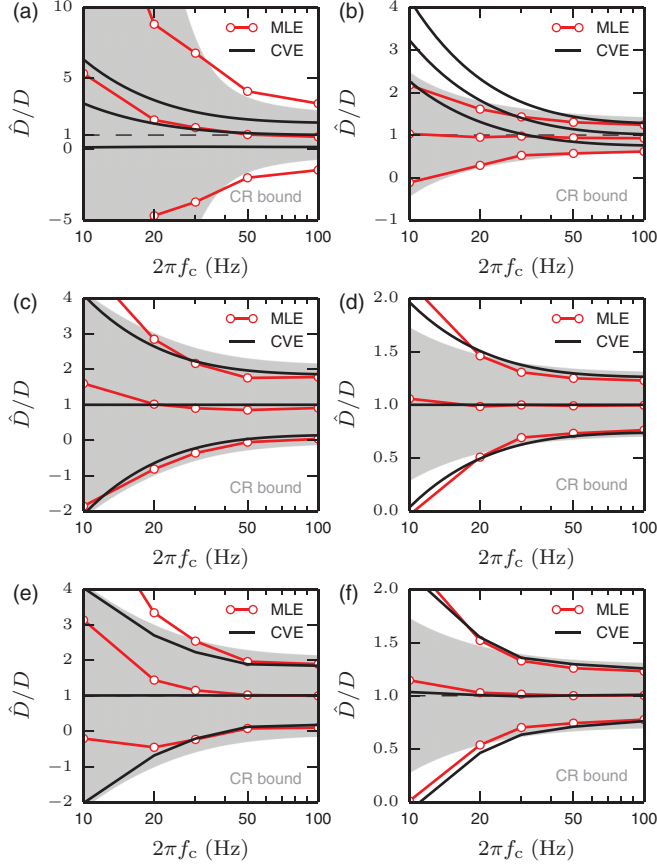


FIG. 8. (Color online) Quality of various estimators for the diffusion coefficient for diffusion on a fluctuating, taut, but flexible one-dimensional substrate. Mean plus/minus standard error of estimators given in units of the true value  $D$ . Estimates  $\hat{D}$  were obtained with MLE and CVE applied to 1000 Monte Carlo generated time series of positions of a particle diffusing on a fluctuating substrate recorded with motion blur and localization errors. (a),(c),(e) Time series of length  $N = 10$ . (b),(d),(f) Time series of length  $N = 100$ . (a),(b) Case of unknown variance  $\sigma^2$  of localization errors and unknown parameters of DNA motion. They are all estimated together with  $D$  from the same time series. Results: CVE is highly biased for large substrate fluctuations ( $=$ small  $2\pi f_c$ ). MLE is optimal for large  $N$ . For small  $N$ , both CVE and MLE fail: Their estimates have systematic errors (bias) larger than the true value of  $D$ . (c),(d) Case where substrate fluctuations have been characterized independently. Results: Both CVE and MLE practically reach the Cramér-Rao bound, but only CVE is unbiased. (e),(f) Case where also localization errors have been characterized independently. Results: Precision of CVE increases negligibly; compare with (c),(d). MLE increases its bias for large substrate fluctuations ( $=$ small  $2\pi f_c$ ), especially at small  $N$ . CVE is unbiased.

linear approximation to  $b_D$ 's dependence on  $\phi$  results in the usual formula for linearized error propagation:

$$\text{var}[b_D(\phi)] = \sum_{ij} \frac{\partial b_D}{\partial \phi_i} \frac{\partial b_D}{\partial \phi_j} \text{cov}(\hat{\phi}_i, \hat{\phi}_j). \quad (63)$$

### C. Testing whether a recorded trajectory describes free diffusion on a fluctuating substrate

As for diffusion on a substrate at rest, we can design a test for diffusion on a fluctuating substrate based on the periodogram of the measured time series of the particle's positions. Fourier transformation of the set of displacements  $\Delta x_1, \Delta x_2, \dots, \Delta x_N$  decorrelates them (Appendix D), which means that their periodogram values,  $\hat{P}_f$ , are uncorrelated (to order  $1/N^2$ ). This makes them well suited for statistical testing of whether a time series describes free diffusion on the fluctuating substrate. The periodogram values  $\hat{P}_f$  should be compared to their expected values, the power spectrum, given by Eq. (40) [Figs. 5(b) and 5(e)], and the distribution of their normalized values  $\epsilon_f = \hat{P}_f / P_f(\hat{\theta})$  to an exponential distribution with unit mean [Figs. 5(c) and 5(f)]. If a  $P$  value is wanted for quantitative comparison, it can be calculated from  $\epsilon_{\Delta f}, \epsilon_{2\Delta f}, \dots, \epsilon_{f_{\text{Nyq}}}$  using Pearson's  $\chi^2$  goodness-of-fit test (Appendix J) with four fitted parameters,  $D, \sigma^2, K_x$ , and  $f_c$  (two if substrate motion has been characterized independently).

## VI. ESTIMATION FOR DIFFUSION ON FLUCTUATING, TAUT DNA

We here derive a MLE of a particle's diffusion coefficient  $D$ , variance  $\sigma^2$  of localization error, and parameters  $\phi$  characterizing substrate motion from a recorded time series of the particle's motion on a taut, fluctuating polymer, e.g., a DNA strand, optimal for long time series (Sec. VIA). For the case when the DNA strand's motion has been characterized independently, we derive an unbiased CVE of  $D$  and  $\sigma^2$ . This CVE estimates the diffusion coefficient much more precisely than the MLE, which also estimates  $\phi$ , and it is optimal even for short time series (Sec. VIB). We develop a rigorous procedure for testing whether a recorded time series describes diffusion on a fluctuating, taut DNA strand (Sec. VIC). Finally, we test the estimators on synthetic Monte Carlo generated data and compare their precision with the information inequality, the Cramér-Rao lower bound on achievable precision (Sec. VID).

### A. Maximum likelihood estimation

We here construct a MLE for the special case of a particle diffusing on a taut, incompressible fiber or polymer such as DNA. This case is special because our observations include substrate motion in dimensions in addition to the one dimension of the substrate. This extra information offers extra insight. Specifically, the substrate is stretched along the  $x$  axis of laboratory coordinates, but in addition to displacements  $\Delta x$  of the particle diffusing along the substrate, we also observe the particle's transversal positions  $y$ . The latter trace the transverse motion of the small region of the substrate that is visited by a particle diffusing on it. In order to use this extra information, we assume that the recorded transversal positions and longitudinal displacements are uncorrelated. This assumption is not completely true, since the longitudinal DNA fluctuations are fully determined by the upstream fluctuations in the two transversal directions  $y$  and  $z$ . However, the contribution of *local* fluctuations in transverse position to longitudinal displacements at the same locality is so small, simulations show (Appendix I), that neglecting them

does not affect estimates noticeably. A comparison between estimator precision and the approximate Cramér-Rao bound shows this (Fig. 8).

The periodograms are calculated as

$$\hat{P}_{\Delta x, f} = |\hat{\Delta x}_f|^2 \Delta f \quad (64)$$

for  $f = \Delta f, 2\Delta f, \dots, f_{\text{Nyq}}$  and

$$\hat{P}_{y, f} = |\hat{y}_f|^2 \Delta f' \quad (65)$$

for  $f = \Delta f', 2\Delta f', \dots, f'_{\text{Nyq}}$ , where  $\Delta f' = 1/[(N+1)\Delta t]$ , while the Nyquist frequencies are  $f_{\text{Nyq}}$  defined above [23] and  $f'_{\text{Nyq}} = N/[2(N+1)\Delta t]$  for  $N$  even and  $f'_{\text{Nyq}} = 1/(2\Delta t) = f_{\text{sample}}/2$  for  $N$  odd.

Given  $\hat{P}_{\Delta x, f}$  and  $\hat{P}_{y, f}$ , the MLE  $\hat{\theta} = (\hat{D}, \hat{\sigma}^2, \hat{\phi})$  returns the set of parameter values  $\theta = (D, \sigma^2, \phi)$ , which maximizes the log-likelihood function

$$\begin{aligned} \ln \mathcal{L}(\theta | (\hat{P}_f)_f) = & \sum_{f=\Delta f}^{f_{\text{Nyq}}} \left( \ln P_{\Delta x, f} + \frac{\hat{P}_{\Delta x, f}}{P_{\Delta x, f}} \right) \\ & + \sum_{f=\Delta f'}^{f'_{\text{Nyq}}} \left( \ln P_{y, f} + \frac{\hat{P}_{y, f}}{P_{y, f}} \right). \end{aligned} \quad (66)$$

The power spectral values  $P_{\Delta x, f}$  and  $P_{y, f}$  are given by Eqs. (51) and (50), respectively.

### 1. Variance of the MLE

The variance of the MLE is, to order  $1/N$ , equal to  $\text{var}(\hat{\theta}) = \mathcal{I}^{-1}$  [18]. Here the Fisher information matrix  $\mathcal{I}$  is given by

$$\begin{aligned} \mathcal{I}_{ij} = & \sum_{f=\Delta f}^{f_{\text{Nyq}}} (P_{\Delta x, f})^{-2} \frac{\partial P_{\Delta x, f}}{\partial \theta_i} \frac{\partial P_{\Delta x, f}}{\partial \theta_j} \\ & + \sum_{f=\Delta f'}^{f'_{\text{Nyq}}} (P_{y, f})^{-2} \frac{\partial P_{y, f}}{\partial \theta_i} \frac{\partial P_{y, f}}{\partial \theta_j}, \end{aligned} \quad (67)$$

where  $P_{\Delta x, f}$  and  $P_{y, f}$  are given by Eqs. (51) and (50), respectively.

### B. Covariance-based estimation

If independent estimates of parameters characterizing substrate motion can be obtained either by direct measurement of the DNA strand's fluctuations or from long time series using MLE as described in Sec. VB, an optimal and unbiased CVE of diffusion coefficients can be constructed. This estimator is generally as precise as the MLE (Sec. VID) and is guaranteed to be unbiased even for short time series.

This unbiased CVE of the diffusion coefficient is given by Eq. (56), where  $b_D$  is given by

$$b_D(K_x, f_c) = \left( \frac{1 - c^2}{4\pi f_c \Delta t} \right)^3 K_x, \quad (68)$$

instead of by Eq. (55).

### 1. Variance of the CVE

The variance of the unbiased CVE is given by Eq. (59) with the expression for  $\mathcal{C}_x(\bar{s}, j)$  [Eq. (35)] replaced with Eq. (49).

### C. Testing whether a recorded trajectory describes free diffusion on fluctuating DNA

To test whether a trajectory describes diffusion along a taut DNA molecule (or any other taut but flexible and fluctuating fiber), the test described in Sec. VC is modified slightly to account for the fact that we also measure the DNA's transversal motion, which is uncontaminated by the particle's diffusive motion. In this case, one should compare the periodograms of the particle's longitudinal, as well as its transversal, motion, respectively  $\hat{P}_{\Delta x, f}$  and  $\hat{P}_{y, f}$ , to their expected values, the power spectra, given by Eq. (51) for  $\Delta x$  [Figs. 5(b) and 5(e) and 6(b)] and Eq. (50) for  $y$  [Figs. 7(b) and 7(e) and 6(e)]. The distributions of the normalized values  $\epsilon_{\Delta x, f} = \hat{P}_{\Delta x, f} / P_{\Delta x, f}(\hat{\theta})$  [Figs. 5(c) and 5(f) and 6(c)] and  $\epsilon_{y, f} = \hat{P}_{y, f} / P_{y, f}(\hat{\theta})$  [Figs. 7(c) and (f) and 6(f)] should be compared to exponential distributions with unit mean. They can be compared quantitatively by calculating a  $P$  value based on Pearson's  $\chi^2$  goodness-of-fit test (Appendix J) taking into account that five parameters,  $D$ ,  $\sigma^2$ ,  $K_x$ ,  $K_y$ , and  $f_c$ , were fitted (two parameters instead of five if the DNA's motion was characterized independently).

### D. Monte Carlo simulations

We tested the estimators on simulated data for a particle diffusing on a taut, but flexible, fluctuating DNA strand. We assumed that all but the lowest mode of the DNA's fluctuations could be ignored as contributors to the particle's movement in laboratory coordinates since this is a signature of taut DNA [Figs. 6(b) and 6(e)] [22]. Consequently, we simulated only the lowest mode.

Since the mathematical description of diffusion on a fluctuating substrate contains two or three additional parameters compared to free diffusion, estimator performance is not determined by a simple SNR, as was the case for diffusion on a substrate at rest (Sec. III), but by a complicated interplay of parameters. Consequently, we did not investigate the whole parameter space, but a subset which fully covers the parameter values that we observed in experimental measurements of hOGG1 proteins diffusing on flow-stretched DNA [22]. The DNA fluctuations are described locally by three parameters: (i) the relaxation rate  $2\pi f_c$ , where  $f_c$  is the corner frequency of the Lorentzian power spectrum of transverse substrate fluctuations; (ii) the amplitude  $K_x(\bar{s})$  of longitudinal substrate fluctuations; and (iii) the amplitude  $K_y(\bar{s})$  of transverse substrate fluctuations.

For the experimental data (Sec. VII), we found that  $K_y(\bar{s})$  does not change between time series though  $2\pi f_c$  varies between 20 and 50 Hz. In contrast,  $K_x(\bar{s})$  scales with  $f_c$  as  $K_x = \alpha_x / (2\pi f_c)$  with coefficient of proportionality  $\alpha_x = 9.2 \pm 1.2 \mu\text{m}^2 \text{s}^{-2}$ . The precisions of estimators are less sensitive to the value of  $\sigma^2$ , while varying  $D$  is mathematically equivalent to a renormalization of the experimental noise, i.e., a change of  $K_x(\bar{s})$  and  $\sigma^2$ . So we kept the parameters  $\alpha_x$ ,  $K_y(\bar{s})$ ,  $D$ , and  $\sigma^2$  constant in our simulations and equal to their experimental mean values:  $D = 0.3 \mu\text{m}^2 \text{s}^{-1}$ ,  $\sigma^2 = 1500 \text{ nm}^2$ ,  $K_y = 0.20 \mu\text{m}^2 \text{s}^{-1}$ , and  $\alpha_x = 9.2 \mu\text{m}^2 \text{s}^{-2}$ , while  $2\pi f_c$  was varied tenfold, between 10 and 100 Hz.

Measured displacements were then simulated as described in Appendix I.

Simulations demonstrate that DNA fluctuations induce a bias in the estimates of the diffusion coefficient, which can be many times larger than the diffusion coefficient itself, if these fluctuations are not accounted for [(Figs. 8(a) and 8(b)]. For long time series [Fig. 8(b)], MLE accounts appropriately for the DNA fluctuations and is practically optimal. For short time series [Fig. 8(a)], both the CVE and the MLE fail to give reliable estimates for the diffusion coefficient: Though CVE outperforms MLE and beats the Cramér-Rao bound by a factor two or more, this is of no practical interest, because its estimates come with a bias that may be larger than the true value of the diffusion coefficient.

In Figs. 8(c) and 8(d), DNA fluctuations were characterized *a priori*. Both CVE and MLE are seen to be practically optimal for short as well as long time series.

In Figs. 8(e) and 8(f), also the variance  $\sigma^2$  of localization errors was determined *a priori*. This results in a slight increase in the precision of the CVE as compared to Figs. 8(c) and 8(d), while it induces a bias in MLE for slow DNA fluctuations. In this case the CVE should be used. There is, however, for the parameter values examined, virtually no increase in precision by knowing  $\sigma^2$  *a priori*, over the case where both  $D$  and  $\sigma^2$  are estimated from the same time series.

## VII. ANALYSIS OF INDIVIDUAL hOGG1 PROTEINS DIFFUSING ON FLUCTUATING DNA

We analyzed a set of hundreds of time series of hOGG1 proteins diffusing on a single  $\lambda$  DNA molecule that was flow stretched over a coverslip. The proteins were fluorescently labeled and visualized with TIRF microscopy; see [21] for details. Measurements were performed at pH 7.5 and 10 mM salt. Images were recorded with a time lapse of  $\Delta t = 11$  ms. A detailed analysis of these data—and of data taken at different pH and salt concentrations—is given in [22]. Here we summarize results from this very practical application of the estimators introduced above. These results include a two-state kinetics of the protein. This kinetics remained hidden in the original analysis of the same data, which was based on MSD.

### A. Comparing MLE and CVE on long time series to remove bias in CVE caused by DNA motion, so CVE can be trusted on short time series, where MLE cannot

We estimated the diffusion coefficient  $D$ , variance of the localization error  $\sigma^2$ , and parameters characterizing the motion of the patch of DNA on which the protein diffused while it was bound to DNA. We did this for time series of length  $N \geq 50$  using the MLE that takes the DNA's motion into account (Sec. VI A). Figure 6 shows the covariance and periodogram of each coordinate in a single such time series. The theoretically expected values for the same covariances and periodograms are also shown. They were calculated using parameter values obtained with MLE.

We also estimated diffusion coefficients from the same time series using CVE, which does not take substrate fluctuations into account (Sec. III C).

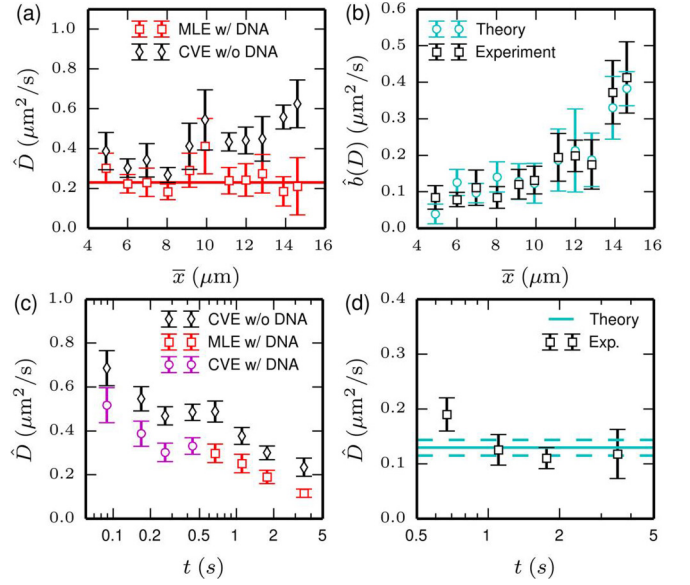


FIG. 9. (Color online) Estimated diffusion coefficients of experimentally recorded hOGG1 proteins diffusing on flow-stretched  $\lambda$  DNA. (a) Estimates  $\hat{D}$  of diffusion coefficients versus protein's mean position  $\bar{x}$  on the DNA strand as measured from the DNA's tethered end. Red squares: estimates obtained with MLE that accounts for the DNA's motion (Sec. VI). Values do not depend on the proteins' position on the DNA ( $P = 0.88$ ). Black diamonds: estimates obtained with CVE that does not account for DNA motion (Sec. III C). An increase in values appears near the DNA's free end. (b) Bias of CVE (black squares) compared to its theoretically expected value calculated from Eq. (68) (cyan circles). The bias increases near the DNA's free end, where fluctuations are larger. (c) Estimates  $\hat{D}$  of the diffusion coefficients as a function of the proteins' residence time  $t$  on DNA. A nontrivial correlation is clearly seen between  $\hat{D}$  and  $t$ . (d) The average bias over the DNA strand of the CVE, which does not take DNA fluctuations into account, does not depend significantly on the protein's residence time on DNA and is equal to its theoretically expected value ( $P = 0.27$ ).  $P$  values were estimated using a  $\chi^2$  test for variance (Appendix J).

Results show that diffusion coefficients estimated using the MLE that takes DNA motion into account, do not depend on a protein's position on the DNA [Fig. 9(a)]. This means that neither the DNA's motion nor a protein's average position in the flow field (which depends on a protein's position on the DNA) affect the value of the diffusion coefficient. If one does not take the DNA's fluctuations into account, however, then estimates of  $D$  are severely biased [Fig. 9(a)]. This bias can be estimated experimentally by subtracting the estimates of  $D$  obtained using the MLE that accounts for DNA fluctuations, from estimates of  $D$  obtained with CVE which does not [Fig. 9(b)]. Since a protein's limited residence time on the DNA limits its movement along the DNA to patches of  $1 \mu\text{m}$  or smaller, we thus found the bias as function of position on the DNA with  $1 \mu\text{m}$  resolution of positions. We also calculate the theoretically predicted bias caused by the DNA's fluctuations at these positions [Eq. (68)] as described in Sec. V B. These theoretically predicted values for the bias of the CVE are in excellent agreement with the experimentally measured values of the bias [Fig. 9(b)]. We see that this bias



does not depend significantly on the proteins' residence time on DNA [Fig. 9(d)]. We used this knowledge of the bias of the CVE to correct biased estimates of diffusion coefficients obtained with CVE from time series shorter than  $N = 50$  [Fig. 9(c)]. For such short series MLE is suboptimal, while CVE is optimal after its bias has been removed.

### B. New results from old data

The procedure just described, using MLE for long and CVE for short time series, allowed us to estimate diffusion coefficients at the single-molecule level. Results obtained in this manner revealed a significant negative correlation between a protein's residence time and its diffusion coefficient during its residence [Fig. 9(c)]. The simplest explanation of this correlation is a two-state kinetics of hOGG1 proteins diffusion on DNA [22]. This interpretation of data is independently confirmed by the fact that residence times are not exponentially distributed, as they would be in a one-state model, since detachment in a one-state model is a Poisson process. Residence times are rather distributed according to a double exponential, as predicted by our two-state model [22].

## VIII. CONCLUSION

We have shown theoretically and on Monte Carlo generated data that the optimal way to estimate diffusion coefficients from short, noisy trajectories of freely diffusing particles is with our unbiased and regression-free CVE. We have described how the attractive properties of this estimator hold also for short, noisy single-particle trajectories on a fluctuating substrate, if the substrate is finite, taut, and some longer trajectories on it also are available. We limited ourselves to describing taut substrates because (a) it implies no loss of generality, since observations of diffusion make sense only on rigid or taut substrates; (b) we consequently arrive at a description that is general and applies to a generic substrate, independently of the specific substrate's known or unknown specific dynamics; and (c) we needed that for a specific problem, a strand of DNA that is tethered by one end to a plane and stretched by a shear flow, for which there is no good theoretical model for its dynamics.

We used that if a substrate is sufficiently taut and its motion sufficiently damped, then its dynamics can be linearized to a good approximation. We also used that a finite substrate's independent degrees of freedom in such a linearized description is a countable set of massively overdamped harmonic oscillators in a heat bath, and their spectrum of relaxation times is discrete and, for a small substrate, well separated. We used that particle tracking with finite time lapse, and especially with motion blur, may resolve only the slowest of these relaxation times, while contributions from all higher modes only increased the localization error.

We consequently modeled the observable transverse motion of a given point on a fluctuating DNA substrate as a massively overdamped harmonic oscillator in a heat bath, a model which was confirmed by fits to experimental data. We derived statistics for this motion which account also for effects due to localization error, finite time-lapse recording, and motion blur due to finite camera shutter time. We used that for

a one-dimensional stretched-out but unstretchable substrate, such as DNA, the substrate's longitudinal motion is given by its transverse motion.

This model for the DNA's fluctuations allowed us to derive a MLE of diffusion coefficients, the variance of localization errors, and parameters describing substrate motion, using both the transverse and the longitudinal coordinates of a recorded trajectory of a diffusing particle. Monte Carlo simulations showed that this MLE is optimal for long time series. We used it on long trajectories to derive the bias of the CVE caused by substrate fluctuations, and we demonstrated that theoretical predictions agree with experimental measurements of the bias, which further confirms our model of the substrate's motion.

We then constructed an unbiased CVE for the case in which substrate fluctuations have been characterized independently, and we showed on Monte Carlo generated data and experimental data that it optimally estimates diffusion coefficients.

Finally, we used it to estimate diffusion coefficients from short, noisy single-particle trajectories of hOGG1 proteins on fluctuating, flow-stretched DNA and showed that one severely overestimates diffusion coefficients if substrate fluctuations are not taken into account.

## ACKNOWLEDGMENTS

The experimental data analyzed here were produced in the laboratory of Xiaoliang Sunney Xie in the Department of Chemistry and Chemical Biology at Harvard University. This work is supported by National Institutes of Health Grant No. ST32 GM07598-25 (to P.C.B. through the Harvard University's Molecular, Cellular, and Chemical Biology training program), by Burroughs Wellcome Career Award at the Scientific Interface to P.C.B., by Human Frontier Science Program Research Grant No. GP0054/2009-C (to H.F.), and by Danish Council for Strategic Research Grant No. 10-092322 (PolyNano) (to H.F.).

## APPENDIX A: SUMMARY OF NOTATION

See the Supplemental Materials (Tables I, II, and III) for summaries of acronyms and notation used in main text [23]. A few symbols that are used only within the section in which they are defined are not included in these tables. Neither are the few new symbols introduced in the appendixes below, as each is used only within the appendix in which it appears.

The constant  $c = \exp(-2\pi f_c \Delta t)$  occurs repeatedly in our aliased Lorentzian power spectra for substrate motion. Since these power spectra come in slightly different flavors, so does  $c$ . We have specified the definition of  $c$  in several places where power spectra are described, and we repeat these definitions in the Supplemental Materials (Table III) [23]. Even without that, the following rule of thumb should make confusion impossible: In any power spectrum written using  $c$ , a corner frequency  $f_c$  also occurs explicitly or, in Appendix G, occurs through  $\lambda_{x,k} = 2\pi f_{c,x}^{(k)}$ . It is always *that* corner frequency which enters in the definition of  $c$  used in the same spectrum.

As stated in the Supplemental Materials (Tables II and III) [23], for  $\hat{D}$ ,  $\hat{\sigma}^2$ ,  $\hat{K}_x$ ,  $\hat{\phi}_i$ ,  $\hat{\bar{\phi}}_i$ ,  $\hat{\theta}$ , and  $\hat{f}_{c,x}$  the "hat" over



the symbols has the conventional meaning of “estimator for parameter(s) denoted by the symbol under the hat.” In a few places where no misunderstanding seems possible, we have used one of these symbols with hat to denote the estimate itself in order to distinguish the estimate of a parameter from the parameter itself.

In order to make our notation agree with that of Ref. [31], which the reader may want to consult for details on the power spectrum analysis used here, a hat over a symbol means the discrete-time, finite-time Fourier transform that one applies to finite time series, in the two cases  $\widehat{\Delta x}_f$  and  $\widehat{y}_f$ , while  $\widehat{P}_{\Delta x, f}$  and  $\widehat{P}_y$  denote the power spectra based on these Fourier transforms, and  $\widehat{\epsilon}_{\Delta x, f}$  denotes the normalized periodogram values based on the same Fourier transforms. Similarly, a “tilde”—as in  $\widetilde{p}$ —denotes the continuous-time, finite-time Fourier transform as in [31], while a “check”—as in  $\check{x}_k$ ,  $\check{P}_k$ , and  $\check{\epsilon}_k$ —refers to sine transforms. All of this is stated also in the Supplemental Materials (Table III) [23].

## APPENDIX B: GENERALIZED LEAST-SQUARES ESTIMATOR

The GLS estimator of  $(D, \sigma^2)$  is defined as

$$\begin{pmatrix} \widehat{D} \\ \widehat{\sigma^2} \end{pmatrix} = (\mathbf{A}^T \Sigma_{\text{msd}}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \Sigma_{\text{msd}}^{-1} \overline{\mathbf{d}^2}, \quad (\text{B1})$$

where  $\overline{\mathbf{d}^2} = (\overline{d_1^2}, \overline{d_2^2}, \dots, \overline{d_N^2})$ ,  $\mathbf{A}$  is a  $2 \times N$  matrix with  $A_{1,n} = 2\Delta t(n-2R)$  and  $A_{2,n} = 2$ , and  $\Sigma_{\text{msd}}$  is a weight matrix proportional to the covariance matrix of  $\overline{\mathbf{d}^2}$ , which is given in [19]. Since  $\Sigma_{\text{msd}}$  depends nonlinearly on  $D$  and  $\sigma^2$ , the GLS of  $(D, \sigma^2)$  must be found using an iterative relaxation algorithm as described in [46].

If  $\Sigma_{\text{msd}}$  is known *a priori*, the variance of the GLS estimator is found by taking the expected value of the outer product of Eq. (B1),

$$\begin{aligned} \text{var}(\widehat{\theta}) &= (\mathbf{A}^T \Sigma_{\text{msd}}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \text{var}(\overline{\mathbf{d}^2}) \mathbf{A} (\mathbf{A}^T \Sigma_{\text{msd}}^{-1} \mathbf{A})^{-1} \\ &= (\mathbf{A}^T \Sigma_{\text{msd}}^{-1} \mathbf{A})^{-1}, \end{aligned} \quad (\text{B2})$$

where  $\widehat{\theta} = (\widehat{D}, \widehat{\sigma^2})^T$ . The iterative GLS estimator uses and estimates  $\widehat{\Sigma}_{\text{msd}}$  instead of the true covariance matrix  $\Sigma_{\text{msd}}$ . The estimated covariance matrix  $\widehat{\Sigma}_{\text{msd}}$  is correlated with  $\overline{\mathbf{d}^2}$ , which means that variance of the iterative GLS does not reduce to  $(\mathbf{A}^T \Sigma_{\text{msd}}^{-1} \mathbf{A})^{-1}$  as in Eq. (B2) and that the variance of the iterative GLS estimator is higher than the variance of the linear GLS estimator [Figs. 1(c) and 1(d)]. The difference is not of order  $\mathcal{O}_p(1/N^2)$ , since  $\overline{\mathbf{d}^2}$  is not bounded; i.e., the variance of the GLS estimator does not necessarily approach the theoretical value [Eq. (B2)] in the large  $N$ -limit.

When  $\sigma^2$  is known *a priori*, the GLS estimator can be reduced to

$$\widehat{D} = (\mathbf{a}^T \Sigma_{\text{msd}-2\sigma^2}^{-1} \mathbf{a})^{-1} \mathbf{a}^T \Sigma_{\text{msd}-2\sigma^2}^{-1} (\overline{\mathbf{d}^2} - 2\widehat{\sigma^2}), \quad (\text{B3})$$

where  $a_n = 2\Delta t(n-2R)$  and

$$\Sigma_{\text{msd}-2\sigma^2} = \text{var}(\overline{\mathbf{d}^2} - 2\widehat{\sigma^2}) = \Sigma_{\text{msd}} + 4\text{var}(\widehat{\sigma^2}). \quad (\text{B4})$$

Even when the noise amplitude  $\sigma$  is known *a priori*,  $\Sigma_{\text{msd}}$  depends nonlinearly on the unknown parameter  $D$ . So one

still needs to use an iterative procedure to find the GLS estimate. The estimator in this case, however, practically reaches the Cramér-Rao lower bound and is unbiased [Figs. 1(e) and 1(f)].

## APPENDIX C: STATISTICS OF COVARIANCES AND THE CVE FOR DIFFUSION ON A SUBSTRATE AT REST

We here derive expressions for the variance of  $\overline{\Delta x_n \Delta x_{n+j}}$  (Sec. 1 of this appendix), the variance of the CVE of  $D$  and  $\sigma^2$  (Sec. 2 of this appendix), and the characteristic function of the CVE of  $D$  (Sec. 3 of this appendix). Finally, we elaborate on the difference between the CVE of  $D$  and the minimal mean-squared-displacement-based estimator that uses only the first two MSDs ( $d_1^2$ , and  $d_2^2$ ) to estimate  $D$  (Sec. 4 of this appendix).

### 1. Variance of estimated covariances

The variance of estimated covariances  $\overline{\Delta x_n \Delta x_{n+j}}$  is derived directly from their definition,  $\overline{\Delta x_n \Delta x_{n+j}} = \sum_{m=1}^{N-j} \Delta x_m \Delta x_{m+j} / (N-j)$ . This gives

$$\begin{aligned} \text{var}(\overline{\Delta x_n \Delta x_{n+j}}) &= \frac{1}{(N-j)^2} \sum_{m,n=1}^{N-j} \text{cov}(\Delta x_m \Delta x_{m+j}, \Delta x_n \Delta x_{n+j}), \end{aligned} \quad (\text{C1})$$

where  $\text{cov}(x, y)$  denotes the covariance of  $x$  and  $y$ . For  $j > 1$ ,  $\langle \Delta x_n \Delta x_{n+j} \rangle = \langle \Delta x_m \Delta x_{m+j} \rangle \langle \Delta x_{m+j} \Delta x_n \rangle = 0$ . So

$$\text{var}(\overline{\Delta x_n \Delta x_{n+j}}) = \frac{1}{(N-j)^2} \sum_{m,n=1}^{N-j} \langle \Delta x_m \Delta x_n \rangle^2. \quad (\text{C2})$$

Using Eqs. (2)–(4) and the definitions of  $\alpha$  and  $\beta$  [23, Table II], we have

$$\begin{aligned} \text{var}(\overline{\Delta x_n \Delta x_{n+j}}) &= \frac{1}{N-j} \langle (\Delta x_n)^2 \rangle^2 \\ &\quad + \frac{2(N-j-1)}{(N-j)^2} \langle \Delta x_n \Delta x_{n+1} \rangle^2 \\ &= \frac{\alpha^2 + 4\alpha\beta + 6\beta^2}{N-j} - \frac{2\beta^2}{(N-j)^2}, \end{aligned} \quad (\text{C3})$$

i.e., Eq. (9) with  $j = |n-m|$ .

### 2. Variance of the CVE

The variance of the CVE of  $D$  is derived in a straightforward manner from Eq. (14),

$$\begin{aligned} \text{var}(\widehat{D}) &= \frac{\text{var}[(\overline{\Delta x_n})^2]}{4(\Delta t)^2} + \frac{\text{var}(\overline{\Delta x_n \Delta x_{n+1}})}{(\Delta t)^2} \\ &\quad + \frac{\text{cov}\left((\overline{\Delta x_n})^2, \overline{\Delta x_n \Delta x_{n+1}}\right)}{(\Delta t)^2}. \end{aligned} \quad (\text{C4})$$

The individual terms of Eq. (C4) are:

$$\text{var}[(\overline{\Delta x_n})^2] = \frac{2\alpha^2 + 8\alpha\beta + 12\beta^2}{N} - \frac{4\beta^2}{N^2}, \quad (\text{C5})$$

where we have used that  $\text{var}[(\Delta x_n)^2] = 2(\alpha + 2\beta)^2$ ,  $\text{cov}((\Delta x_n)^2, (\Delta x_{n+1})^2) = 2\beta^2$ , and  $\text{cov}((\Delta x_n)^2, (\Delta x_m)^2) = 0$  for  $m > n + 1$ ;

$$\text{var}(\overline{\Delta x_n \Delta x_{n+1}}) = \frac{\alpha^2 + 4\alpha\beta + 7\beta^2}{N-1} - \frac{2\beta^2}{(N-1)^2}, \quad (\text{C6})$$

since  $\text{var}(\Delta x_n \Delta x_{n+1}) = (\alpha + 2\beta)^2 + \beta^2$ ,  $\text{cov}(\Delta x_n \Delta x_{n+1}, \Delta x_{n+1} \Delta x_{n+2}) = \beta^2$ , and  $\text{cov}(\Delta x_n \Delta x_{n+1}, \Delta x_m \Delta x_{m+1}) = 0$

for  $m > n + 1$ ;

$$\text{cov}(\overline{(\Delta x_n)^2}, \overline{\Delta x_n \Delta x_{n+1}}) = -\frac{4\alpha\beta + 8\beta^2}{N}, \quad (\text{C7})$$

since  $\text{cov}((\Delta x_n)^2, \Delta x_n \Delta x_{n\pm 1}) = 2\alpha\beta - 4\beta^2$  and  $\text{cov}((\Delta x_n)^2, \Delta x_m \Delta x_{m+1}) = 0$  for  $m > n$ .

Inserting Eqs. (C5)–(C7) in Eq. (C4) gives Eq. (17). The variance of  $\hat{\sigma}^2$  and the covariance of  $\hat{D}$  and  $\hat{\sigma}^2$  are derived in the same manner and are, to second order in  $1/N$ ,

$$\begin{aligned} \text{var}(\hat{\sigma}^2) &= \frac{(1 - 4R + 6R^2)\alpha^2 + 4(1 - 2R + 2R^2)\alpha\beta + (7 - 12R + 8R^2)\beta^2}{N} \\ &\quad + \frac{(1 - 2R)^2\alpha^2 + 4(1 - 2R)^2\alpha\beta + (5 - 20R + 16R^2)\beta^2}{N^2}, \end{aligned} \quad (\text{C8})$$

and

$$\text{cov}(\hat{D}, \hat{\sigma}^2) = -\frac{(1 - 2R)(\alpha^2 + 2\alpha\beta + 3\beta^2) + \alpha^2/2 - \beta^2}{N\Delta t} - \frac{(1 - 2R)(\alpha + 2\beta)^2 + \beta^2}{N^2\Delta t}. \quad (\text{C9})$$

The variance for the case when the localization error has been characterized independently [Eq. (18)] is found in the same manner.

### 3. Characteristic function of the CVE

The characteristic function of the CVE,  $\tilde{p}$ , is given by

$$\begin{aligned} \ln \tilde{p} &= \frac{N+1}{2} \ln \left[ \frac{4}{A} \right] + \frac{1}{4} \ln[(A+C)^2 - B^2] \\ &\quad - \frac{1}{2}(S_{1,+} + S_{1,-} + S_{2,+} + S_{2,-}), \end{aligned} \quad (\text{C10})$$

where

$$S_{1,\pm} = \frac{N+1}{2} \ln \left[ 1 + \sqrt{1 - \left( \frac{B \pm \sqrt{B^2 - 4AC}}{2A} \right)^2} \right], \quad (\text{C11})$$

$$S_{2,\pm} = \frac{1}{2} \ln \left\{ 1 - \left[ \frac{1 - \sqrt{1 - \left( \frac{B \pm \sqrt{B^2 - 4AC}}{2A} \right)^2}}{1 + \sqrt{1 - \left( \frac{B \pm \sqrt{B^2 - 4AC}}{2A} \right)^2}} \right]^{N+1} \right\}, \quad (\text{C12})$$

and  $A$ ,  $B$ , and  $C$  are functions of  $\omega$ ,

$$A(\omega) = 1 + \frac{2\omega D}{N}(1 + \varepsilon), \quad (\text{C13a})$$

$$B(\omega) = \frac{2\omega D}{N-1} \left[ 2 + \left( 1 + \frac{1}{N} \right) \varepsilon \right], \quad (\text{C13b})$$

$$C(\omega) = -\frac{4\omega D}{N} \varepsilon. \quad (\text{C13c})$$

### 1. Derivation

To derive the characteristic function of the CVE  $\hat{D}$  of  $D$ , we first note the trivial equality for the distribution of  $\hat{D}$  around the true value  $D$ ,

$$\begin{aligned} p(\hat{D}|D, \sigma^2) &= \int p(\Delta \mathbf{x}|D, \sigma^2) \delta \\ &\quad \times \left[ \hat{D} - \frac{(\overline{\Delta x_i})^2 + 2\overline{\Delta x_i \Delta x_{i+1}}}{2\Delta t} \right] \mathcal{D}\Delta \mathbf{x}, \end{aligned} \quad (\text{C14})$$

where  $\Delta \mathbf{x} = (\Delta x_1, \Delta x_2, \dots, \Delta x_N)^T$ , and  $\mathcal{D}\Delta \mathbf{x} = \prod_{i=1}^N d\Delta x_i$ . We use the Dirac  $\delta$  function written in terms of its Fourier transform,  $\delta(x - x_0) = \int_{-\infty}^{\infty} e^{i(x-x_0)\omega} d\omega / (2\pi)$ , and that  $\overline{(\Delta x_n)^2} / 2 = \Delta \mathbf{x}^T \mathbf{I} \Delta \mathbf{x} / N$  and  $\overline{\Delta x_n \Delta x_{n+1}} = \Delta \mathbf{x}^T \mathbf{C} \Delta \mathbf{x} / (N-1)$ , where  $\mathbf{I}$  is the identity matrix and  $C_{ij} = \delta_{i,j\pm 1}$ . This gives

$$\begin{aligned} p(\hat{D}|D, \sigma^2) &= \int_{-\infty}^{\infty} \frac{e^{i\hat{D}\omega}}{2\pi} \int \frac{e^{-\frac{1}{2}\Delta \mathbf{x}^T \{ \Sigma_{\Delta x}^{-1} + \frac{i\omega}{\Delta t} (\frac{1}{N} + \frac{C}{N-1}) \} \Delta \mathbf{x}}}{(2\pi)^{N/2} \sqrt{\text{Det} \Sigma_{\Delta x}}} \\ &\quad \times \mathcal{D}\Delta \mathbf{x} d\omega, \end{aligned} \quad (\text{C15})$$

where the covariance matrix of  $\Delta \mathbf{x}$ ,  $\Sigma_{\Delta x}$ , is given by Eqs. (2)–(4).

We use that both  $\Sigma_{\Delta x}$  and  $\mathbf{C}$  can be diagonalized using a normalized version of the DST given by the orthogonal transformation matrix  $\mathbf{U}$  with entries

$$U_{ij} = \sqrt{\frac{2}{N+1}} \sin \left( \frac{\pi i j}{N+1} \right). \quad (\text{C16})$$

We define  $c_k = [\mathbf{U}^{-1} \mathbf{C} \mathbf{U}]_{kk} = 2 \cos \theta_k$  and note that  $P_k = \langle \hat{P}_k \rangle = \Delta t [\mathbf{U}^{-1} \Sigma_{\Delta x} \mathbf{U}]_{kk}$ , where  $\theta_k = \frac{\pi k}{N+1}$ . Then

$$p(\hat{D}|D, \sigma^2) = \int_{-\infty}^{\infty} \frac{e^{i\hat{D}\omega} d\omega}{2\pi \sqrt{\prod_{k=1}^N \lambda_k(\omega)}}, \quad (\text{C17})$$

where

$$\lambda_k(\omega) = 1 + \frac{i\omega P_k}{(\Delta t)^2} \left( \frac{1}{N} + \frac{c_k}{N-1} \right). \quad (\text{C18})$$

The characteristic function of the CVE,  $\tilde{p}$ , is defined as the Fourier transform of the probability density  $p$ . So  $\tilde{p}(\omega) = \prod_{k=1}^N \lambda_k(\omega)^{-\frac{1}{2}}$ .

From Eq. (C18) we see that  $\lambda_k$  is a second-degree polynomial in  $\cos \theta_k$ . So it can be written in the form  $\lambda_k = A + B \cos \theta_k + C \cos^2 \theta_k$ , where  $A$ ,  $B$ , and  $C$  are given in Eq. (C13). The logarithm of the characteristic function is thus given by

$$-2 \ln \tilde{p} = \sum_{k=1}^N \ln(A + B \cos \theta_k + C \cos^2 \theta_k). \quad (\text{C19})$$

We use that

$$\sum_{k=1}^N f(\cos \theta_k) = -\frac{1}{2}[f(1) + f(-1)] + \frac{1}{2} \sum_{k=1}^{2N+2} f(\cos \theta_k) \quad (\text{C20})$$

to rewrite Eq. (C19) as

$$\begin{aligned} -2 \ln \tilde{p} &= -\frac{1}{2} \ln[(A + B + C)(A - B + C)] \\ &\quad + \frac{1}{2} \sum_{k=1}^{2N+2} \ln \lambda_k. \end{aligned} \quad (\text{C21})$$

The last term can be further simplified,

$$\begin{aligned} \frac{1}{2} \sum_{k=1}^{2N+2} \ln \lambda_k &= \frac{1}{2} \sum_{k=1}^{2N+2} \ln \left[ C \left( \cos \theta_k + \frac{a_+}{2C} \right) \left( \cos \theta_k + \frac{a_-}{2C} \right) \right] \\ &= \frac{1}{2} \sum_{k=1}^{2N+2} [\ln(a_+ + 2C \cos \theta_k) + \ln(a_- + 2C \cos \theta_k)] \\ &\quad - (N+1) \ln(4C), \end{aligned} \quad (\text{C22})$$

where we have defined  $a_{\pm} = B \pm \sqrt{B^2 - 4AC}$ .

To simplify the last two terms of Eq. (C22), we define the sum

$$\begin{aligned} S(a, b) &= \frac{1}{2} \sum_{k=1}^{2N+2} \ln(a + b \cos \theta_k) \\ &= \frac{1}{2} \sum_{k=1}^{2N+2} \ln[c(\beta + e^{i\theta_k})(\beta + e^{-i\theta_k})] \\ &= (N+1) \ln \left( \frac{b\beta_{\pm}}{2} \right) + \frac{1}{2} \sum_{k=1}^{2N+2} \ln \left( 1 + \frac{e^{i\theta_k}}{\beta_{\pm}} \right) \\ &\quad + \frac{1}{2} \sum_{k=1}^{2N+2} \ln \left( 1 + \frac{e^{-i\theta_k}}{\beta_{\pm}} \right), \end{aligned} \quad (\text{C23})$$

with  $c = \frac{b}{2\beta}$  and  $\beta_{\pm} = \frac{a}{b} \pm \sqrt{(\frac{a}{b})^2 - 1}$ . The two last terms in Eq. (C23) can be rewritten as

$$\begin{aligned} \sum_{k=1}^{2N+2} \ln \left( 1 + \frac{e^{\pm i\theta_k}}{\beta_{\pm}} \right) &= - \sum_{k=1}^{2N+2} \sum_{l=1}^{\infty} \frac{\left( -\frac{e^{\pm i\theta_k}}{\beta_{\pm}} \right)^l}{l} \\ &= - \sum_{l=1}^{\infty} \frac{(-\beta_{\pm})^{-l}}{l} \sum_{k=1}^{2N+2} e^{\pm il\theta_k}. \end{aligned} \quad (\text{C24})$$

The sum  $\sum_{k=1}^{2N+2} e^{\pm il\theta_k}$  is equal to zero because its argument makes a full circle in the complex plane, except when  $l$  is equal to an integer times  $2N+2$ . Thus,

$$\begin{aligned} \sum_{k=1}^{2N+2} \ln \left( 1 + \frac{e^{\pm i\theta_k}}{\beta_{\pm}} \right) &= -2(N+1) \sum_{m=1}^{\infty} \frac{[(-\beta_{\pm})^{-2(N+1)}]}{2(N+1)m} \\ &= \ln \left( 1 - \beta_{\pm}^{-2(N+1)} \right). \end{aligned} \quad (\text{C25})$$

Inserting Eq. (C25) in Eq. (C23) gives

$$\begin{aligned} S(a, b) &= (N+1) \ln \left( \frac{b\beta_{\pm}}{2} \right) + \ln(1 - \beta_{\pm}^{-2(N+1)}) \\ &= \ln[(a \pm \sqrt{a^2 - b^2})^{N+1} - (a \mp \sqrt{a^2 - b^2})^{N+1}] \\ &\quad - (N+1) \ln 2, \end{aligned} \quad (\text{C26})$$

where we have used that  $\beta_+ \beta_- = 1$ .

We insert this result into Eq. (C22) and use Eq. (C21) to get

$$\begin{aligned} \ln \tilde{p} &= \frac{1}{2} (N+1) \ln(4C) + \frac{1}{4} \ln[(A+C)^2 - B^2] \\ &\quad - \frac{1}{2} S(a_+, 2C) - \frac{1}{2} S(a_-, 2C), \end{aligned} \quad (\text{C27})$$

with  $a_{\pm} = B \pm \sqrt{B^2 - 4AC}$ .

To avoid numerical problems, we rewrite Eq. (C27) to get Eq. (C10).

#### 4. Difference between CVE and MSD

The CVE of the diffusion coefficient  $D$  was constructed as a maximally simple unbiased estimator of  $D$  of a particle based on estimates of the covariance function of the particle's recorded single-time-lapse displacements. In this spirit it is also possible to construct a maximally simple unbiased estimator  $\hat{D}_{\text{msd}}$  of  $D$  based on estimates of the first two MSDs,  $\overline{d_1^2}$  and  $\overline{d_2^2}$ . While this estimator, given by  $\hat{D}_{\text{msd}} = (\overline{d_2^2} - \overline{d_1^2}) / (2\Delta t)$ , has the same expected value as the CVE of  $D$ —they are both unbiased—the two estimators have different precision,  $\hat{D}_{\text{msd}}$  being less precise than CVE. This is seen as follows.

From Eq. (8) we have

$$\begin{aligned} 2\Delta t \hat{D}_{\text{msd}} &= \frac{(\Delta x_1)^2 + (\Delta x_N)^2 + (N+1) \sum_{n=2}^{N-1} (\Delta x_n)^2}{N(N-1)} \\ &\quad + \frac{2 \sum_{n=1}^{N-1} \Delta x_n \Delta x_{n+1}}{N-1}, \end{aligned} \quad (\text{C28})$$

while Eq. (14) gives us

$$2\Delta t \hat{D}_{\text{cve}} = \frac{\sum_{n=1}^N (\Delta x_n)^2}{N} + \frac{2 \sum_{n=1}^{N-1} \Delta x_n \Delta x_{n+1}}{N-1}. \quad (\text{C29})$$

Comparison of Eqs. (C28) and (C29) reveals the subtle difference between  $\hat{D}_{\text{msd}}$  and  $\hat{D}_{\text{cve}}$ : while the CVE of  $D$  weights all the squared displacements evenly [first term in Eq. (C29)], the maximally simple MSD-based estimator weights the contributions from the displacements at the start and the end of the time series,  $\Delta x_1$  and  $\Delta x_N$ , by only  $1/(N-1)$  and the rest of the displacements by  $(N+1)/(N-1)$  [first term in Eq. (C28)]. These unequal weights on equivalent terms represent a suboptimal use of information: A simple calculation proves that it gives  $\hat{D}_{\text{msd}}$  a larger variance than  $\hat{D}_{\text{cve}}$  has.

#### APPENDIX D: TRANSFORMING THE MEASURED DATA INTO STATISTICALLY INDEPENDENT DATA

We here show how discrete sine transformation of the measured set of diffusive displacements on a substrate at rest produces statistically independent data. Exact statistical tests and practical success with maximum likelihood estimation (Sec. 1 of this appendix) build on this simplification. We then review how DFT produces approximately independent data (independent to order  $1/N$ ) for a much larger class of data (Sec. 2 of this appendix).

##### 1. Discrete sine transformation produces independent data for diffusion on a substrate at rest

The covariance matrix for the set  $\Delta \mathbf{x} = (\Delta x_1, \Delta x_2, \dots, \Delta x_N)$  of measured displacements of a particle diffusing in a medium or on a substrate at rest [Eqs. (2)–(4)] is a symmetric, tridiagonal Toeplitz matrix. It can be written

$$\Sigma_{\Delta x} = [2(1 - 2R)D\Delta t + 2\sigma^2]\mathbf{I} - (\sigma^2 - 2RD\Delta t)\mathbf{M}, \quad (\text{D1})$$

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{M}_{ij} = \delta_{i,j\pm 1}$ . It is easy to verify that the set of  $N$  vectors  $(\mathbf{q}^{(k)})_{k=1}^N$  with entries

$$q_n^{(k)} = \Delta t \sin[\pi kn/(N+1)] \quad (\text{D2})$$

are eigenvectors of  $\mathbf{M}$  and consequently eigenvectors of  $\Sigma_{\Delta x}$ . The associated eigenvalues are

$$\lambda_k = 2\{1 - \cos[\pi k/(N+1)]\}. \quad (\text{D3})$$

Comparison of Eq. (D2) with Eq. (11) reveals that the matrix  $\mathbf{M}^{-1}$  composed of  $(\mathbf{q}^{(k)})_k$  as columns is the matrix equivalent of the DST, i.e.,  $\mathbf{M}^{-1}\Delta \mathbf{x} = \widetilde{\Delta \mathbf{x}}$ . Consequently,

$$\langle \widetilde{\Delta \mathbf{x}} \widetilde{\Delta \mathbf{x}}^T \rangle = \mathbf{M}^{-1} \Sigma_{\Delta x} (\mathbf{M}^{-1})^T = \mathbf{P}, \quad (\text{D4})$$

where  $\mathbf{P}$  is a diagonal matrix with diagonal entries  $(P_k)_{k=1}^N$  given in Eq. (10). This proves that  $\widetilde{\Delta x}_k$  and  $\widetilde{\Delta x}_l$  are uncorrelated for  $k \neq l$ . Since  $(\Delta x_n)_n$  are Gaussian distributed and independent,  $(\Delta x_k)_k$ , being linear combinations of  $(\Delta x_n)_n$ , are jointly Gaussian distributed. This implies that  $\widetilde{\Delta x}_k$  and  $\widetilde{\Delta x}_l$  are statistically independent for  $k \neq l$ , the reason being that uncorrelated jointly Gaussian distributed variables are independent.

##### 2. Discrete Fourier transformation approximately removes statistical dependencies in data

We briefly review how DFT of a stochastic time series results in approximately independent data under quite general assumptions. These assumptions are satisfied by time-lapse recorded displacements of a single diffusing particle.

We consider a zero-mean stationary stochastic process  $z(t)$ , which we sample at constant time lapse  $\Delta t$ . The autocovariance of the resulting sequence  $\mathbf{z} = (z_1, z_2, \dots, z_N)$  depends only on the relative separation in time,

$$\langle z_m z_n \rangle = C_{|m-n|}. \quad (\text{D5})$$

We assume that  $j|C_j|$  is summable, i.e.,

$$\sum_{j=0}^{\infty} j|C_j| < \infty. \quad (\text{D6})$$

This assumption is satisfied by time series with, e.g., finite or exponentially decreasing autocovariance. Consequently, it is satisfied by stationary series from linear dissipative systems driven by white noise, such as diffusion on a substrate that is fluctuating or at rest.

Let  $\hat{\mathbf{z}}$  denote the DFT of  $\mathbf{z}$ . The off-diagonal elements of the correlation matrix of  $\hat{\mathbf{z}}/\sqrt{N}$  are bounded by  $2 \sum_{j=0}^N j|C_j|/N$ ; thus, by Eq. (D6), they vanish except for terms of order  $1/N$ :  $\langle \hat{z}_k \hat{z}_l^* \rangle / N = \mathcal{O}(1/N)$  for  $k \neq l$ , while  $\langle |\hat{z}_k|^2 \rangle / N = \mathcal{O}(1)$  [47]. Consequently, for large  $N$ , (almost) all information in the time series is contained in the periodogram  $\hat{P}_k = |\hat{z}_k|^2 / (N\Delta t)$ , and we can ignore *correlations* between different spectral values. In general, absence of correlations does not imply independence. However, if  $(z_1, z_2, \dots, z_N)$  are jointly Gaussian distributed—as is the case for our time series—so are the real and imaginary parts of  $(\hat{z}_1, \hat{z}_2, \dots, \hat{z}_N)$  (modulo redundancies due to  $\mathbf{z}$  being real), and *then* they are independent to the extent that they are uncorrelated. Thus,  $\hat{z}_k/\sqrt{N}$  and  $\hat{z}_l/\sqrt{N}$  are independent to lowest order in  $1/N$  for  $k \neq l$ .

We do not need to assume that  $(z_1, z_2, \dots, z_N)$  are jointly Gaussian distributed, however. A central limit theorem for the DFT [48] only requires that  $(z_1, z_2, \dots)$  is an ergodic process—and ours is—to ensure that  $(\hat{z}_k/\sqrt{N})_{k=\dots}$  are *independent* and Gaussian distributed to order  $1/N$ . Since the periodogram values are proportional to  $|\hat{z}_k|^2/N$ , they thus are independent, except for terms of order  $1/N^2$  or higher.

#### APPENDIX E: MAXIMUM LIKELIHOOD ESTIMATOR FOR DIFFUSION ON A SUBSTRATE AT REST

We here derive an efficient algorithm for maximum likelihood estimation (Sec. 1 of this appendix) and investigate the source of the bias of the MLE (Sec. 2 of this appendix) for diffusion in a medium or on a substrate at rest.

##### 1. Fast MLE algorithm

Finding the MLE of  $D$  and  $\sigma^2$  is a two-dimensional optimization problem. We reduce it to a one-dimensional problem by using the scale-invariance of diffusion. To this end, we introduce two new parameters  $(\lambda, \phi)$ ,

$$\lambda = (2D\Delta t + 2\sigma^2)\Delta t, \quad \phi = \text{arccot}(\text{SNR}). \quad (\text{E1})$$

The parameter  $\phi$  obviously is a measure of the SNR. The parameter  $\lambda$  describes the overall scale of power spectra and



periodograms for  $R = 0$ , i.e., in the absence of motion blur. Specifically,  $\lambda$  is the power spectral density of free diffusion (recorded with localization errors) evaluated in the middle of its frequency range, at  $f = f_{\text{Nyq}}/2$ ; see Eq. (13).

With  $P_k = \langle \tilde{P}_k \rangle$ , Eq. (E1) used in Eq. (10) gives  $P_k = \lambda F_k(\phi)$ , where  $F_k(\phi) = a_k \cos^2 \phi + b_k \sin^2 \phi$  with  $a_k = 1 - 2R(1 - \cos \frac{\pi k}{N+1})$  and  $b_k = 1 - \cos \frac{\pi k}{N+1}$ . The log likelihood then reads

$$\ln \mathcal{L}((\lambda, \phi | (\Delta x_n)_{n=1}^N)) = -\frac{1}{2} \sum_{k=1}^N \left\{ \frac{(\widetilde{\Delta x_k})^2}{\lambda F_k(\phi)} + \ln[\lambda F_k(\phi)] \right\}. \quad (\text{E2})$$

Stationarity of this log likelihood with respect to  $\lambda$  requires that

$$\lambda = \lambda(\phi) = \frac{1}{N} \sum_{k=1}^N \frac{(\widetilde{\Delta x_k})^2}{F_k(\phi)}. \quad (\text{E3})$$

We use this result to eliminate  $\lambda$  in the log-likelihood function in Eq. (E2) and are left with a one-dimensional optimization problem. The value for  $\phi$  that solves this problem for a given time series is our estimate  $\hat{\phi}$ . Once it has been found by numerical maximization of Eq. (E2), Eq. (E3) gives the MLE for  $\hat{\lambda}$ . With  $\hat{\phi}$  and  $\hat{\lambda}$  thus determined, the MLE for  $(D, \sigma^2)$  is

$$\hat{D} = \frac{\hat{\lambda} \cos^2 \hat{\phi}}{2D(\Delta t)^2}, \quad \hat{\sigma}^2 = \frac{\hat{\lambda} \sin^2 \hat{\phi}}{2\Delta t}. \quad (\text{E4})$$

The one-dimensional optimization problem for  $\phi$  is considerably easier for a computer to solve than the two-dimensional problem of maximizing the likelihood with respect to  $D$  and  $\sigma^2$ . In practice, this reduction of dimension only speeds up the MLE algorithm by a factor of two to three, since each calculation of the likelihood function takes longer, as it now involves a double sum and calls to transcendental functions. However, the reduction to one dimension allows us to visualize the likelihood landscape as a simple plot of a function of one parameter,  $\phi$  [Fig. 3(b)].

## 2. Bias of the MLE

Monte Carlo simulations show that the MLEs for  $D$  and  $\sigma^2$  are biased [Figs. 4(a) and 4(b)]. We show here that this is so because they do not admit negative values for  $D$  and  $\sigma^2$ . Such negative values would be physically meaningless, so it makes sense that the two estimators by design avoid negative values. The cost of this meaningful design is bias, however.

A common source of bias in ML estimators is asymmetry of the likelihood function with respect to the true values, e.g., skewness. We derive an approximation to the bias of the MLE due to skewness by Taylor expanding the stationarity condition  $\partial_{\theta_i} \ln[\mathcal{L}(\theta)] = 0$ , where  $\theta = (D, \sigma^2)^T$  and  $\partial_{\theta_i}$  is the partial derivative with respect to  $\theta_i$ .

To order  $1/N$ , this bias is

$$b_i(\theta) = - \sum_{jkm} \mathcal{I}^{ij} \mathcal{I}^{km} \left\{ \left[ \partial_{\theta_j} \partial_{\theta_k} \ln[\mathcal{L}(\theta)] \right] \partial_{\theta_m} \ln[\mathcal{L}(\theta)] \right\} + \frac{1}{2} \left\{ \partial_{\theta_j} \partial_{\theta_k} \partial_{\theta_m} \ln[\mathcal{L}(\theta)] \right\}, \quad (\text{E5})$$

where  $\mathcal{I}$  is the Fisher information matrix and  $\mathcal{I}^{ij} \equiv (\mathcal{I}^{-1})_{ij}$ . The Fisher information matrix is determined using Eq. (22). Since  $\mathcal{L}$  is Gaussian and  $P_k$  is a first-degree polynomial in  $D$  and  $\sigma^2$ , we have that

$$\langle \partial_{\theta_i} \partial_{\theta_j} \partial_{\theta_k} \ln[\mathcal{L}(\theta)] \rangle = -2 \sum_{n=1}^N \frac{\partial_{\theta_i} P_n \partial_{\theta_j} P_n \partial_{\theta_k} P_n}{P_n^3}$$

and

$$\langle \partial_{\theta_i} \partial_{\theta_j} \ln[\mathcal{L}(\theta)] \partial_{\theta_k} \ln[\mathcal{L}(\theta)] \rangle = \sum_{n=1}^N \frac{\partial_{\theta_i} P_n \partial_{\theta_j} P_n \partial_{\theta_k} P_n}{P_n^3}.$$

Thus, the bias due to skewness is zero (to order  $1/N$ ), but we introduce a bias when we require  $\hat{D}$  and  $\hat{\sigma}^2$  to be positive. For a given experimental realization of the measurements  $\Delta \mathbf{x}$ , there is a finite probability that the maximum of  $\ln \mathcal{L}$  is found at a physically meaningless negative value of either  $D$  or  $\sigma^2$ . This probability is not symmetric in  $(D, \sigma^2)$ , the maximum is more likely to be in the region  $\sigma^2 < 0$  for high SNR and more likely to be in the region  $D < 0$  for low SNR. Since we require  $D$  and  $\sigma^2$  to be positive, we introduce positive bias in  $\hat{\sigma}^2$  and negative bias in  $\hat{D}$  for high SNR, and vice versa for low SNR. Since this bias stems from the dispersion of the measured likelihood function around its true maximum, we expect this bias to be of the same order as the standard error of the estimates, i.e., that it decreases as  $N^{-1/2}$ . This is confirmed by numerical results [Figs. 4(a) and 4(b)].

## APPENDIX F: THE EFFECT OF FINITE TIME-LAPSE RECORDING AND MOTION BLUR ON A POWER SPECTRUM

In order to obtain a power spectrum from an experimentally measured time series  $(z_n)_{n=1}^N$ , we compute the DFT of the time series as

$$\hat{z}_k = \Delta t \sum_{n=0}^{N-1} e^{-i2\pi kn/N} z_n. \quad (\text{F1})$$

Here the mode number  $k$  corresponds to the frequency  $f_k = k\Delta f$ . From Eq. (F1) and the identity  $z_n = \frac{1}{\Delta t} \int_0^{\Delta t} z(t_n - t) dt$ , which gives  $z_n$  with motion blur due to a shutter time equal to the time lapse, we have

$$\begin{aligned} \hat{z}_k &= \sum_{n=0}^{N-1} e^{-i2\pi kn/N} \int_0^{\Delta t} z(t_n - t) dt \\ &= \sum_{n=0}^{N-1} e^{-i2\pi kn/N} \int_0^{\Delta t} \frac{1}{t_N} \sum_{k'=-\infty}^{\infty} e^{i2\pi k'(t_n - t)/t_N} \tilde{z}_{k'} dt \\ &= \frac{1}{t_N} \sum_{k'=-\infty}^{\infty} \tilde{z}_{k'} \int_0^{\Delta t} e^{-i2\pi k't/t_N} dt \sum_{n=0}^{N-1} e^{-i2\pi(k-k')n/N} \\ &= \sum_{m=-\infty}^{\infty} \frac{1 - e^{-i2\pi k/N}}{i2\pi(k/N + m)} \tilde{z}_{k+mN}. \end{aligned} \quad (\text{F2})$$

Here we have used the infinite Fourier series

$$z(t) = \frac{1}{t_N} \sum_{k=-\infty}^{\infty} e^{i2\pi kt/t_N} \tilde{z}_k, \quad (\text{F3})$$

where the finite-continuous-time Fourier transforms  $\tilde{z}_k$  are defined as

$$\tilde{z}_k = \int_0^{t_N} z(t) e^{-i2\pi kt/t_N} dt. \quad (\text{F4})$$

We also used that  $\sum_{n=0}^{N-1} e^{i2\pi kn/N} = 0$ , except for  $k = mN$  with  $m \in \mathbb{Z}$ .

We can then express the expected value of the measured periodogram, the power spectrum  $P_f$ , in terms of the power spectrum  $P_f^{(\text{cont})}$  of the corresponding continuous-time process  $z(t)$  as derived in the absence of finite sampling-time effects and localization errors,

$$P_f^{(\text{sub})} = \frac{1 - \cos(2\pi f \Delta t)}{2\pi^2(\Delta t)^2} \sum_{m=-\infty}^{\infty} \frac{P_{f+m f_{\text{sample}}}^{(\text{sub,cont})}}{(f + m f_{\text{sample}})^2}. \quad (\text{F5})$$

This equation gives the expected value of the power spectrum for a time-lapse recorded stationary process measured with maximally open shutter. The same result was found in [49] for time series of infinite duration. We have here shown that it is valid for time series of any finite duration as well. This is no surprise: For a smooth power spectrum without spikes and other abrupt changes in value, leakages [50] is no issue, so finite time of measurement simply makes the spectrum discrete, such that  $f$  only takes values that are integer multiples of  $\Delta f$ , the inverse measurement time, while the spectral values remain unchanged [31].

The infinite sum over  $m$  in Eq. (F5) is a so-called *periodic summation*. It makes  $\tilde{F}^{(\text{per})}$  a periodic function of  $f$  with period  $f_{\text{sample}} = 1/\Delta t$ . It can be evaluated analytically. It is of the form

$$\tilde{F}^{(\text{per})}(f) = \sum_{m \in \mathbb{Z}} \tilde{F}(f + m f_{\text{sample}}). \quad (\text{F6})$$

Since  $P^{(\text{sub,cont})}$  is a Lorentzian, we have, apart from  $m$ -independent factors,

$$\tilde{F}(f) = \frac{1}{f^2(f_c^2 + f^2)} = \frac{1}{f_c^2} \left( \frac{1}{f^2} - \frac{1}{f_c^2 + f^2} \right). \quad (\text{F7})$$

From [31, Sec. VI] we have that

$$\begin{aligned} \sum_{m \in \mathbb{Z}} \frac{1}{f_c^2 + (f + m f_{\text{sample}})^2} \\ = \frac{\pi \Delta t}{f_c} \frac{1 - c^2}{1 + c^2 - 2c \cos(2\pi f \Delta t)}, \end{aligned} \quad (\text{F8})$$

with  $c = \exp(-2\pi f_c \Delta t)$ . By taking the limit  $f_c \rightarrow 0$  on both sides in this result, we also have that

$$\sum_{m \in \mathbb{Z}} \frac{1}{(f + m f_{\text{sample}})^2} = \frac{2\pi^2 \Delta t^2}{1 - \cos(2\pi f \Delta t)}. \quad (\text{F9})$$

Thus,

$$\tilde{F}^{(\text{per})}(f) = \frac{2\pi^2(\Delta t)^2 f_c^{-2}}{1 - \cos(2\pi f \Delta t)} - \frac{\pi \Delta t (1 - c^2) f_c^{-3}}{1 + c^2 - 2c \cos(2\pi f \Delta t)}. \quad (\text{F10})$$

From Eqs. (28) and (F5) we have that

$$P_f^{(\text{sub})} = \frac{K_x [1 - \cos(2\pi f \Delta t)]}{4\pi^4 (\Delta t)^2} \tilde{F}^{(\text{per})}(f), \quad (\text{F11})$$

which with some rearrangement of terms gives Eq. (38).

## APPENDIX G: ESTIMATION FOR DIFFUSION ON A FLUCTUATING SUBSTRATE

We here give estimators of the diffusion coefficient  $D$  and parameters characterizing experimental conditions,  $\sigma^2$  and  $\phi$ , for more general cases than the estimators presented in the main text (Secs. V and VI). Section 1 of this appendix presents a CVE of  $D$  for the case when both  $\phi$  and  $\sigma^2$  have been estimated *a priori*. Section 2 of this appendix gives estimators (both CVE and MLE) for the case when more than one mode is needed to describe the substrate's fluctuations. Finally, Section 3 of this appendix generalizes the estimators given in Section 2 to the case when the camera shutter is only kept open for a fraction of the time lapse.

### 1. CVE for independently characterized localization errors

If the substrate fluctuations and the localization errors both have been characterized independently, the bias of the CVE for free diffusion [Eq. (16)] due to substrate fluctuations is, with  $c = \exp(-2\pi f_{c,x} \Delta t)$ ,

$$b_D(K_x, f_{c,x}) = \frac{4\pi f_{c,x} \Delta t - 3 + 4c - c^2}{(1 - 2R)(2\pi f_{c,x} \Delta t)^3} K_x. \quad (\text{G1})$$

So we can construct an unbiased CVE as

$$\hat{D} = \frac{\overline{(\Delta x_n)^2} - 2\sigma^2}{2(1 - 2R)\Delta t} - b_D(\hat{K}_x, \hat{f}_{c,x}), \quad (\text{G2})$$

with  $b_D$  given by Eq. (G1).

#### a. Diffusion on DNA

In the special case of diffusion on a taut polymer such as DNA (Sec. IV D), the bias is given by

$$b_D(K_x, f_c) = \frac{8\pi f_c \Delta t - 3 + 4c^2 - c^4}{(1 - 2R)(4\pi f_c \Delta t)^3} K_x. \quad (\text{G3})$$

### 2. Including higher modes of substrate fluctuations

#### a. Power spectrum

If multiple modes are needed to describe the substrate's motion, these are included in the MLE by replacing the term of the power spectrum corresponding to substrate fluctuations [Eqs. (40), (50), (51)] with a multimode term.

In the case of diffusion on a generic multidimensional fluctuating substrate [Eq. (40)], the power spectrum then is

$$P_{\Delta x, f} = 2\zeta(\bar{s})^2 D(\Delta t)^2 + [2\sigma^2 \Delta t - 2\zeta(\bar{s})^2 D(\Delta t)^2/3][1 - \cos(2\pi f \Delta t)] + \frac{k_B T}{\gamma} \sum_{k=1}^K \left\{ \frac{2x_k(\bar{s})^2}{\lambda_{x,k}^3 \Delta t c_k} [1 - \cos(2\pi f \Delta t)] \left[ c_k^2 - 1 + 2\lambda_{x,k} \Delta t c_k + \frac{(1+c_k)(1-c_k)^3}{1+c_k^2 - 2c_k \cos(2\pi f \Delta t)} \right] \right\}, \quad (G4)$$

where  $\gamma$  is the substrate's drag coefficient,  $x_k(\bar{s}), k = 1, 2, \dots$  are the spatial eigenfunctions of the substrate's motion with the slowest relaxation rates,  $\lambda_{x,k} = 2\pi f_{c,x}^{(k)}$ , and  $c_k = \exp(-\lambda_{x,k} \Delta t)$ . If the system is linear,  $\lambda_{x,k}$  does not depend on  $\bar{s}$ .  $K$  is chosen such that all modes that contribute significantly to the motion, are included. If no theory exists or is known for the eigenmodes of the substrate motion, the parameters  $(\lambda_{x,k})_{k=1}^K$  and  $[x_k(\bar{s})]_{k=1}^K$  can be fitted as independent parameters for a fixed value of  $\bar{s}$ , i.e., for a given "point" on the substrate. By a point we here mean a range or area or volume with linear extent large enough to contain the observed trajectory of the diffusion particle. We assume that this range or area or volume is small, and hence a point, compared to the characteristic length scales of those eigenmodes of substrate motion which contribute to the observed motion of the tracked particle. We define taut to mean that this condition is satisfied.

For diffusion on a fluctuating, taut, but flexible fiber or polymer, such as DNA, the power spectrum of the transversal laboratory coordinate is

$$P_{y, f} = \sigma^2 \Delta t + \frac{k_B T}{\gamma_{\perp}} \sum_{k=1}^K \left\{ \frac{y_k(\bar{s})^2}{\lambda_k^3 \Delta t c_k} \left[ c_k^2 - 1 + 2\lambda_k \Delta t c_k + \frac{(1+c_k)(1-c_k)^3}{1+c_k^2 - 2c_k \cos(2\pi f \Delta t)} \right] \right\}, \quad (G5)$$

with  $c_k = \exp(-\lambda_k \Delta t)$ , while the power spectrum of displacements in the longitudinal laboratory coordinate is

$$P_{\Delta x, f} = 2\zeta(\bar{s})^2 D(\Delta t)^2 + [2\sigma^2 \Delta t - 2\zeta(\bar{s})^2 D(\Delta t)^2/3][1 - \cos(2\pi f \Delta t)] + \left( \frac{k_B T}{\gamma_{\perp}} \right)^2 \sum_{k,l=1}^K \frac{2x_{k,l}(\bar{s})^2}{\lambda_k \lambda_l (\lambda_k + \lambda_l)^2 \Delta t c_{k,l}} [1 - \cos(2\pi f \Delta t)] \left[ c_{k,l}^2 - 1 + 2(\lambda_k + \lambda_l) \Delta t c_{k,l} + \frac{(1+c_{k,l})(1-c_{k,l})^3}{1+c_{k,l}^2 - 2c_{k,l} \cos(2\pi f \Delta t)} \right], \quad (G6)$$

where  $c_{k,l} = \exp[-(\lambda_k + \lambda_l) \Delta t]$ .

As above,  $y_k(\bar{s})$  are spatial eigenfunctions of the DNA's transversal motion, and  $x_{k,l}(\bar{s}) = \int_0^{\bar{s}} y'_k(s) y'_l(s) ds$ , where a prime denotes derivative with respect to  $s$ , while  $\gamma_{\perp}$  is the drag coefficient of the DNA in its perpendicular direction. Note that the number of contributing modes of substrate motion in  $P_{\Delta x, f}$  is  $K^2$  when the number of contributing modes of substrate motion in  $P_{y, f}$ , and hence by assumption in  $P_{z, f}$ , is  $K$ . This is because the fixed contour length of the DNA makes the longitudinal motion of a point on the DNA a function of its transversal motion.

When a theory gives the spatial eigenmodes of the DNA's motion (see [22]), it suffices to fit two parameters, the DNA's drag coefficient for transversal motion  $\gamma_{\perp}$  and the protein's mean position  $\bar{s}$  on the DNA. This assumes  $k_B T$  is known. If it is not, it should be fitted as well. If no such theory exists,  $(\lambda_k)_{k=1}^K$ ,  $[x_{k,l}(\bar{s})]_{k,l=1}^K$ , and  $[y_k(\bar{s})]_{k=1}^K$  must all be fitted as independent parameters for each given value of  $\bar{s}$ .

### b. CVE

If the substrate fluctuations have been characterized independently, an unbiased CVE that takes multiple modes of the substrate's motion into account, is constructed by replacing  $b_D$  in Eq. (56) with

$$b_D(\phi) = \sum_{k=1}^K \left( \frac{1-c_k}{\lambda_k \Delta t} \right)^3 x_k(\bar{s})^2. \quad (G7)$$

As for the MLE of the previous section,  $K$  is chosen such that all modes that contribute significantly to the motion, are included.

For diffusion on taut, inextensible, but flexible fiber, such as DNA, the bias of the CVE is

$$b_D(\phi) = \sum_{k,l=1}^K \frac{(1-c_{k,l})^3}{\lambda_k \lambda_l (\lambda_k + \lambda_l)^2 (\Delta t)^3} x_{k,l}(\bar{s})^2. \quad (G8)$$

If also the localization errors have been characterized independently,  $b_D$  is given by

$$b_D(\phi) = \sum_{k,l=1}^K \frac{2(\lambda_k + \lambda_l) \Delta t - 3 + 4c_{k,l} - c_{k,l}^2}{(1-2R)\lambda_k \lambda_l (\lambda_k + \lambda_l)^2 (\Delta t)^3} x_{k,l}(\bar{s})^2, \quad (G9)$$

which is inserted in Eq. (G2).

### 3. Camera shutter time shorter than time lapse

In experiments the camera shutter is usually kept open during the whole experiment; i.e., the camera shutter time is equal to the time lapse [25]. In almost all relevant scenarios this is optimal since it maximizes the number of photons recorded in a frame and thus minimizes the localization error. It may be necessary, e.g., due to excessive motion blur, to open the shutter only for part of the time lapse, i.e., for a time interval  $\tau < \Delta t$ . In this case the statistics of the measured time series change. We give below the power spectrum and covariance which in

this case should be used in the MLE and CVE instead of the corresponding expressions given in Sec. IV.

### a. Power spectrum

If the shutter is held open for the duration  $\tau < \Delta t$ , Eq. (36) must be replaced with

$$P_{\Delta s, f} = 2D\Delta t \{ \Delta t - [1 - \cos(2\pi f \Delta t)]\tau/3 \} \quad (\text{G10})$$

and Eq. (38) with

$$\begin{aligned} P_f^{(\text{sub})} &= \sum_{m=-\infty}^{\infty} \frac{1 - \cos[2\pi(f + mf_{\text{sample}})\tau]}{2\pi^2\tau^2(f + mf_{\text{sample}})^2} P_{f+mf_{\text{sample}}}^{(\text{sub}, \text{cont})} \\ &= \frac{2K_x(\bar{s})\Delta t}{(2\pi f_c)^3\tau^2} \left\{ 2\pi f_c\tau - \sinh(2\pi f_c\tau) \right. \\ &\quad \left. + \frac{(1 - c^2)[\cosh(2\pi f_c\tau) - 1]}{1 + c^2 - 2c \cos(2\pi f \Delta t)} \right\}, \end{aligned} \quad (\text{G11})$$

where the infinite sum was done like that in Eq. (F5), by *Poisson resummation*. The power spectra must be changed accordingly in Eq. (40), Eqs. (50) and (51), and Eqs. (G4)–(G6).

### b. Covariance

If the shutter is kept open only for some fraction  $\tau/\Delta t$  of the time lapse, the autocovariance of the substrate's fluctuations is

given by

$$\begin{aligned} C_x(\bar{s}, j \Delta t; \tau) &= \begin{cases} \sum_{k=1}^K \frac{[2\lambda_k\tau - 2(1 - c_k)]}{\lambda_k^3\tau^2} x_k(\bar{s})^2 & \text{for } j = 0, \\ \sum_{k=1}^K \frac{2[\cosh(\lambda_k\tau) - 1]}{\lambda_k^3\tau^2} x_k(\bar{s})^2 c_k^{|j|} & \text{elsewise.} \end{cases} \end{aligned} \quad (\text{G12})$$

From these expressions,  $b_D$  can be calculated as  $b_D = [C_x(\bar{s}, \Delta t; \tau) - C_x(\bar{s}, 2\Delta t; \tau)]/\Delta t$  when  $\sigma$ , the amplitude of localization errors, is unknown, and as  $b_D = [C_x(\bar{s}, 0; \tau) - C_x(\bar{s}, \Delta t; \tau)]/\Delta t$  when the localization errors have been characterized independently and  $\sigma$  hence is known. The bias  $b_D$  can then be inserted into Eq. (56) or Eq. (G2), respectively, to construct an unbiased CVE of  $D$ .

## APPENDIX H: VARIANCE OF ESTIMATOR FOR A FLUCTUATING SUBSTRATE

We here derive the variance of estimators  $\hat{D}$  and  $\hat{\sigma}^2$  for the diffusion coefficient  $D$  and the variance  $\sigma^2$  of localization errors. We do this for the case when substrate fluctuations have been characterized independently and for the case when both substrate fluctuations and localization error have been characterized independently. Section 1 of this appendix details the derivation of the variance of CVE, while Section 2 of this appendix details the derivation of the variance of MLE.

### 1. Derivation of the variance of the CVE

We here derive the variance of the unbiased CVE [Eq. (56)] for diffusion on a fluctuation substrate. We ignore the contribution to the variance from uncertainties in  $\zeta(\bar{s})$  in the following calculations. Their contribution is found afterwards by standard propagation of errors. From Eq. (56) the variance of the unbiased CVE is then

$$\text{var}(\hat{D}) = \frac{\text{var}[(\Delta x_n)^2]}{4(\Delta t)^2} + \frac{\text{var}(\Delta x_n \Delta x_{n+1})}{(\Delta t)^2} + \frac{\text{cov}((\Delta x_n)^2, \Delta x_n \Delta x_{n+1})}{(\Delta t)^2} + \text{var}(\widehat{b_D}). \quad (\text{H1})$$

The variance of  $\widehat{b_D}$  is found by standard propagation of errors as given by Eq. (63). The other three terms are calculated as in Sec. III C, where we here must take contributions from DNA fluctuations into account as well,

$$\begin{aligned} \text{var}[(\Delta x_n)^2] &= \frac{\text{var}[(\Delta x_n)^2]}{N} + \frac{2}{N^2} \sum_{j=1}^{N-1} (N-j) \text{cov}((\Delta x_n)^2, (\Delta x_{n+j})^2) \\ &= \frac{8(\alpha + \beta)^2 + 4(1 - 1/N)\beta^2}{N} + \frac{8(\alpha + \beta)\mathcal{C}_0 - 8(1 - 1/N)\beta\mathcal{C}_1}{N} + \frac{2\mathcal{C}_0^2}{N} + \frac{4}{N^2} \sum_{j=1}^{N-1} (N-j)\mathcal{C}_j^2, \end{aligned} \quad (\text{H2})$$

since

$$\begin{aligned} \text{var}[(\Delta x_n)^2] &= 8(\alpha + \beta)^2 + 8(\alpha + \beta)\mathcal{C}_0 + 2\mathcal{C}_0^2, \quad \text{cov}((\Delta x_n)^2, (\Delta x_{n+1})^2) = 2\beta^2 - 4\beta\mathcal{C}_1 + 2\mathcal{C}_1^2, \\ \text{cov}((\Delta x_n)^2, (\Delta x_{n+j})^2) &= 2\mathcal{C}_j^2, \quad j > 1, \end{aligned}$$

where  $\alpha \equiv \zeta(\bar{s})^2 D \Delta t$  and  $\beta \equiv \sigma^2 - \zeta(\bar{s})^2 D \Delta t/3$ ;

$$\begin{aligned} \text{var}(\Delta x_n \Delta x_{n+1}) &= \frac{\text{var}(\Delta x_n \Delta x_{n+1})}{N-1} + \frac{2}{(N-1)^2} \sum_{j=1}^{N-2} (N-j-1) \text{cov}(\Delta x_n \Delta x_{n+1}, \Delta x_{n+j} \Delta x_{n+j+1}) \\ &= \frac{4(\alpha + \beta)^2 + 3\beta^2}{N-1} - \frac{2\beta^2}{(N-1)^2} + \frac{4(\alpha + \beta)\mathcal{C}_0 - 6\beta\mathcal{C}_1 + 4(\alpha + \beta)\mathcal{C}_2 - 2\beta\mathcal{C}_3}{N-1} + \frac{4\beta\mathcal{C}_1 - 4(\alpha + \beta)\mathcal{C}_2 + 4\beta\mathcal{C}_3}{(N-1)^2} \\ &\quad + \frac{\mathcal{C}_0^2 + \mathcal{C}_1^2}{N-1} + \frac{2}{(N-1)^2} \sum_{j=1}^{N-2} (N-j-1) (\mathcal{C}_{j-1}\mathcal{C}_{j+1} + \mathcal{C}_j^2), \end{aligned} \quad (\text{H3})$$



since

$$\begin{aligned}
 \text{var}(\Delta x_n \Delta x_{n+1}) &= 4(\alpha + \beta)^2 + \beta^2 + 4(\alpha + \beta)\mathcal{C}_0 - 2\beta\mathcal{C}_1 + \mathcal{C}_0^2 + \mathcal{C}_1^2, \\
 \text{cov}(\Delta x_n \Delta x_{n+1}, \Delta x_{n+2} \Delta x_{n+3}) &= \beta^2 - 2\beta\mathcal{C}_1 + 2(\alpha + \beta)\mathcal{C}_2 + \mathcal{C}_0\mathcal{C}_2 + \mathcal{C}_1^2, \\
 \text{cov}(\Delta x_n \Delta x_{n+1}, \Delta x_{n+2} \Delta x_{n+3}) &= -\beta\mathcal{C}_3 + \mathcal{C}_1\mathcal{C}_3 + \mathcal{C}_2^2, \\
 \text{cov}(\Delta x_n \Delta x_{n+1}, \Delta x_{n+j} \Delta x_{n+j+1}) &= \mathcal{C}_{j-1}\mathcal{C}_{j+1} + \mathcal{C}_j^2, \quad j > 2; \\
 \text{cov}(\overline{(\Delta x_n)^2}, \overline{\Delta x_n \Delta x_{n+1}}) &= \frac{2}{N(N-1)} \sum_{j=0}^{N-2} (N-j-1) \text{cov}((\Delta x_n)^2, \Delta x_{n+j} \Delta x_{n+j+1}) \\
 &= -\frac{8(\alpha + \beta)\beta}{N} + \frac{-4\beta\mathcal{C}_0 + 8(\alpha + \beta)\mathcal{C}_1 - 4[1 - 1/(N-1)]\beta\mathcal{C}_2}{N} \\
 &\quad + \frac{4}{N(N-1)} \sum_{j=0}^{N-2} (N-j-1) \mathcal{C}_j \mathcal{C}_{j+1}, \tag{H4}
 \end{aligned}$$

since

$$\begin{aligned}
 \text{cov}(\overline{(\Delta x_n)^2}, \overline{\Delta x_n \Delta x_{n+1}}) &= -4(\alpha + \beta)\beta - 2\beta\mathcal{C}_0 + 4(\alpha + \beta)\mathcal{C}_1 + 2\mathcal{C}_0\mathcal{C}_1, \\
 \text{cov}(\overline{(\Delta x_n)^2}, \overline{\Delta x_{n+1} \Delta x_{n+2}}) &= -2\beta\mathcal{C}_2 + 2\mathcal{C}_1\mathcal{C}_2, \\
 \text{cov}(\overline{(\Delta x_n)^2}, \overline{\Delta x_{n+j} \Delta x_{n+j+1}}) &= 2\mathcal{C}_j \mathcal{C}_{j+1}, \quad j > 1.
 \end{aligned}$$

Combining Eqs. (H2)–(H4) gives the variance [Eq. (59)] of the unbiased CVE defined by Eq. (56). The variance [Eq. (H8)] of the CVE for *a priori* determined noise amplitude [Eq. (G2)] is derived in a similar fashion.

## 2. Variance of $\hat{D}$ for known variance $\sigma^2$ of localization errors

### a. MLE

We here derive an expression for the variance of the MLE  $\hat{D}$  for the case of diffusion on a fluctuating substrate. We do this for the case when the substrate fluctuations have been characterized independently. We also do it for the case when substrate fluctuations and localization errors both have been characterized independently.

Let  $\phi$  denote the parameters that have already been estimated independently, and let  $\hat{\phi}$  denote the estimates. Let  $\theta$  denote the parameters that we want to estimate using MLE. When we use the estimates  $\hat{\phi}$  as fixed parameters in the estimation of  $\theta$ , the errors on  $\hat{\phi}$ ,  $\Delta\phi = \hat{\phi} - \phi^*$ , propagate to  $\hat{\theta}$ . Here  $\phi^*$  denotes the true value of  $\phi$ . Since the stationarity condition

$$l_{\theta_i}(\hat{\theta}|\hat{\phi}, \mathbf{x}) \equiv \frac{\partial \ln \mathcal{L}(\hat{\theta}|\hat{\phi}, \mathbf{x})}{\partial \theta_i} = 0 \tag{H5}$$

gives  $\hat{\theta}$  only as an implicit function of  $\phi$ , we cannot use classical propagation of errors to calculate the variance of  $\hat{\theta}$ . We can, however, follow a derivation similar to the one that shows the approximate equality between the inverse Fisher information and the variance of the MLE, and derive a first-order approximation (in  $1/N$ ) of the variance of  $\hat{\theta}$ .

From the stationarity condition [Eq. (H5)],

$$\begin{aligned}
 0 &= l_{\theta_i}(\hat{\theta}|\hat{\phi}, \mathbf{x}) = l_{\theta_i}(\theta^*|\phi^*, \mathbf{x}) + l_{\theta_i\theta_j}(\theta^*|\phi^*, X)\Delta\theta_j + l_{\theta_i\phi_k}(\theta^*|\phi^*, X)\Delta\phi_k + \mathcal{O}_p(1) \\
 &= l_{\theta_i}(\theta^*|\phi^*, X) + (\mathcal{I}_\theta)_{ij}\Delta\theta_j + (\mathcal{I}_\phi^\theta)_{ik}\Delta\phi_k + \mathcal{O}_p(1),
 \end{aligned}$$

since the central limit theorem dictates that  $l_{\theta_i\theta_j}(\theta^*|\phi^*, X) = (\mathcal{I}_\theta)_{ij} + \mathcal{O}_p(1)$ . Thus,

$$\Delta\theta_m = -(\mathcal{I}_\theta)^{mi} [l_{\theta_i}(\theta^*|\phi^*, X) + (\mathcal{I}_\phi^\theta)_{ik}\Delta\phi_k] + \mathcal{O}_p(N^{-1}), \tag{H6}$$

where  $\mathcal{I}^{mi} = (\mathcal{I}^{-1})_{mi}$ , and we sum over repeated indices. Then

$$\begin{aligned}
 \text{var}(\hat{\theta})_{mn} &= \langle \Delta\theta_m \Delta\theta_n \rangle \simeq (\mathcal{I}_\theta)^{mi} \left[ \langle l_{\theta_i}(\theta^*|\phi^*, X) l_{\theta_j}(\theta^*|\phi^*, X) \rangle + (\mathcal{I}_\phi^\theta)_{ik} \langle \Delta\phi_k \Delta\phi_l \rangle (\mathcal{I}_\phi^\theta)_{lj} \right. \\
 &\quad \left. + (\mathcal{I}_\phi^\theta)_{il} \langle l_{\theta_i}(\theta^*|\phi^*, X) \Delta\phi_l \rangle + (\mathcal{I}_\phi^\theta)_{jk} \langle l_{\theta_j}(\theta^*|\phi^*, X) \Delta\phi_k \rangle \right] (\mathcal{I}_\theta)^{jn} \\
 &= (\mathcal{I}_\theta)^{mn} + (\mathcal{I}_\theta)^{mi} (\mathcal{I}_\phi^\theta)_{ik} \text{var}(\hat{\phi})_{kl} (\mathcal{I}_\phi^\theta)_{lj} (\mathcal{I}_\theta)^{jn}, \tag{H7}
 \end{aligned}$$

since  $l_{\theta_j}(\theta^*|\phi^*, X)$  and  $\hat{\phi}_k$  are uncorrelated.

### b. CVE

When also  $\sigma^2$  has been determined independently, the variance of the unbiased CVE is derived as for unknown  $\sigma^2$ . It is

$$\frac{\text{var}(\hat{D})}{D^2} = \frac{\text{var}[(\Delta x_n)^2]}{[2\zeta(\bar{s})^2 D(1-2R)\Delta t]^2} + \frac{\text{var}(\hat{\sigma}^2)}{[\zeta(\bar{s})^2 D(1-2R)\Delta t]^2} + \frac{\text{var}[b_D(\hat{\sigma}^2, \hat{\phi})]}{[\zeta(\bar{s})^2 D]^2} + \frac{\text{var}[\zeta(\bar{s})^2]}{\zeta(\bar{s})^4}, \quad (\text{H8})$$

where now, differing from Eq. (H2) and with  $\varepsilon = \sigma^2/[\zeta(\bar{s})^2 D\Delta t] - 2R$ ,

$$\begin{aligned} \text{var}[(\Delta x_n)^2] &= [2\zeta(\bar{s})^2 D\Delta t]^2 \frac{2(1+\varepsilon)^2 + (1-1/N)\varepsilon^2}{N} + 8\zeta(\bar{s})^2 D\Delta t \frac{(1+\varepsilon)\mathcal{C}_0 - (1-1/N)\varepsilon\mathcal{C}_1}{N} \\ &\quad + \frac{2\mathcal{C}_0^2}{N} + \frac{4}{N^2} \sum_{j=1}^{N-1} (N-j)\mathcal{C}_j^2. \end{aligned} \quad (\text{H9})$$

## APPENDIX I: MONTE CARLO SIMULATIONS OF DIFFUSION ON DNA

We simulated the laboratory motion of a point on the DNA as the motion of a massively overdamped harmonic oscillator in a heat bath [51]. Since we thus simulate only one mode of the DNA's motion, we can simulate the coupling between the two transversal modes and the longitudinal mode exactly without having to simulate the motion of the entire DNA strand. We use that [22]

$$\begin{aligned} x(s, t) &= s - \frac{1}{2} \int_0^s [y'(s', t)^2 + z'(s', t)^2] ds' \\ &= s - \frac{1}{2} \int_0^s [A_{1,y}(t)^2 + A_{1,z}(t)^2] y_1'(s')^2 ds' \\ &= s - \frac{y(s', t)^2 + z(s', t)^2}{2} \int_0^s \frac{y_1'(s')^2}{y_1(s')^2} ds' \\ &= s - \frac{\sqrt{\pi f_c K_x}}{2K_y} [y(s', t)^2 + z(s', t)^2], \end{aligned} \quad (\text{I1})$$

since the motions in the two transversal directions are equivalent such that  $y_1 = z_1$  and  $f_c K_x$  and  $K_y$  are constant along the DNA.

Experimental data are usually measured with the camera shutter kept open for the duration of a time lapse  $\Delta t$ . So to simulate positions we need to integrate over the full time lapse,

$$y_{n+1}^{\text{DNA}} = \frac{1}{\Delta t} \int_0^{\Delta t} y^{\text{DNA}}(t_n + t) dt, \quad (\text{I2})$$

We do this by approximating the integral with a sum,

$$y_{n+1}^{\text{DNA}} \simeq h \sum_{q=1}^{1/h} y^{\text{DNA}}(t_n + qh\Delta t), \quad (\text{I3})$$

where  $1/h$  is an integer and Eq. (I3) approaches Eq. (I2) as  $h \rightarrow 0$ . The motion of a transversal mode is equivalent to the motion of a Brownian particle trapped in an optical trap. We can thus simulate  $y^{\text{DNA}}$  according to [31],

$$y^{\text{DNA}}(t_{m+1}) = e^{-2\pi f_c h \Delta t} y^{\text{DNA}}(t_m) + \Delta y_h \eta_m^{\text{DNA}}, \quad (\text{I4})$$

where  $m = 1, 2, \dots, (N+1)/h$ ,  $\eta_m^{\text{DNA}}$  is a normalized Gaussian white noise, and

$$\Delta y_h \equiv \sqrt{\frac{(1 - e^{-4\pi f_c h \Delta t}) K_y(s)}{4\pi f_c}}.$$

The  $y$  position of the DNA measured at time  $t_n$  is thus

$$y_n^{\text{DNA}} = h \sum_{q=1}^{1/h} y^{\text{DNA}}(t_n/h + q). \quad (\text{I5})$$

The diffusive movement of the protein is simulated in a similar fashion,

$$x^{\text{Diff}}(t_{m+1}) = x^{\text{Diff}}(t_m) + \sqrt{2Dh\Delta t} \eta_m^{\text{Diff}}, \quad (\text{I6})$$

where  $m = 1, 2, \dots, (N+1)/h$  and  $\eta_m^{\text{Diff}}$  is a normalized Gaussian white noise.

We simulated three time series,  $\{y^{\text{DNA}}(t_m)\}_{m=-Q/h}^{(N+1)/h}$ ,  $\{z^{\text{DNA}}(t_m)\}_{m=-Q/h}^{(N+1)/h}$ , and  $\{x^{\text{Diff}}(t_m)\}_{m=-Q/h}^{(N+1)/h}$ , where we set  $x(t_{-Q/h}) = y(t_{-Q/h}) = z(t_{-Q/h}) = 0$  and  $1/h = 100$ , and  $Q$  was chosen such that the DNA thermalized before we sampled the time series. We calculated the longitudinal positions of the DNA,  $x^{\text{DNA}}$ , from Eq. (I1) and calculated the motion blurred positions  $\{y_n^{\text{DNA}}\}_{n=0}^N$ ,  $\{x_n^{\text{DNA}}\}_{n=0}^N$ , and  $\{x_n^{\text{Diff}}\}_{n=0}^N$  using Eq. (I5).

We finally summed the DNA motion and the diffusive movement and added positional noise to obtain the “measured” positions,

$$x_n = x_n^{\text{Diff}} + x_n^{\text{DNA}} + \sigma \xi_n, \quad (\text{I7})$$

where  $\xi_n$  is standard white noise. We also calculated the transversal positions,

$$y_n = y_n^{\text{DNA}} + \sigma \xi_n. \quad (\text{I8})$$

Figures 8(a) and 8(b) present numerical results for the performance of the MLE that explicitly accounts for DNA fluctuations, and for the CVE, which does not, as a function of the relaxation rate  $2\pi f_c$ . Figures 8(c) and 8(d) compare the performances of the MLE and CVE in the case when the DNA's fluctuations have been characterized *a priori*. Figures 8(e) and 8(f) show the performances of the estimators for the case when both the DNA's motion and the localization errors have been characterized independently.

### 1. Correlations between transversal and longitudinal DNA fluctuations

In order to extract as much information as possible about the motion of the DNA from the experimental time series, we use the periodograms of both the longitudinal and the transversal motion,  $(P_{\Delta x})_f$  and  $(P_y)_f$ . To do this in practice, we assume that  $\hat{P}_{y,f}$  and  $\hat{P}_{\Delta x,f}$  are independent. However, since the

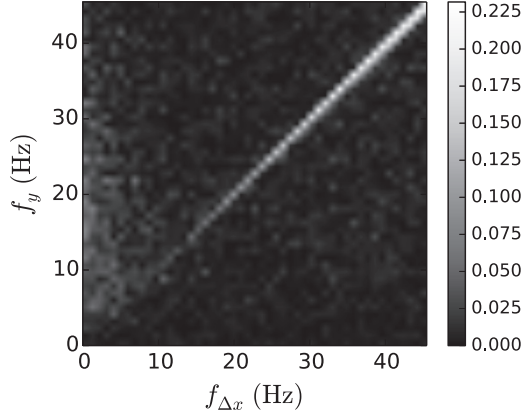


FIG. 10. Heat map of correlations between  $\hat{P}_{y,f}$  and  $\hat{P}_{\Delta x,f}$  (a value of 1 signifies that  $\hat{P}_{y,f}$  and  $\hat{P}_{\Delta x,f}$  are perfectly correlated while a value of 0 signifies complete decorrelation). From Monte Carlo generated data with  $D = 0.3 \mu\text{m}^2/\text{s}$ ,  $f_c = 30 \text{ Hz}$ ,  $K_x = 2.1 \mu\text{m}$ ,  $K_y = 0.21 \mu\text{m}$ , and  $\sigma^2 = 1500 \text{ nm}$ . The ensemble size is  $M = 10\,000$ , the length of time series is  $N = 100$ , and the shutter is kept open for the whole time lapse  $\tau = \Delta t$ . The plot shows that overall it is a good approximation to assume that  $\hat{P}_{y,f}$  and  $\hat{P}_{\Delta x,f}$  are uncorrelated.

longitudinal motion of the DNA is dependent on its transversal motion,  $\hat{P}_{y,f}$  and  $\hat{P}_{\Delta x,f}$  are not completely independent. We calculated the correlations between  $\hat{P}_{y,f}$  and  $\hat{P}_{\Delta x,f}$  for an ensemble of 10 000 Monte Carlo generated time series of length  $N = 100$ . The correlations between  $\hat{P}_{y,f}$  and  $\hat{P}_{\Delta x,f}$  are, in general, small and can globally be ignored (Fig. 10). The maximal value of the correlations is 0.23, the mean correlation coefficient is  $0.0085 \pm 0.0003$ , while the mean absolute value of the correlation coefficient is  $0.0133 \pm 0.0002$ . Furthermore, the MLE performs practically optimally even when these correlations are ignored (Fig. 8).

## APPENDIX J: HYPOTHESIS TESTING

We used the following two standard statistical tests to compare theory with experimental data: *Pearson's  $\chi^2$  goodness-of-fit test* (Sec. 1 of this appendix) and the  *$\chi^2$  test for variance* (Sec. 2 of this appendix).

### 1. Pearson's $\chi^2$ goodness-of-fit test

Pearson's  $\chi^2$  goodness-of-fit test is used to compare a measured distribution to a theoretical distribution. For use in Pearson's  $\chi^2$  test, the measured data are divided into a number of bins  $r$ . The number of bins should be large enough

to ensure that the observed number of counts  $O_i$  in any bin  $i$  is Poisson distributed. Moreover, the expected number of counts,  $E_i$ , should be large enough in every bin to ensure that the Poisson distribution of  $O_i$  is approximately Gaussian. The test statistics is

$$X^2 = \sum_{i=1}^r \frac{(O_i - E_i)^2}{E_i}. \quad (\text{J1})$$

If the theory that provides the values for  $E_i$  is correct,  $X^2$  follows a  $\chi^2$  distribution in the limit of many counts in each bin. A common and not-too-demanding rule of thumb demands  $O_i \geq 5$  in all bins. If no parameters were fitted, the expected  $\chi^2$  distribution has  $r - 1$  degrees of freedom. If  $n$  parameters in the theory were determined by fitting the theory to the data before  $\chi^2$  is evaluated, the number of degrees of freedom is  $r' = r - 1 - n$ . The number 1 is subtracted to account for the fact that the total count is known. This number  $r'$  and  $X^2$  are used to calculate the statistical support for the theory, also known as its  $P$  value, defined as the probability that a repetition of the experiment and the fitting to the data will result in a new  $\chi^2$  value that is larger than the value we obtained by fitting to the existing data.

### 2. $\chi^2$ test for variance

The  $\chi^2$  test for variance consists of estimating whether the dispersion of a number  $q$  of observed data points  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$  with known variances  $\sigma_i^2$  is consistent with the assumption that the points are independently and normally distributed around their expected values  $\theta_1, \theta_2, \dots, \theta_q$ , which are predicted by a theoretical model. This consistency is the null hypothesis of the test.

The  $\chi^2$  test statistic is

$$X^2 = \sum_{i=1}^q \frac{(\hat{\theta}_i - \theta_i)^2}{\sigma_i^2},$$

and follows a  $\chi^2$  distribution with  $q - p$  degrees of freedom, where  $p$  is the number of model parameters that were estimated from the data if the null hypothesis is true.

In practice we bin measurements and use the sample variances  $s_i^2$  as estimates of the true variances  $\sigma_i^2$  of the estimated means  $\theta_i$ . Thus, the test statistic is not exactly  $\chi^2$  distributed. However, since the number of measurements in each bin is larger than 10, the difference between the actual distribution and the  $\chi^2$  distribution is negligible if the null hypothesis is true, and consequently  $\theta_i$  is Gaussian distributed for all  $i = 1, \dots, q$ .

- [1] A. Tafvizi, L. A. Mirny, and A. M. van Oijen, *Chem. Phys. Chem.* **12**, 1481 (2011).
- [2] J. Gorman and E. C. Greene, *Nat. Struct. Mol. Biol.* **15**, 768 (2008).
- [3] S. E. Halford and J. F. Marko, *Nucleic Acids Res.* **32**, 3040 (2004).
- [4] R. D. Vale, D. R. Soll, and I. R. Gibbons, *Cell* **59**, 915 (1989).

- [5] J. Helenius, G. Brouhard, Y. Kalaidzidis, S. Diez, and J. Howard, *Nature (London)* **441**, 115 (2006).
- [6] I. Minoura, E. Katayama, K. Sekimoto, and E. Muto, *Biophys. J.* **98**, 1589 (2010).
- [7] A. W. Sonesson, U. M. Elofsson, T. H. Callisen, and H. Brismar, *Langmuir* **23**, 8352 (2007).
- [8] S. Wieser and G. J. Schütz, *Methods* **46**, 131 (2008).

- [9] D. Lasne, G. A. Blab, S. Berciaud, M. Heine, L. Groc, D. Choquet, L. Cognet, and B. Lounis, *Biophys. J.* **91**, 4598 (2006).
- [10] S. Y. Nishimura, S. J. Lord, L. O. Klein, K. A. Willets, M. He, Z. Lu, R. J. Twieg, and W. E. Moerner, *J. Phys. Chem. B* **110**, 8151 (2006).
- [11] M. B. Smith, E. Karatekin, A. Gohlke, H. Mizuno, N. Watanabe, and D. Vavylonis, *Biophys. J.* **101**, 1794 (2011).
- [12] H. Bornfleth, P. Edelmann, D. Zink, and T. Cremer, *Biophys. J.* **77**, 2871 (1999).
- [13] M. Goulian and S. M. Simon, *Biophys. J.* **79**, 2188 (2000).
- [14] H. Qian, M. P. Sheetz, and E. L. Elson, *Biophys. J.* **60**, 910 (1991).
- [15] X. Michalet, *Phys. Rev. E* **82**, 041914 (2010).
- [16] M. J. Saxton, *Biophys. J.* **72**, 1744 (1997).
- [17] This esoteric acronym should not be confused with OLS, the standard acronym for ordinary least squares.
- [18] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. (Wiley, 1973).
- [19] X. Michalet and A. J. Berglund, *Phys. Rev. E* **85**, 061916 (2012).
- [20] A. J. Berglund, *Phys. Rev. E* **82**, 011917 (2010).
- [21] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, and X. S. Xie, *Proc. Natl. Acad. Sci. USA* **103**, 5752 (2006).
- [22] C. L. Vestergaard, P. C. Blainey, and H. Flyvbjerg (unpublished).
- [23] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.89.022726> for a summary of acronyms and notation used in the article.
- [24] K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, *Nat. Methods* **7**, 377 (2010).
- [25] C. L. Vestergaard (unpublished).
- [26] P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi, and X. S. Xie, *Nat. Struct. Mol. Biol.* **16**, 1224 (2009).
- [27] A. Granéli, C. C. Yeykal, R. B. Robertson, and E. C. Greene, *Proc. Natl. Acad. Sci. USA* **103**, 1221 (2006).
- [28] M. Vrljic, S. Y. Nishimura, S. Brasselet, W. E. Moerner, and H. M. McConnell, *Biophys. J.* **83**, 2681 (2002).
- [29] A. Einstein, *Ann. Phys.* **17**, 549 (1905).
- [30] A. Biebricher, W. Wende, C. Escudé, A. Pingoud, and P. Desbiolles, *Biophys. J.* **96**, L50 (2009).
- [31] K. Berg-Sørensen and H. Flyvbjerg, *Rev. Sci. Instrum.* **75**, 594 (2004).
- [32] Note that MATLAB and SciPy's `dst` and `fft` algorithms are defined without the  $\Delta t$  prefactor. The result returned by `dst` or `fft` should thus be multiplied by  $\Delta t$  to obtain the DST and DFT given by Eqs. (11) and (12), respectively.
- [33] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, UK, 2001).
- [34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, UK, 2002).
- [35] W. Ying, G. Huerta, S. Steinberg, and M. Zúñiga, *Bull. Math. Biol.* **71**, 1967 (2009).
- [36] S. F. Nørrelykke and H. Flyvbjerg, *Rev. Sci. Instrum.* **81**, 075103 (2010).
- [37] Here and in Appendix C we make an exception and let  $\hat{D}$  denote the estimate, while everywhere else in this article it denotes the estimator. We prefer this potentially confusing, but commonly used, notation here rather than introducing yet more notation.
- [38] F. A. Espinoza, M. J. Wester, J. M. Oliver, B. S. Wilson, N. L. Andrews, D. S. Lidke, and S. L. Steinberg, *Bull. Math. Biol.* **74**, 1857 (2012).
- [39] C. Bouchiat, M. D. Wang, J. Allemand, T. Strick, S. M. Block, and V. Croquette, *Biophys. J.* **76**, 409 (1999).
- [40] T. Q. Nguyen and H.-H. Kausch, eds., *Flexible Polymer Chains in Elongational Flow: Theory and Experiment* (Springer-Verlag, New York, 1999).
- [41] F. Brochard-Wyart, *Europhys. Lett.* **30**, 387 (1995).
- [42] K. Laube, K. Günther, and M. Mertig, *J. Phys. Condens. Matter* **23**, 184119 (2011).
- [43] C. E. Sing and A. Alexander-Katz, *Macromolecules* **44**, 9020 (2011).
- [44] Y. Zhang, A. Donev, T. Weisgraber, B. J. Alder, M. D. Graham, and J. J. de Pablo, *J. Chem. Phys.* **130**, 234902 (2009).
- [45] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed. (McGraw-Hill, New York, 2003).
- [46] H. Madsen, *Time Series Analysis* (Chapman & Hall/CRC, Boca Raton, FL, 2008).
- [47] M. Hamidi and J. Pearl, *IEEE Trans. Inf. Theory* **21**, 480 (1975).
- [48] M. Peligrad and W. B. Wu, *Ann. Prob.* **38**, 2009 (2010).
- [49] W. P. Wong and K. Halvorsen, *Opt. Express* **14**, 12517 (2006).
- [50] K. Berg-Sørensen, L. Oddershede, E.-L. Florin, and H. Flyvbjerg, *J. Appl. Phys.* **93**, 3167 (2003).
- [51] S. F. Nørrelykke and H. Flyvbjerg, *Phys. Rev. E* **83**, 041103 (2011).