Time-averaged quadratic functionals of a Gaussian process

Denis S. Grebenkov*

Laboratoire de Physique de la Matière Condensée (UMR 7643), CNRS – Ecole Polytechnique, F-91128 Palaiseau, France, Laboratoire Poncelet (UMI 2615), CNRS – Independent University of Moscow, Bolshoy Vlasyevskiy Pereulok 11, 119002 Moscow, Russia, and Chebyshev Laboratory, Saint Petersburg State University, 14th line of Vasil'evskiy Ostrov 29, Saint Petersburg, Russia

(Received 16 February 2011; published 14 June 2011)

The characterization of a stochastic process from its single random realization is a challenging problem for most single-particle tracking techniques which survey an individual trajectory of a tracer in a complex or viscoelastic medium. We consider two quadratic functionals of the trajectory: the time-averaged mean-square displacement (MSD) and the time-averaged squared root mean-square displacement (SRMS). For a large class of stochastic processes governed by the generalized Langevin equation with arbitrary frictional memory kernel and harmonic potential, the exact formulas for the mean and covariance of these functionals are derived. The formula for the mean value can be directly used for fitting experimental data, e.g., in optical tweezers microrheology. The formula for the variance (and covariance) allows one to estimate the intrinsic fluctuations of measured (or simulated) time-averaged MSD or SRMS for choosing the experimental setup appropriately. We show that the time-averaged SRMS has smaller fluctuations than the time-averaged MSD, in spite of much broader applications of the latter one. The theoretical results are successfully confirmed by Monte Carlo simulations of the Langevin dynamics. We conclude that the use of the time-averaged SRMS would result in a more accurate statistical analysis of individual trajectories and more reliable interpretation of experimental data.

DOI: 10.1103/PhysRevE.83.061117 PACS number(s): 02.50.−r, 05.60.−k, 05.10.−a, 02.70.Rr

I. INTRODUCTION

Single-particle tracking (SPT) techniques which survey individual trajectories of tracer particles, assess the most detailed information about dynamics in complex or viscoelastic media, notably in living cells [1–5]. The mean-square displacement (MSD) of a tracer fluctuating inside the substance is often used for extracting dynamical quantities (e.g., diffusion coefficient or subdiffusion scaling exponent) and microrheological quantities (e.g., stiffness [5], compliance and dynamic moduli [6,7], or viscosities [8]). By its nature, SPT techniques provide a limited number (from few to few hundred) of random realizations of the underlying stochastic process. Even if many realizations are available, statistical "ensemble" averaging may still be problematic for tracers that move in spatially heterogeneous and time-evaluating media such as living cells. One faces therefore a challenging problem of characterizing the process from several, or even a single random realization. For comparison, fluorescence photobleaching recovery or proton nuclear magnetic resonance imaging [9–11], in which the signal is formed by an extremely large number of molecules, provides the ensemble-averaged information about the dynamics, and the aforementioned problem does not appear.

The standard approach consists of replacing ensemble averages by time averages [12–14]. If the sample duration t_m is long enough, a tracer is expected to experience various features of the dynamics that would be somehow reflected in the trajectory. In terms of ergodic theory, if the trajectory is long enough to sample densely the whole phase space, time-averaged characteristics of the process are expected to be close to their ensemble-averaged counterparts.

A practical implementation of time averaging faces several problems. The ergodicity is a strong assumption which may or may not be valid. For instance, continuous-time random walks (CTRWs) exhibiting aging [15–18] are known to be nonergodic, e.g., the time-averaged and ensemble-averaged MSD are different [19–21]. The main difficulty is that the ergodicity of a system is not known *a priori* and has to be tested. Another, more practical, problem is to know whether the sample duration t_m is long enough for a sufficient reduction of fluctuations in a time-averaged quantity. In order to illustrate this point, we consider the time-averaged MSD which is the most often used characteristic of diffusion:

$$\mathcal{M}(t,t_m) = \frac{1}{t_m - t}\int_0^{t_m - t} dt_0 m(t,t_0), \qquad m(t,t_0) = [X(t_0+t) - X(t_0)]^2, \tag{1}$$

where $X(t)$ is the position of a tracer at time t. In spite of time averaging, this is still a random variable (we stress that the canonical term "*mean*-squared displacement" is misleading here since there is no expectation or ensemble averaging). It means that two acquisitions under identical conditions lead to different values of $\mathcal{M}(t,t_m)$. Only in the theoretical limit of infinitely long sample ($t_m \to \infty$), the random variable $\mathcal{M}(t,t_m)$ approaches its deterministic ensemble-averaged mean value if the dynamics is ergodic. Note also that the unavoidable presence of experimental noise, static localization uncertainty, the blurring of a particle's position over camera integration times, and medium heterogeneities introduce artifacts into measurements and thus require optimization of acquisition techniques and data analysis [22–24]. Although these issues are important for a reliable interpretation of experimental data, they could not be included in the paper and will be addressed in a future work. In turn, we focus on the intrinsic

*denis.grebenkov@polytechnique.edu

statistical uncertainty of time-averaged functionals of random trajectories.

In practice, the sample duration $t_{\rm m}$ is always finite, and the fluctuations of $\mathcal{M}(t,t_{\rm m})$ can be quantified by its variance. By increasing $t_{\rm m}$, one expects to reduce the statistical uncertainty in the determination of the ensemble-averaged value of $\mathcal{M}(t,t_{\rm m})$ by its empirical (random) realization extracted from a measured individual trajectory. In other words, although $\mathcal{M}(t,t_{\rm m})$ remains random, its values for different random trajectories become less fluctuating and closer to each other and to the mean value as $t_{\rm m}$ increases. This condition may be a quantitative criterion for choosing the appropriate sample duration, e.g., in optical tweezers microrheology. Alternatively, the knowledge of the variance for experimentally accessible $t_{\rm m}$ allows one to characterize the accuracy of time-averaged measurements. Although the time-averaged MSD is routinely used for characterizing diffusion in living cells, only few theoretical studies on the statistical uncertainty are available, and uniquely in the case of normal diffusion [12–14,22].

In this paper, we derive exact explicit formulas for the mean and covariance of the time-averaged MSD for a tracer undergoing the dynamics described by the generalized Langevin equation (GLE) (Sec. II A). The analytical formula for the mean $\mathcal{M}(t,t_{\rm m})$ can be directly used for fitting experimental MSD curves. As intuitively expected, the variance of the time-averaged MSD is shown to increase with the lag time t as the averaging window between 0 and $t_{\rm m}-t$ shrinks. For a large part of the data sample (when the condition $t \ll t_{\rm m}$ is not fulfilled), the time-averaged MSD is too much fluctuating and thus inaccurate. For this reason, other functionals for extracting information about the dynamics have been recently suggested, e.g., the maximum likelihood estimator [24] and the mean maximal excursion [25].

Searching for alternative characterizations of a stochastic process from its single random trajectory, we consider another *quadratic* functional of $X(t)$, the time-averaged squared root mean-square displacement (SRMS):

$$\mathcal{R}(t,t_{\rm m}) \equiv \frac{1}{t_{\rm m}-t}\int_0^{t_{\rm m}-t} dt_0 r(t,t_0), \tag{2}$$

where the SRMS between times t_0 and t_0+t is defined as

$$\begin{aligned} r(t,t_0) &= \frac{1}{t}\int_0^t dt'\left(X(t'+t_0)-\frac{1}{t}\int_0^t dt'' X(t''+t_0)\right)^2 \\ &= \left(\frac{1}{t}\int_0^t dt' X^2(t'+t_0)\right)-\left(\frac{1}{t}\int_0^t dt'' X(t''+t_0)\right)^2 \end{aligned} \tag{3}$$

[in fact, $\sqrt{r(t,t_0)}$ is the root mean square or quadratic mean of tracer's positions between t_0 and t_0+t]. In contrast to MSD, for which the integrand function $m(t,t_0)$ contained uniquely the values of $X(t)$ at two end points t_0 and t_0+t of the averaging window, the new integrand function $r(t,t_0)$ includes the time averaging inside this window. Such a "double" time averaging in Eqs. (2) and (3) is then expected to yield a smaller variance for $\mathcal{R}(t,t_{\rm m})$.

We derive exact explicit formulas for the mean and covariance of $\mathcal{R}(t,t_{\rm m})$. As for the time-averaged MSD, the analytical formula for the mean $\mathcal{R}(t,t_{\rm m})$ can be directly used for fitting experimental time-averaged SRMS curves. Most importantly, we shall show that the variance of $\mathcal{R}(t,t_{\rm m})$ is smaller than that of $\mathcal{M}(t,t_{\rm m})$, especially at long times. As a consequence, the use of the time-averaged SRMS would result in a more accurate statistical analysis of individual trajectories and more reliable interpretation of experimental data.

The paper is organized as follows. In Sec. II, we retrieve the solution of the GLE in a convolution form and then derive the explicit formulas for time-averaged MSD and SRMS for an arbitrary memory kernel. The special case of massless tracers with a power-law friction memory kernel is considered in detail in Sec. III. In Sec. IV, a numerical solution of the GLE via Monte Carlo simulations is used to illustrate fluctuations of time-averaged characteristics. Appendixes present some technical derivations.

II. ARBITRARY MEMORY KERNEL

A. Generalized Langevin equation (GLE)

We consider a large class of dynamics described by a GLE,

$$m\ddot{X}(t)+\int_0^t dt'\gamma(t-t')\dot{X}(t')+kX(t)=F(t), \tag{4}$$

where m is the tracer mass, k is the spring constant of the trapping harmonic potential (e.g., optical trap stiffness), $\gamma(t)$ is the friction memory kernel, and $F(t)$ is the Gaussian thermal force with mean zero and the covariance determined by the fluctuation-dissipation theorem [26,27]:

$$\langle F(t)F(t')\rangle = k_B T\gamma(|t-t'|), \tag{5}$$

with $\langle\cdots\rangle$ denoting the ensemble average over random realizations of the thermal force, k_B being the Boltzmann's constant, and T the absolute temperature. The statistical properties of solutions of the GLE have been intensively studied by many authors (see, e.g., [28–34]). In particular, Desposito and Vinales derived explicit formulas for the limiting MSD, velocity autocorrelation function, and some other functionals in the case of power-law memory kernels [35–38]. Following their approach, we first recall the derivation of a formal solution of Eq. (4).

The Laplace transform of the GLE (4) is

$$[ms^2+\tilde{\gamma}(s)s+k]\tilde{X}(s)=\{mv_0+[ms+\tilde{\gamma}(s)]x_0\}+\tilde{F}(s),$$

where tilde denotes Laplace-transformed quantities, and x_0 and v_0 are the initial position and velocity, from which

$$\tilde{X}(s)=\left[x_0\frac{1}{s}+mv_0\tilde{G}(s)-kx_0\frac{\tilde{G}(s)}{s}\right]+\tilde{G}(s)\tilde{F}(s), \tag{6}$$

where

$$\tilde{G}(s) = \frac{1}{ms^2 + s\tilde{\gamma}(s) + k}. \tag{7}$$

The inverse Laplace transform yields a formal solution,

$$X(t) = X_0(t) + \int_0^t dt' G(t-t') F(t'), \tag{8}$$

where $X_0(t)$ is a solution of the homogeneous Langevin equation without thermal force, i.e., the inverse Laplace transform of the first term in Eq. (6),

$$X_0(t) = x_0 + m v_0 G(t) - k x_0 G^{(1)}(t), \tag{9}$$

with $G(t)$ being the inverse Laplace transform of $\tilde{G}(s)$, and $G^{(n)}(t)$ the nth iterative integral of $G(t)$:

$$G^{(n)}(t) \equiv \int_0^t dt' G^{(n-1)}(t') \quad (n = 1,2,\ldots).$$

Being related to the Gaussian force $F(t)$ through linear Eq. (8), the process $X(t)$ is itself Gaussian. Throughout this paper, we focus on this class of Gaussian processes which are defined by setting the friction memory kernel $\gamma(t)$.

B. Time-averaged MSD

Substituting Eq. (8) into Eq. (1), one can split the mean time-averaged MSD in two terms,

$$\langle \mathcal{M}(t,t_{\rm m}) \rangle = \mathcal{M}_1(t,t_{\rm m}) + \langle \mathcal{M}_2(t,t_{\rm m}) \rangle,$$

where

$$\mathcal{M}_1(t,t_{\rm m}) = \frac{1}{t_{\rm m}-t} \int_0^{t_{\rm m}-t} dt_0 \left[X_0(t_0+t) - X_0(t_0)\right]^2,$$

$$\mathcal{M}_2(t,t_{\rm m}) = \frac{1}{t_{\rm m}-t} \int_0^{t_{\rm m}-t} dt_0 \left[\int_0^{t_0+t} dt_1 G(t_0+t-t_1) F(t_1) - \int_0^{t_0} dt_1 G(t_0-t_1) F(t_1) \right]^2, \tag{10}$$

where the condition $\langle F(t) \rangle = 0$ was used. The explicit form (9) for the homogeneous solution $X_0(t)$ yields

$$\mathcal{M}_1(t,t_{\rm m}) = \frac{1}{t_{\rm m}-t} \int_0^{t_{\rm m}-t} dt_0 [U_0(t,t_0)]^2, \tag{11}$$

with

$$U_0(t,t_0) = m v_0 [G(t_0+t) - G(t_0)] - k x_0 [G^{(1)}(t_0+t) - G^{(1)}(t_0)]. \tag{12}$$

The ensemble average of the second contribution can be written as

$$\langle \mathcal{M}_2(t,t_{\rm m}) \rangle = \frac{k_B T}{t_{\rm m}-t} \int_0^{t_{\rm m}-t} dt_0 U(t,t,t_0,t_0), \tag{13}$$

where

$$U(t_1,t_2,t_{01},t_{02}) = K(t_{01}+t_1,t_{02}+t_2) - K(t_{01},t_{02}+t_2) - K(t_{01}+t_1,t_{02}) + K(t_{01},t_{02}) \tag{14}$$

and

$$K(t_1,t_2) = \int_0^{t_1} dt_1' G(t_1-t_1') \int_0^{t_2} dt_2' G(t_2-t_2') \gamma(|t_1'-t_2'|). \tag{15}$$

The double Laplace transform of the function $K(t_1,t_2)$ with respect to t_1 and t_2 is

$$\tilde{K}(s_1,s_2) = \tilde{G}(s_1)\tilde{G}(s_2) \frac{\tilde{\gamma}(s_1) + \tilde{\gamma}(s_2)}{s_1+s_2}.$$

Introducing $g(t) \equiv dG(t)/dt$ for which

$$\tilde{g}(s) = s\tilde{G}(s) = \frac{s}{ms^2 + \tilde{\gamma}(s)s + k},$$

and writing explicitly the sum of $\tilde{g}(s_1)$ and $\tilde{g}(s_2)$, one obtains

$$s_1 s_2 \tilde{K}(s_1,s_2) = \frac{\tilde{g}(s_1) + \tilde{g}(s_2)}{s_1+s_2} - m\tilde{g}(s_1)\tilde{g}(s_2) - k\tilde{G}(s_1)\tilde{G}(s_2).$$

The inverse Laplace transform of the right-hand side is

$$H(t_1,t_2) \equiv g(|t_1-t_2|) - m g(t_1) g(t_2) - k G(t_1) G(t_2), \tag{16}$$

from which

$$K(t_1,t_2) = \int_0^{t_1} dt_1' \int_0^{t_2} dt_2' H(t_1',t_2') = G^{(1)}(t_1) + G^{(1)}(t_2) - G^{(1)}(|t_2-t_1|) - m G(t_1) G(t_2) - k G^{(1)}(t_1) G^{(1)}(t_2),$$

and we assumed that $G(0) = 0$ here and throughout this section (see Sec. III for discussion). After the substitution of this formula into Eqs. (13) and (14), many terms cancel each other, and one gets

$$\langle \mathcal{M}_2(t,t_{\rm m}) \rangle = M(t) - \frac{k_B T}{t_{\rm m}-t} \int_0^{t_{\rm m}-t} dt_0 \times [m [G(t_0+t) - G(t_0)]^2 + k[G^{(1)}(t_0+t) - G^{(1)}(t_0)]^2], \tag{17}$$

where

$$M(t) = 2 k_B T G^{(1)}(t). \tag{18}$$

In the special case of massless tracers moving without harmonic potential (i.e., $m = 0$ and $k = 0$), the second term vanishes, and the mean time-averaged MSD becomes independent of the sample duration $t_{\rm m}$. Note that Eq. (18) can potentially be used to access the friction memory kernel $\gamma(t)$ by measuring $\langle \mathcal{M}(t,t_{\rm m}) \rangle$.

The mean time-averaged MSD from Eq. (17) is also reduced to $M(t)$ in the limit of infinitely long sample ($t_{\rm m} = \infty$). The

same expression (18) was derived by Desposito and Vinales for the limiting MSD [38]:

$$\lim_{t_0\to\infty}\langle[X(t_0+t)-X(t_0)]^2\rangle = M(t).$$

Although the limiting MSD is convenient for theoretical description, it is inaccessible from experimental measurements. In turn, the time-averaged MSD is directly related to the acquired trajectory of a tracer. Equation (17) proves that these two quantities become identical in the limit of infinitely long time averaging ($t_{\rm m}\to\infty$).

The covariance of the time-averaged MSD is derived in Appendix A:

$$\begin{aligned} C_{\mathcal{M}}(t_1,t_2,t_{\rm m}) &\equiv \langle\mathcal{M}(t_1)\mathcal{M}(t_2)\rangle - \langle\mathcal{M}(t_1)\rangle\langle\mathcal{M}(t_2)\rangle \\ &= \frac{2(k_BT)^2}{(t_{\rm m}-t_1)(t_{\rm m}-t_2)}\int_0^{t_{\rm m}-t_1}dt_{01}\int_0^{t_{\rm m}-t_2}dt_{02} \\ &\quad\times\left([U(t_1,t_2,t_{01},t_{02})]^2\right. \\ &\quad\left.+\frac{2}{k_BT}U_0(t_1,t_{01})U_0(t_2,t_{02})U(t_1,t_2,t_{01},t_{02})\right), \end{aligned} \tag{19}$$

where the functions U_0 and U are given by Eqs. (12) and (14). The variance of the time-average MSD is simply $C_{\mathcal{M}}(t,t,t_{\rm m})$. If the initial conditions are irrelevant, setting $x_0=0$ and $v_0=0$ yields a more compact formula for the variance:

$$\mathrm{var}\{\mathcal{M}(t,t_{\rm m})\} = \frac{2(k_BT)^2}{(t_{\rm m}-t)^2}\int_0^{t_{\rm m}-t}dt_{01}\int_0^{t_{\rm m}-t}dt_{02}[U(t,t,t_{01},t_{02})]^2. \tag{20}$$

C. SRMS

We start by noting that the definition (3) of SRMS between t_0 and t_0+t can equivalently be written as

$$r(t,t_0) = \frac{1}{t}\int_0^t dt'\left(X(t'+t_0)-\frac{1}{t'}\int_0^{t'}dt''X(t''+t_0)\right)^2 \tag{21}$$

[the equivalence is shown by taking the difference between Eqs. (3) and (21) and integrating by parts]. By technical reasons, we shall use Eq. (21) in the following derivations.

Denoting the time-averaged position of the tracer as

$$Y(t,t_0)\equiv\frac{1}{t}\int_0^t dt'X(t'+t_0),$$

using Eq. (8) and integrating by parts, one obtains

$$\begin{aligned} Y(t,t_0) &= \frac{1}{t}\int_0^t dt'X_0(t'+t_0)+\int_0^{t+t_0}dt'G(t+t_0-t')F(t') \\ &\quad-\frac{1}{t}\int_0^t dt't'\int_0^{t'+t_0}dt''g(t'+t_0-t'')F(t''), \end{aligned} \tag{22}$$

from which

$$\begin{aligned} r(t,t_0) &= \frac{1}{t}\int_0^t dt'\left(X_0(t'+t_0)-\frac{1}{t'}\int_0^{t'}dt''X_0(t''+t_0)\right. \\ &\quad\left.+\frac{1}{t'}\int_0^{t'}dt''t''\int_0^{t''+t_0}dt'''g(t''+t_0-t''')F(t''')\right)^2. \end{aligned} \tag{23}$$

The mean value of the SRMS from Eq. (21) can be split in two terms:

$$\langle r(t,t_0)\rangle = r_1(t,t_0)+\langle r_2(t,t_0)\rangle,$$

where

$$r_1(t,t_0) = \frac{1}{t}\int_0^t\frac{dt'}{t'^2}\,[J_0(t',t_0)]^2,$$

$$\begin{aligned} r_2(t,t_0) &= \frac{1}{t}\int_0^t dt_1 \\ &\quad\times\left(\frac{1}{t_1}\int_0^{t_1}dt't'\int_0^{t'+t_0}dt''g(t'+t_0-t'')F(t'')\right)^2, \end{aligned} \tag{24}$$

with

$$\begin{aligned} J_0(t,t_0) &\equiv tX_0(t+t_0)-\int_0^t dt'X_0(t'+t_0) \\ &= mv_0\hat{G}^{(0)}(t,t_0)-kx_0\hat{G}^{(1)}(t,t_0) \end{aligned} \tag{25}$$

and

$$\begin{aligned} \hat{G}^{(n)}(t,t_0) &\equiv t\ \mathrm{sgn}(t+t_0)G^{(n)}(|t+t_0|) \\ &\quad-G^{(n+1)}(|t+t_0|)+G^{(n+1)}(|t_0|) \end{aligned} \tag{26}$$

[the explicit form (9) and the condition $\langle F(t)\rangle=0$ were used]. After numerous integrations by parts of the first relation in Eq. (24), one gets

$$\begin{aligned} r_1(t,t_0) &= -\left(mv_0\frac{G^{(1)}(t+t_0)-G^{(1)}(t_0)}{t}\right. \\ &\quad\left.-kx_0\frac{G^{(2)}(t+t_0)-G^{(2)}(t_0)}{t}\right)^2 \\ &\quad+\frac{1}{t}\int_{t_0}^{t+t_0}dt'(mv_0G(t')-kx_0G^{(1)}(t'))^2. \end{aligned} \tag{27}$$

The second relation in Eq. (24) can be explicitly written by using the fluctuation-dissipation relation (5) as

$$\langle r_2(t,t_0)\rangle = \frac{k_BT}{t}\int_0^t dt'\frac{J(t',t',t_0,t_0)}{t'^2}, \tag{28}$$

where

$$J(t_1,t_2,t_{01},t_{02})\equiv\int_0^{t_1}dt_1't_1'\int_0^{t_2}dt_2't_2'H(t_1'+t_{01},t_2'+t_{02}), \tag{29}$$

and the function $H(t_1,t_2)$, which was given by Eq. (16), represents

$$H(t_1,t_2)=\int_0^{t_1}dt_1' g(t_1-t_1')\int_0^{t_2}dt_2' g(t_2-t_2')\gamma(|t_1'-t_2'|).$$

Substituting Eq. (16) into Eq. (29) and integrating by parts, one finds

$$\begin{aligned}J(t_1,t_2,t_{01},t_{02})&=\hat{G}^{(2)}(t_1,t_{01}-t_{02})+\hat{G}^{(2)}(t_2,t_{02}-t_{01})\\&\quad-\hat{G}^{(2)}(t_1-t_2,t_{01}-t_{02})-t_1t_2G^{(1)}(|t_1-t_2+t_{01}-t_{02}|)\\&\quad-m\hat{G}^{(0)}(t_1,t_{01})\hat{G}^{(0)}(t_2,t_{02})-k\hat{G}^{(1)}(t_1,t_{01})\hat{G}^{(1)}(t_2,t_{02}).\end{aligned}\tag{30}$$

In particular, one gets

$$J(t,t,t_0,t_0)=2\hat{G}^{(2)}(t,0)-m[\hat{G}^{(0)}(t,t_0)]^2-k[\hat{G}^{(1)}(t,t_0)]^2.\tag{31}$$

Substituting this result into Eq. (28) and integrating by parts the first term, one finally obtains

$$\begin{aligned}\langle r_2(t,t_0)\rangle&=R(t)+k_BT\Bigg(m\frac{[G^{(1)}(t+t_0)-G^{(1)}(t_0)]^2}{t^2}\\&\quad+k\frac{[G^{(2)}(t+t_0)-G^{(2)}(t_0)]^2}{t^2}\\&\quad-\frac{1}{t}\int_{t_0}^{t+t_0}dt'\{m[G(t')]^2+k[G^{(1)}(t')]^2\}\Bigg),\end{aligned}\tag{32}$$

where

$$R(t)=2k_BT\frac{G^{(3)}(t)}{t^2}.\tag{33}$$

These equations provide a general solution for the mean value of the SRMS for arbitrary memory kernel $\gamma(t)$.

D. Time-averaged SRMS

Since the SRMS involves only a part of the sample, high fluctuations are expected at short lag times t. Averaging this quantity in Eq. (2) over the whole sample reduces the fluctuations. Substituting Eqs. (27) and (32) into Eq. (2) and integrating by parts the last term in order to eliminate the double integral, one obtains the mean $\mathcal{R}(t,t_m)$ as the sum of two terms:

$$\begin{aligned}\mathcal{R}_1(t,t_m)&=-\frac{1}{t^2}\frac{1}{t_m-t}\int_0^{t_m-t}dt_0\{\big[mv_0[G^{(1)}(t_0+t)-G^{(1)}(t_0)]-kx_0[G^{(2)}(t_0+t)-G^{(2)}(t_0)]\big]^2\\&\quad+tt_0\big[[mv_0G(t+t_0)-kx_0G^{(1)}(t+t_0)]^2-[mv_0G(t_0)-kx_0G^{(1)}(t_0)]^2\big]\}+\frac{1}{t}\int_{t_m-t}^{t_m}dt'(mv_0G(t')-kx_0G^{(1)}(t'))^2,\\\langle\mathcal{R}_2(t,t_m)\rangle&=R(t)+\frac{k_BT}{t^2}\frac{1}{t_m-t}\int_0^{t_m-t}dt_0\{m\big[[G^{(1)}(t_0+t)-G^{(1)}(t_0)]^2+tt_0([G(t_0+t)]^2-[G(t_0)]^2)\big]\\&\quad+k\big[[G^{(2)}(t_0+t)-G^{(2)}(t_0)]^2+tt_0([G^{(1)}(t_0+t)]^2-[G^{(1)}(t_0)]^2)\big]\}-\frac{k_BT}{t}\int_{t_m-t}^{t_m}dt'(m[G(t')]^2+k[G^{(1)}(t')]^2).\end{aligned}\tag{34}$$

Although these expressions are cumbersome, their numerical computation for a given kernel $G(t)$ is rather straightforward. In what follows, we shall show that, if the sample duration t_m is long enough, the main contribution to $\langle\mathcal{R}(t,t_m)\rangle$ is provided by the simple term $R(t)$. More importantly, the time-averaged SRMS will be shown to have smaller fluctuations than $\mathcal{M}(t,t_m)$. For this purpose, the covariance of the time-averaged SRMS is computed in Appendix B:

$$\begin{aligned}C_{\mathcal{R}}(t_1,t_2,t_m)&\equiv\langle\mathcal{R}(t_1)\mathcal{R}(t_2)\rangle-\langle\mathcal{R}(t_1)\rangle\langle\mathcal{R}(t_2)\rangle\\&=\frac{2(k_BT)^2}{t_1t_2(t_m-t_1)(t_m-t_2)}\int_0^{t_m-t_1}dt_{01}\int_0^{t_m-t_2}dt_{02}\\&\quad\times\int_0^{t_1}\frac{dt_1'}{(t_1')^2}\int_0^{t_2}\frac{dt_2'}{(t_2')^2}\Bigg([J(t_1',t_2',t_{01},t_{02})]^2\\&\quad+\frac{2}{k_BT}J_0(t_1',t_{01})J_0(t_2',t_{02})J(t_1',t_2',t_{01},t_{02})\Bigg).\end{aligned}\tag{35}$$

One can note that this expression resembles Eq. (19) for the covariance of the time-averaged MSD, but two additional integrals (over t_1' and t_2') stand for the moving average inside the chosen window. The variance of the time-averaged SRMS is simply $C_{\mathcal{R}}(t,t,t_m)$. If the initial conditions are irrelevant, setting $x_0=0$ and $v_0=0$ yields a more compact formula for the variance:

$$\begin{aligned}\mathrm{var}\{\mathcal{R}(t,t_m)\}&=\frac{2(k_BT)^2}{t^2(t_m-t)^2}\int_0^{t_m-t}dt_{01}\int_0^{t_m-t}dt_{02}\\&\quad\times\int_0^t\frac{dt_1}{t_1^2}\int_0^t\frac{dt_2}{t_2^2}[J(t_1,t_2,t_{01},t_{02})]^2.\end{aligned}\tag{36}$$

III. SUBDIFFUSION WITH POWER-LAW MEMORY KERNEL

A power-law memory kernel

$$\gamma(t)=\frac{\gamma_\alpha t^{-\alpha}}{\Gamma(1-\alpha)}\quad(0<\alpha<1)\tag{37}$$

is often considered as a paradigm of strongly correlated processes and is known to lead to subdiffusive behavior with

the scaling exponent α (here γ_α is a generalized friction coefficient in units kg sec$^{\alpha-2}$).[1] For this case, Desposito and Vinales used the recipes from [39] to provide the explicit representation of the kernels $g(t)$, $G(t)$, and $G^{(1)}(t)$ [35], e.g.,

$$G(t) = \frac{t}{m}\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}(t^2k/m)^n E^{(n)}_{2-\alpha,2+\alpha n}(-\gamma_\alpha t^{2-\alpha}/m), \tag{38}$$

where $E^{(n)}_{\alpha,\beta}(z)$ is the nth order derivative of the Mittag-Leffler function $E_{\alpha,\beta}(z)$,

$$E^{(n)}_{\alpha,\beta}(z) = \sum_{j=0}^{\infty}\frac{(j+n)!z^j}{j!\Gamma[\alpha(j+n)+\beta]},$$

and the Mittag-Leffler function is

$$E_{\alpha,\beta}(z) = \sum_{n=0}^{\infty}\frac{z^n}{\Gamma(\alpha n+\beta)}, \tag{39}$$

with $\Gamma(z)$ being the Gamma function. It is worth stressing that $G(0) = 0$ whenever $m > 0$ as we assumed in Sec. II.

Although the representation (38) of the kernel $G(t)$ is exact and explicit, a theoretical investigation of the mean and variance of the time-averaged MSD and SRMS is rather sophisticated. To further proceed, we introduce three time scales of the GLE (4):

$$\tau_{\rm in,fr} = \left(\frac{m}{\gamma_\alpha}\right)^{1/(2-\alpha)}, \quad \tau_{\rm in,ha} = \sqrt{\frac{m}{k}}, \quad \tau_{\rm fr,ha} = \left(\frac{\gamma_\alpha}{k}\right)^{1/\alpha}.$$

These time scales describe the balance between the inertial term and the frictional force, the inertial term and the harmonic force, and the frictional and harmonic forces, respectively (after normalization by the thermal force). Comparing t with these time scales, one can distinguish different regimes of the Langevin dynamics. An estimation of the time scales may allow one to simplify the problem by neglecting the forces which are irrelevant for given experimental setup. In what follows, we consider two limiting cases which are important for practical applications.

(i) When there is no harmonic potential ($k = 0$), two time scales are infinite, $\tau_{\rm in,ha} = \tau_{\rm fr,ha} = \infty$, while the remaining $\tau_{\rm in,fr}$ controls the Langevin dynamics. All the terms in Eq. (38) vanish except that for $n = 0$:

$$G(t) = \frac{t}{m}E_{2-\alpha,2}(-\gamma_\alpha t^{2-\alpha}/m).$$

Using the identity [39]

$$\frac{\partial}{\partial t}t^{\beta-1}E_{\alpha,\beta}(-ct^\alpha) = t^{\beta-2}E_{\alpha,\beta-1}(-ct^\alpha), \tag{40}$$

one deduces

$$G^{(n)}(t) = \frac{t^{1+n}}{m}E_{2-\alpha,2+n}(-\gamma_\alpha t^{2-\alpha}/m) \tag{41}$$

[note that $g(t)$ is obtained for $n = -1$]. We shall briefly discuss this case in Appendix C.

(ii) In many microrheological experiments, the size of tracer particles is below a micron, and the inertial effects can often be neglected. In this overdamped case ($m = 0$), two time scales are zero, $\tau_{\rm in,fr} = \tau_{\rm in,ha} = 0$, while the remaining $\tau_{\rm fr,ha}$ controls the Langevin dynamics. Taking the limit $m \to 0$ and substituting $\tilde\gamma(s) = \gamma_\alpha s^{\alpha-1}$ in Eq. (7), one gets $\tilde G(s) = 1/(\gamma_\alpha s^\alpha + k)$, from which

$$G(t) = \frac{t^{\alpha-1}}{\gamma_\alpha}E_{\alpha,\alpha}(-kt^\alpha/\gamma_\alpha), \tag{42}$$

while the integrals and derivatives of $G(t)$ can also be represented through Mittag-Leffler functions using the identity (40):

$$G^{(n)}(t) = \frac{t^{\alpha+n-1}}{\gamma_\alpha}E_{\alpha,\alpha+n}(-kt^\alpha/\gamma_\alpha). \tag{43}$$

Note that Eq. (42) could also be directly retrieved from Eq. (38) in the limit $m \to 0$ by using the asymptotic formula for Mittag-Leffler functions:

$$E_{\alpha,\beta}(-z) \simeq \sum_{n=1}^{N}\frac{(-1)^{n-1}}{z^n\Gamma(\beta-\alpha n)} \qquad (z\to\infty), \tag{44}$$

with an appropriate order N. Although this asymptotic series diverges as $N \to \infty$, it can still be used for an accurate computation of Mittag-Leffler functions for large negative arguments.

In contrast to the general case, for which $G(0) = 0$, the kernel $G(t)$ from Eq. (42) diverges at $t = 0$ for massless particles with $\alpha < 1$. This is related to the fact that the instantaneous velocity of Brownian motion is not well defined. Nevertheless, although the assumption $G(0) = 0$ fails for massless tracers, the results of Sec. II are correct. For instance, a more careful revision of the derivation for $\langle r_2(t,t_0)\rangle$ in Sec. II C shows that supplementary terms containing $G(0)$ which would appear in Eq. (22) after integration by parts will be canceled by similar terms which would also appear in Eq. (31). One can consider the inertial term (with small $m > 0$) as a regularization of the problem. Once the results are derived, one can take the limit $m \to 0$ to eliminate this regularization.

In what follows, we illustrate the applications of the time-averaged MSD and SRMS by considering subdiffusion of massless tracers in a harmonic potential.

A. Time-averaged MSD

Neglecting the mass in Eqs. (11) and (17) and using the explicit formulas for $G(t)$ and $G^{(1)}(t)$ and the identity

$$xE_{\alpha,\alpha+\beta}(-x) = 1/\Gamma(\beta) - E_{\alpha,\beta}(-x), \tag{45}$$

one gets

$$\langle\mathcal{M}(t,t_{\rm m})\rangle = M(t) - \left(\frac{k_BT}{k} - x_0^2\right)M_\alpha(t/\tau,t_{\rm m}/\tau), \tag{46}$$

where $\tau = (\gamma_\alpha/k)^{1/\alpha}$ denotes the time scale $\tau_{\rm fr,ha}$,

$$M(t) = \frac{2k_BT}{k}[1 - E_{\alpha,1}(-kt^\alpha/\gamma_\alpha)], \tag{47}$$

[1]Note that for normal diffusion without memory (with $\alpha = 1$), Eq. (37) is replaced by $\gamma(t) = 2\gamma_1\delta(t)$, where $\delta(t)$ is the Dirac distribution.

and

$$M_\alpha(z,\zeta) = \frac{1}{\zeta - z}\int_0^{\zeta - z} dx (E_{\alpha,1}(-x^\alpha) - E_{\alpha,1}[-(x+z)^\alpha])^2. \tag{48}$$

Although this integral is not reduced to tabulated special functions, its numerical computation is straightforward.

For normal diffusion ($\alpha = 1$), one gets

$$M_1(z,\zeta) = \frac{1 - e^{-2(\zeta - z)}}{2(\zeta - z)}(1 - e^{-z})^2. \tag{49}$$

The first term in Eq. (46) is independent of the sample duration t_m and provides the main contribution to the mean time-averaged MSD which was previously given by Desposito and Vinales [38]. As shown in Appendix D, the function $M_\alpha(z,\zeta)$ decreases as ζ^{-1} for large ζ (large t_m), so that $M(t)$ becomes exact for infinitely long samples. In practice, the sample duration is finite, and the correction $M_\alpha(z,\zeta)$ can be significant when t is comparable to t_m. Figure 1 illustrates this effect and shows that missing the correction term may lead to underestimation of the plateau level $2k_BT/k$ when the sample duration is not too long as compared to the intrinsic time τ. The explicit form (48) allows one to investigate the role of this correction and the ranges of applicability of the approximation $M(t)$.

In the short-time limit, the main contribution comes again from $M(t)$, while the function $M_\alpha(z,\zeta)$ provides the next-order correction, as shown in Appendix D:

$$M_\alpha(z,\zeta) \simeq \frac{z^2}{\zeta}\int_0^\zeta dx \frac{[E_{\alpha,0}(-x^\alpha)]^2}{x^2} + O(z^3). \tag{50}$$

When $1/2 < \alpha < 1$, the integral converges in the limit $\zeta \to \infty$ so that this term remains in the order of z^2/ζ. One retrieves therefore the classical result for subdiffusion:

$$\langle \mathcal{M}(t,t_m)\rangle \simeq \frac{2k_BT}{\gamma_\alpha}\frac{t^\alpha}{\Gamma(1+\alpha)} + O(t^{2\alpha}), \tag{51}$$

where the prefactor k_BT/γ_α is called the generalized diffusion coefficient. Moreover, Eq. (51) becomes exact (without correction terms) in the limit $k \to 0$ (no harmonic potential).

In the long-time limit $t \to t_m$, the function $M_\alpha(z,\zeta)$ is shown to be comparable to $M(t)$ (see Appendix D). Using Eq. (D3), one gets

$$\langle \mathcal{M}(t,t_m)\rangle \simeq \frac{2k_BT}{k}\left(\frac{1}{2} + \frac{k(t_m - t)^\alpha/\gamma_\alpha}{\Gamma(\alpha+2)}\right) + O\left(t_m^{-\alpha}\right)$$

(here the initial condition was omitted by setting $x_0 = 0$). In particular, the limiting value $\lim_{t_m\to\infty}\langle \mathcal{M}(t_m,t_m)\rangle$ is twice smaller than the expected value $2k_BT/k$. This behavior is clearly seen in Fig. 1 where three curves (solid, dashed, and dash-dotted) approach the level $1/2$ and even descend below it (because $t_m/\tau = 10$ is not large enough for omitting next-order corrections).

B. Time-averaged SRMS

For massless particles, the time-averaged SRMS is obtained by substituting Eq. (43) into Eqs. (34), using the identity (45)

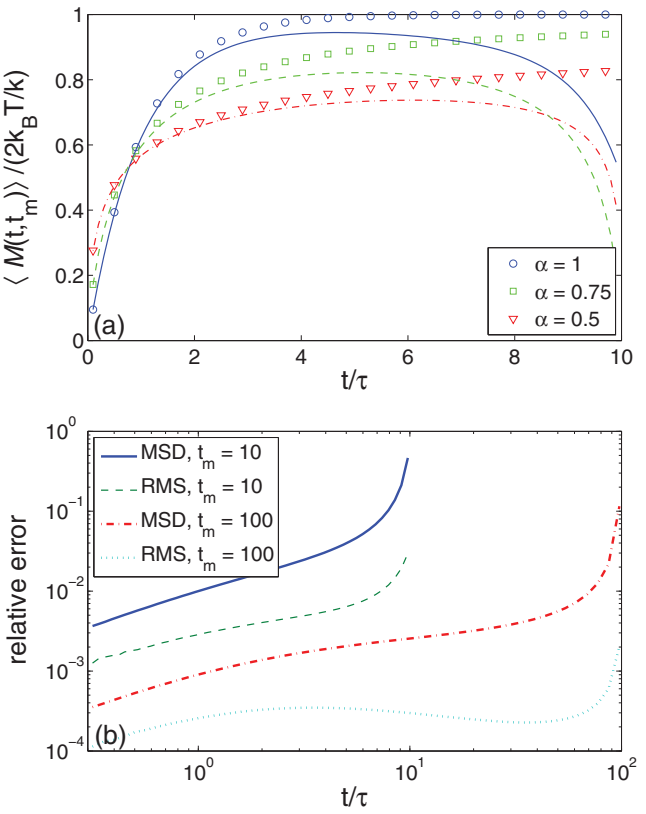

FIG. 1. (Color online) (a) Mean time-averaged MSD as a function of the lag time t for $t_m = 10\tau$ and three values of α: 1, 0.75, and 0.5. Symbols show the main contribution $M(t)$ from Eq. (47), while lines present the exact formula (46) in the presence of the function $M_\alpha(z,\zeta)$. When $t \ll t_m$, the correction is small. For larger t, a significant deviation from the plateau is observed. The approximation by $M(t)$ may underestimate the factor $2k_BT/k$ in Eq. (46). (b) The ratio $M_\alpha(t/\tau,t_m/\tau)/\langle \mathcal{M}(t,t_m)\rangle$ [respectively $R_\alpha(t/\tau,t_m/\tau)/\langle \mathcal{R}(t,t_m)\rangle$] which represents the relative error of using $M(t)$ [respectively $R(t)$] instead of the mean value $\langle \mathcal{M}(t,t_m)\rangle$ [respectively $\langle \mathcal{R}(t,t_m)\rangle$], for $\alpha = 0.75$ and two sample durations $t_m = 10\tau$ and $t_m = 100\tau$. As expected, the increase of t_m leads to smaller relative errors. The approximation of the time-averaged SRMS is more accurate than that of the time-averaged MSD.

and integrating by parts:

$$\langle \mathcal{R}(t,t_m)\rangle = R(t) - \left(\frac{k_BT}{k} - x_0^2\right)R_\alpha(t/\tau,t_m/\tau), \tag{52}$$

where the main contribution is

$$R(t) = \frac{k_BT}{k}[1 - 2E_{\alpha,3}(-kt^\alpha/\gamma_\alpha)], \tag{53}$$

while the correction term is determined by the function

$$R_\alpha(z,\zeta) \equiv \frac{1}{z}\int_{\zeta - z}^\zeta dx[E_{\alpha,1}(-x^\alpha)]^2 - \frac{1}{z^2(\zeta - z)} \times \int_0^{\zeta - z} dx\{((x+z)E_{\alpha,2}[-(x+z)^\alpha] - xE_{\alpha,2}(-x^\alpha))^2 + xz(\{E_{\alpha,1}[-(x+z)^\alpha]\}^2 - [E_{\alpha,1}(-x^\alpha)]^2)\}. \tag{54}$$

In Appendix D, the short-time asymptotic behavior of the correction term for $\zeta \gg 1$ is shown to be

$$R_\alpha(z,\zeta) \simeq \frac{z^2}{12\zeta}\int_0^\zeta dx \frac{[E_{\alpha,0}(-x^\alpha)]^2}{x^2} + O(z^3) \quad (z \ll 1), \tag{55}$$

which is 12 times smaller than similar asymptotic (50) for the time-averaged MSD. In the opposite limit $z \to \zeta$ (i.e., $t \to t_{\rm m}$), $R_\alpha(z,\zeta) \sim \zeta^{-1}$ (when $1/2 < \alpha < 1$). We also checked numerically that the correction term $R_\alpha(z,\zeta)$ monotonously increases with z. This means that $R_\alpha(z,\zeta)$ is the next-order correction to $R(t)$ in both short-time and long-time limits (except for $\alpha = 1$, which can be considered separately). Neglecting the correction term $R_\alpha(z,\zeta)$, the asymptotic behavior of the mean time-averaged SRMS is derived from that of the function $R(t)$:

$$\begin{aligned} \langle \mathcal{R}(t,t_{\rm m})\rangle &\simeq \frac{2k_BT}{\gamma_\alpha}\frac{t^\alpha}{\Gamma(\alpha+3)} + O(t^{2\alpha}) \quad (t \ll \tau), \\ \langle \mathcal{R}(t,t_{\rm m})\rangle &\simeq \tfrac{k_BT}{k}\left(1 - \frac{2(kt^\alpha/\gamma_\alpha)^{-1}}{\Gamma(3-\alpha)}\right) + O(t^{-1}) (t \gg \tau). \end{aligned} \tag{56}$$

Note that the next-order correction $O(t^{-1})$ in the last relation becomes of the same order as the main term when $\alpha = 1$.

It is worth noting that the starting point x_0 stands in front of the correction term $R_\alpha(z,\zeta)$ whose contribution can be neglected in both asymptotic limits. The initial condition has therefore a weak influence on the solution, as expected.

For normal diffusion ($\alpha = 1$), the identity

$$E_{1,k}(x) = \frac{1}{x^{k-1}}\left(e^x - \sum_{n=0}^{k-2}\frac{x^n}{n!}\right) \tag{57}$$

yields

$$\begin{aligned} R(t) &= \frac{k_BT}{k}\left(1 - \frac{2}{kt/\gamma_1} + 2\frac{1-e^{-kt/\gamma_1}}{(kt/\gamma_1)^2}\right), \\ R_1(z,\zeta) &= \frac{1-e^{-2(\zeta-z)}}{2(\zeta-z)}\left(\frac{1-e^{-2z}}{2z} - \frac{(1-e^{-z})^2}{z^2}\right). \end{aligned} \tag{58}$$

C. Variance

It is worth recalling that both $\mathcal{M}(t,t_{\rm m})$ and $\mathcal{R}(t,t_{\rm m})$ are random variables, as they are defined through a random trajectory $X(t)$. Although the time averaging is used to reduce fluctuations, quantitative measures of these fluctuations are needed to know whether the empirical time-averaged MSD or SRMS is a reliable estimate of its ensemble-averaged mean. For this purpose, we consider the ratio between the standard deviation (the square root of the variance) and the mean of $\mathcal{M}(t,t_{\rm m})$ or $\mathcal{R}(t,t_{\rm m})$. This ratio can be called the *statistical uncertainty* because it quantifies how accurately a single random realization of $\mathcal{M}(t,t_{\rm m})$ or $\mathcal{R}(t,t_{\rm m})$ determines the mean value: the smaller this ratio is, the better the estimation of the mean is. The knowledge of the statistical uncertainty would help one to choose the experimental setup appropriately and to interpret the acquired data correctly.

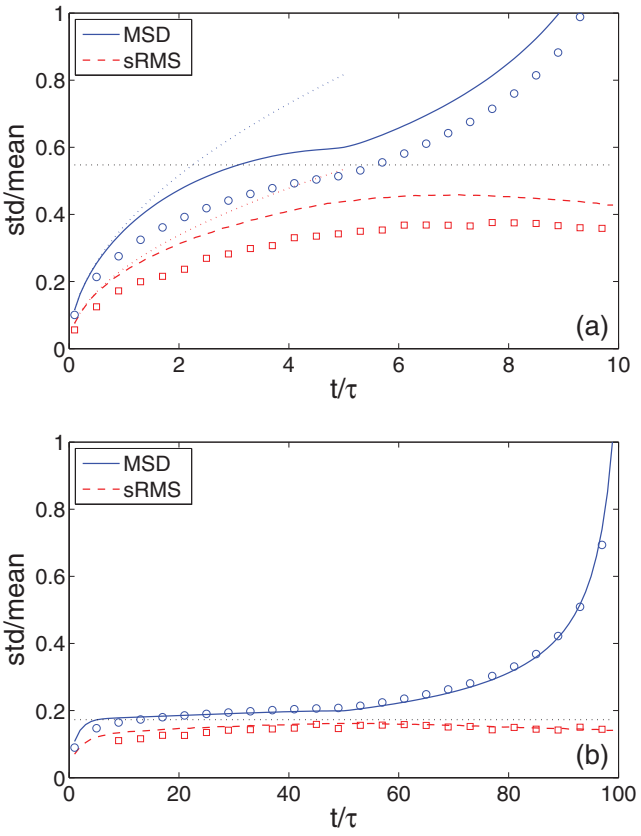

FIG. 2. (Color online) Statistical uncertainty of the time-averaged MSD (solid blue curve) and SRMS (dashed red curve) with $\alpha = 1$, their short-time asymptotics (59), (61), and intermediate level (60) shown by dotted lines, respectively. The sample duration is $t_{\rm m} = 10\tau$ (a) and $t_{\rm m} = 100\tau$ (b). For comparison, the statistical uncertainty of MSD and SRMS for subdiffusion with $\alpha = 0.75$ is shown by circles and squares, respectively. Although the curves are different, they feature similar behavior. Interestingly, the statistical uncertainty is smaller for $\alpha = 0.75$.

In Sec. II, we gave the general formulas (20) and (36) for the variances of the time-averaged MSD and SRMS, respectively. These formulas express the variance as double and quadruple integral of an explicit function which is related to the kernel $G(t)$ and its iterative integrals. For massless particles, one can compute the variance numerically by using the explicit formula (43) of the kernel. Although further simplifications seem to be feasible, a direct numerical integration is used in this paper.

In what follows, we consider the case when $t_{\rm m} \gg \tau$ [$\tau = (\gamma_\alpha/k)^{1/\alpha}$], while the lag time t varies between 0 and $t_{\rm m}$.

1. *Time-averaged MSD*

We first consider normal diffusion ($\alpha = 1$) for which double integrals in Eqs. (20) can be evaluated analytically. Figure 2 shows the statistical uncertainty of the time-averaged MSD as a function of the lag time t. One can distinguish three different regimes for which the approximate relations are derived in Appendix A.

(i) For very small lag time t, the statistical uncertainty behaves as

$$\frac{\mathrm{std}\{\mathcal{M}(t,t_\mathrm{m})\}}{\langle\mathcal{M}(t,t_\mathrm{m})\rangle} \approx \frac{2}{\sqrt{3}}\sqrt{t/t_\mathrm{m}} \quad (t \ll \tau \ll t_\mathrm{m}). \tag{59}$$

In this limit, the empirical time-averaged MSD is a reliable measure of its mean because fluctuations are small. For instance, the results with 10% statistical uncertainty are acquired for $t \leqslant 0.015 t_\mathrm{m}$. This inequality can be used for choosing the appropriate sample duration t_m. Note that in this regime, an accurate study of the behavior up to time t requires the acquisition of an approximately 100 times longer data sample.

(ii) When the lag time becomes in the order of τ (but still much smaller than t_m), the MSD curve starts to approach a plateau, and so does the variance. In this regime, the statistical uncertainty approaches a constant level:

$$\frac{\mathrm{std}\{\mathcal{M}(t,t_\mathrm{m})\}}{\langle\mathcal{M}(t,t_\mathrm{m})\rangle} \approx \sqrt{3\tau/t_\mathrm{m}} \quad (\tau \ll t \ll t_\mathrm{m}). \tag{60}$$

For instance, in order to measure the plateau with 10% statistical uncertainty, one has to choose the sample duration $t_\mathrm{m} \simeq 300\tau$.

(iii) Finally, when the lag time approaches t_m, the statistical uncertainty further increases and reaches the limit value $\sqrt{2} \approx 1.41\ldots$, independently of τ and t_m. In this case, an estimate of the mean value from a random $\mathcal{M}(t,t_\mathrm{m})$ is meaningless.

Although the above analysis was performed for normal diffusion ($\alpha = 1$), the conclusions are applicable for subdiffusion as well ($\alpha < 1$). As shown in Fig. 2, the statistical uncertainty behaves similarly for $\alpha = 0.75$. Moreover, the statistical uncertainty seems to decrease when α decreases.

2. *Time-averaged SRMS*

Even for normal diffusion, an analytical computation of the quadruple integral in Eq. (36) is cumbersome. For this reason, we restricted the analysis to the two limiting cases: no harmonic potential ($k = 0$) which is equivalent to the short-time limit, and the limit $t \to t_\mathrm{m}$.

(i) When there is no harmonic potential, the variance of the time-averaged SRMS is computed analytically and given by Eq. (B1). In the short-time limit, $R(t) = (k_B T/\gamma_1)t/3$ and the statistical uncertainty behaves as

$$\frac{\mathrm{std}\{\mathcal{R}(t,t_\mathrm{m})\}}{\langle\mathcal{R}(t,t_\mathrm{m})\rangle} \approx \frac{2}{\sqrt{7}}\sqrt{t/t_\mathrm{m}}. \tag{61}$$

The comparison with Eq. (59) for the time-averaged MSD shows a smaller statistical uncertainty for SRMS (prefactor $2/\sqrt{7}$ instead of $2/\sqrt{3}$). Figure 2 illustrates this short-time behavior (see dotted curve with smaller amplitude).

(ii) In the limit $t \to t_\mathrm{m}$, the formula for the variance can be simplified (see Appendix B). For normal diffusion, the statistical uncertainty is shown to be

$$\frac{\mathrm{std}\{\mathcal{R}(t_\mathrm{m},t_\mathrm{m})\}}{\langle\mathcal{R}(t_\mathrm{m},t_\mathrm{m})\rangle} \simeq \sqrt{2\tau/t_\mathrm{m}}, \tag{62}$$

in sharp contrast to the case of MSD for which the statistical uncertainty approached $\sqrt{2}$. Taking the sample duration large enough (in comparison to the time scale τ), the statistical uncertainty of the time-averaged SRMS can be made small at long lag times. We conclude that the time-averaged SRMS provides a better estimator for random trajectories.

The statistical uncertainty for subdiffusion ($\alpha < 1$), which was computed numerically, exhibits very similar features (Fig. 2). Interestingly, smaller values of α lead to smaller statistical uncertainties. Note that the asymptotic formula (62) which was derived for $\alpha = 1$, can still be used for $\alpha < 1$ as a rough estimate of fluctuations.

IV. NUMERICAL SIMULATIONS

In order to illustrate the fluctuations of the time-averaged MSD and SRMS, we simulate random trajectories of subdiffusing tracers governed by the GLE (4). We focus on massless tracers ($m = 0$) undergoing the dynamics with a power-law memory kernel (37). In this case, one can use the explicit representation (8) of the trajectory $X(t)$ in the form of a convolution between the kernel $G(t)$ given by Eq. (42) and the Gaussian correlated noise $F(t)$. We set $x_0 = 0$ to explicitly eliminate the irrelevant contribution $X_0(t)$. For a small time step δ, the vector of positions $x_n = X(n\delta)$ is approximated as

$$x_n \approx \delta \sum_{n'=0}^{n-1} \frac{\delta^{\alpha-1}}{\gamma_\alpha}(n-n')^{\alpha-1} E_{\alpha,\alpha}\left(-\frac{k\delta^\alpha}{\gamma_\alpha}(n-n')^\alpha\right) F_0 f_{n'},$$

where $F_0 = \sqrt{k_B T \gamma_\alpha \delta^{-\alpha}/\Gamma(1-\alpha)}$ and $\{f_n\}$ is the vector of Gaussian random variables with mean zero and covariance $\langle f_n f_{n'}\rangle = |n-n'|^{-\alpha}$ for $n \neq n'$, and $\langle f_n^2\rangle = f^{(0)}$. The choice of the constant $f^{(0)}$ is a subtle point. In fact, in the continuous case, the covariance of the thermal force diverges at $t = t'$ according to Eq. (37). We fix the constant $f^{(0)}$ by approximating the integral of the memory kernel from 0 to t_m. On the one hand,

$$\int_0^{t_\mathrm{m}} dt\gamma(t) = \frac{\gamma_\alpha}{\Gamma(1-\alpha)}\int_0^{t_\mathrm{m}} dt t^{-\alpha} = \frac{\gamma_\alpha t_\mathrm{m}^{1-\alpha}}{\Gamma(2-\alpha)}.$$

On the other hand, the same integral can be approximated as

$$\int_0^{t_\mathrm{m}} dt\gamma(t) \approx \delta\left[\frac{\gamma^{(0)}}{2} + \frac{\gamma_\alpha}{\Gamma(1-\alpha)}\left(\sum_{n=1}^{N-1} n^{-\alpha} + \frac{N^{-\alpha}}{2}\right)\right]$$

(where $N = t_\mathrm{m}/\delta$), from which one gets the discrete approximation $\gamma^{(0)}$ of the divergent $\gamma(0)$. Since $k_B T\gamma^{(0)} = F_0 f^{(0)}$ according to Eq. (5), one obtains

$$f^{(0)} = 2\left(\frac{N^{1-\alpha}}{1-\alpha} - \frac{N^{-\alpha}}{2} - \sum_{n=1}^{N-1} n^{-\alpha}\right).$$

Although the constant $f^{(0)}$ formally depends on N (discretization of time), we checked numerically that the right-hand side rapidly converges to a limit when N increases.

A numerical generation of the correlated Gaussian noise $\{f_n\}$ with a given covariance matrix was realized by the Davies and Harte method [40–42]. This method relies on the embedding of the Toepliz covariance matrix in a non-negative definite symmetric circulant matrix for which fast Fourier transform allows one to generate the correlated Gaussian noise within $O(N\ln N)$ operations.

The time-averaged MSD and SRMS of a simulated trajectory $\{x_n\}$ are computed as discrete analogs of Eqs. (1) and (2):

$$\mathcal{M}(t,t_{\rm m}) = \frac{1}{N-n+1}\sum_{k=0}^{N-n}(x_{k+n}-x_k)^2,$$

$$\mathcal{R}(t,t_{\rm m}) = \frac{1}{N-n+1}\sum_{k=0}^{N-n}\left[\frac{1}{n}\sum_{j=1}^{n}x_{k+j}^2 - \left(\frac{1}{n}\sum_{j=1}^{n}x_{k+j}\right)^2\right],$$

where $n = t/\delta$ and $N = t_{\rm m}/\delta$. For a faster numerical computation of the second relation, one can introduce two auxiliary vectors,

$$y_k = \sum_{j=1}^{n} x_{k+j}, \qquad z_k = \sum_{j=1}^{n} x_{k+j}^2,$$

which can be computed iteratively:

$$y_0 = \sum_{j=1}^{n} x_j, \quad y_k = y_{k-1} + x_{k+n} - x_k \quad (k = 1,2,\ldots N-n)$$

(similarly for z_k), from which

$$\mathcal{R}(t,t_{\rm m}) = \frac{1}{N-n+1}\left[\frac{1}{n}\sum_{k=0}^{N-n} z_k - \frac{1}{n^2}\sum_{k=0}^{N-n} y_k^2\right].$$

For numerical simulations, the dimensionless physical parameters are used: $k_BT = 1$, $\gamma_\alpha = 1$, and $k = 1$. The time step δ is set to be 0.0001, while the sample duration is either $t_{\rm m} = 10$ or $t_{\rm m} = 100$ (with $N = 10^5$ and $N = 10^6$ discrete steps, respectively).

For subdiffusion with $\alpha = 0.75$, Figs. 3 and 4 show the time-averaged MSD and SRMS divided by k_BT/k, as a function of the normalized time t/τ, where $\tau = (\gamma_\alpha/k)^{1/\alpha} = 1$. The theoretical mean value and the range within plus and minus one standard deviation are shown and compared to the empirical time-averaged MSD and SRMS computed from five generated trajectories.

For a moderate sample duration $t_{\rm m}/\tau = 10$, one concludes the following about the time-averaged MSD:

1. The correction term in Eq. (46) is significant so that the mean time-averaged MSD does not reach the asymptotic plateau $2k_BT/k$ (there is a gap between the bold solid curve and dashed level at 2); in this regime, the use of the simplified formula (47) would lead to an underestimation of the plateau value $2k_BT/k$.
2. The fluctuations are strong, the standard deviation being on the order of the mean value; fitting the simulated data would give a broad distribution of the fitting parameters (i.e., the plateau $2k_BT/k$, the exponent α, and the time scale τ).
3. As expected, the fluctuations grow with time.

For a longer sample ($t_{\rm m} = 100$), the plateau is better established, with smaller fluctuations. In this case, the simplified formula (47) would give a reasonable fit.

The time-averaged SRMS exhibits different features. For both durations $t_{\rm m} = 10$ and $t_{\rm m} = 100$, the empirical SRMS curves are much smoother than the empirical MSD curves. Moreover, the empirical SRMS curves are closer to the mean curve obtained from Eq. (52) resulting in a more accurate fitting. We conclude that the time-averaged SRMS provides a more reliable way for the statistical analysis of individual single-particle trajectories. A potential drawback of this functional is a slower convergence to a plateau level k_BT/k, which is expected from Eqs. (53) and (58).

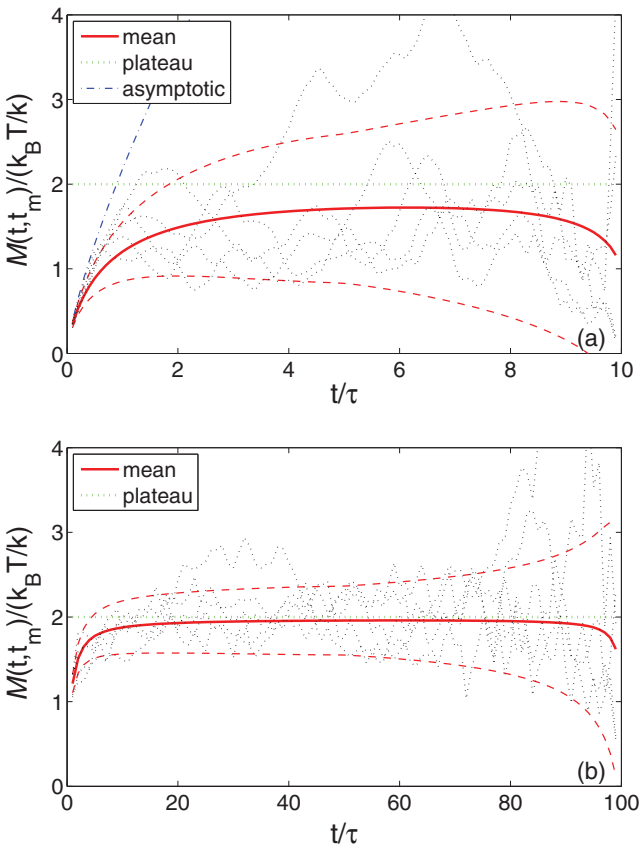

FIG. 3. (Color online) The normalized time-averaged MSD for subdiffusion with $\alpha = 0.75$, $\gamma_\alpha = 1$, $k = 1$, $k_BT = 1$, $\tau = 1$, $t_{\rm m} = 10$ (a), and $t_{\rm m} = 100$ (b). The mean MSD from Eq. (46) (solid red curve) is compared to its asymptotic short-time behavior (59) (dash-dotted blue curve) and to five empirical time-averaged MSDs from simulated trajectories (dotted black curves). The dashed red curves show the mean MSD plus or minus its standard deviation computed from Eq. (20). The expected plateau level at 2 is shown by dotted green line.

At this point, we emphasize again the challenge in the characterization of a stochastic process from its single trajectory. Even for simulated data, for which the underlying model of thermal fluctuations and the setup parameters are precisely known, while other sources of fluctuations (e.g., measurement errors) are excluded, fitting the empirical curves may be problematic at long times.

The presence of a harmonic potential which keeps a tracers in a bounded region may be advantageous or not. On one hand, it leads to saturation of the mean time-averaged MSD and SRMS at long times. Here, the time-averaged SRMS benefits from this saturation by reducing the fluctuations, while the time-averaged MSD suffers from a shorter averaging window. For this reason, the time-averaged SRMS results in a more accurate statistical analysis at long times in the presence of a harmonic potential.

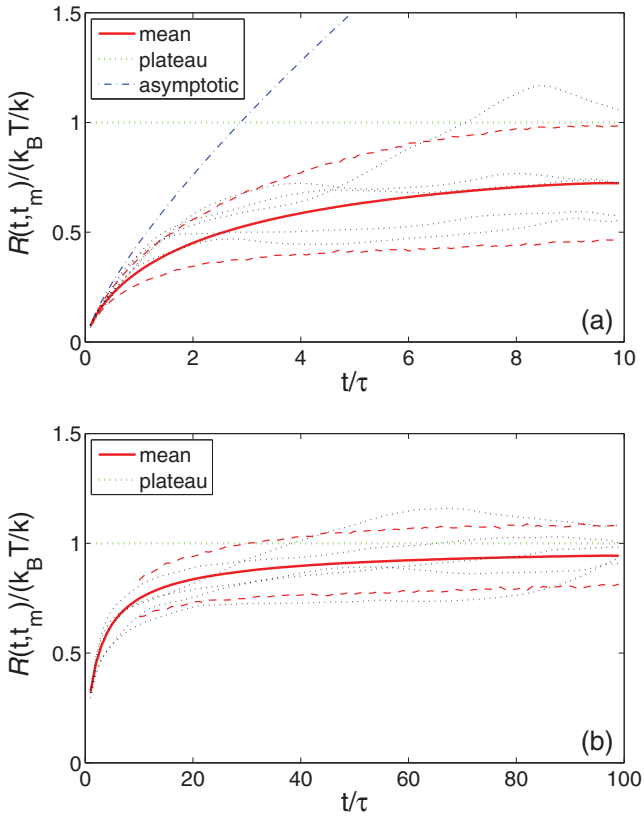

FIG. 4. (Color online) The normalized time-averaged SRMS for subdiffusion with $\alpha = 0.75$, $\gamma_\alpha = 1$, $k = 1$, $k_BT = 1$, $\tau = 1$, $t_m = 10$ (a), and $t_m = 100$ (b). The mean SRMS from Eq. (52) (solid red curve) is compared to its asymptotic short-time behavior (56) (dash-dotted blue curve) and to five empirical time-averaged SRMSs from simulated trajectories (dotted black curves). The dashed red curves show the mean SRMS plus or minus its standard deviation computed from Eq. (36). The expected plateau level 1 is shown by dotted green line.

On the other hand, the harmonic potential and the induced trapping of tracers affect the intrinsic dynamics (e.g., passive or active transport in living cells) and may shadow or alter certain long-time effects. From this point of view, one may wish to weaken the harmonic potential or, equivalently, to consider only the short-time behavior. In this situation, the time-averaged MSD and SRMS behave similarly, the latter showing slightly smaller fluctuations. Since the statistical uncertainty was shown to behave as $\sqrt{t/t_m}$ (for $\alpha = 1$), high rate acquisitions (e.g., by optical tweezers' techniques [43–45], which typically provide data samples on time scales from microseconds to seconds), may be preferred.

V. CONCLUSION

We dealt with the problem of characterizing a stochastic process from its single random realization. This is a generic problem for single-particle tracking techniques, which have become nowadays standard tools in microbiology. We focused on stochastic processes which can be described by a generalized Langevin equation with a frictional memory kernel, harmonic potential, and Gaussian thermal force satisfying the fluctuation-dissipation theorem. Our main results are the explicit formulas for the mean and covariance of two quadratic functionals of the process: the time-averaged MSD and SRMS. If the first functional is routinely used in processing and analysis of the data from SPT techniques, the second one has received little attention so far.

The explicit formulas for the means can be directly used for fitting empirical MSD and SRMS curves extracted from individual trajectories of tracers. As both functionals are still random variables, their empirical values can provide reasonable estimates of the means only if the related variances are small enough. In other words, the mean values are estimated within a statistical uncertainty which, in a first approximation, is defined as the ratio between the standard deviation and the mean. We showed that the statistical uncertainty of the time-averaged SRMS is smaller than that of the time-averaged MSD, in spite of much broader applications of the latter one. These theoretical results have been successfully confirmed by Monte Carlo simulations of the Langevin dynamics. We conclude that the use of the time-averaged SRMS would result in a more accurate statistical analysis of single trajectories and a more reliable interpretation of experimental data.

ACKNOWLEDGMENTS

Numerous fruitful discussions with E. Bertseva are gratefully acknowledged. The work has been partly supported by the ANR program DYOPTRI.

APPENDIX A: COVARIANCE OF MSD

In this appendix, we compute the covariance of the time-averaged MSD:

$$C_{\mathcal{M}}(t_1,t_2,t_m) \equiv \langle \mathcal{M}(t_1,t_m)\mathcal{M}(t_2,t_m)\rangle - \langle \mathcal{M}(t_1,t_m)\rangle\langle \mathcal{M}(t_2,t_m)\rangle.$$

Eqs. (10) yield

$$\langle \mathcal{M}(t_1,t_m)\mathcal{M}(t_2,t_m)\rangle = \frac{1}{(t_m-t_1)(t_m-t_2)} \times \int_0^{t_m-t_1} dt_{01} \int_0^{t_m-t_2} dt_{02} \times \langle (A_1+B_1)^2(A_2+B_2)^2\rangle,$$

where

$$A_i = X_0(t_{0i}+t_i) - X_0(t_{0i}) = U_0(t_i,t_{0i}),$$

$$B_i = \int_0^{t_{0i}+t_i} dt' G(t_{0i}+t_i-t')F(t') - \int_0^{t_{0i}} dt' G(t_{0i}-t')F(t'),$$

with the function U_0 defined by Eq. (12). The condition $\langle F(t)\rangle = 0$ implies

$$\begin{aligned}\left\langle (A_1+B_1)^2(A_2+B_2)^2\right\rangle &= \left(A_1^2+\langle B_1^2\rangle\right)\left(A_2^2+\langle B_2^2\rangle\right) \\ &\quad + 4A_1A_2\langle B_1B_2\rangle + \left\langle B_1^2B_2^2\right\rangle - \langle B_1^2\rangle\langle B_2^2\rangle. \end{aligned} \quad \text{(A1)}$$

The first term represents $\langle \mathcal{M}(t_1,t_m)\rangle\langle \mathcal{M}(t_2,t_m)\rangle$ and will be canceled from the covariance. The second term, which contains the initial conditions, involves A_i and

$$\langle B_1 B_2\rangle = k_B T U(t_1,t_2,t_{01},t_{02}),$$

with the function U defined by Eq. (14).

We focus now on the last two terms in Eq. (A1). Expanding the quadratic polynomials in $\langle B_1^2 B_2^2\rangle$, we have

$$\begin{aligned}\langle B_1^2 B_2^2\rangle = {} & I(t_{01}+t_1,t_{01}+t_1,t_{02}+t_2,t_{02}+t_2)\\ & -2I(t_{01}+t_1,t_{01}+t_1,t_{02}+t_2,t_{02})\\ & +I(t_{01}+t_1,t_{01}+t_1,t_{02},t_{02})\\ & -2I(t_{01}+t_1,t_{01},t_{02}+t_2,t_{02}+t_2)\\ & +4I(t_{01}+t_1,t_{01},t_{02}+t_2,t_{02})\\ & -2I(t_{01}+t_1,t_{01},t_{02},t_{02})\\ & +I(t_{01},t_{01},t_{02}+t_2,t_{02}+t_2)\\ & -2I(t_{01},t_{01},t_{02}+t_2,t_{02})+I(t_{01},t_{01},t_{02},t_{02}),\end{aligned}$$

where

$$\begin{aligned}I(t_1,t_2,t_3,t_4) \equiv {} & \int_0^{t_1} dt_1' G(t_1-t_1') \int_0^{t_2} dt_2' G(t_2-t_2')\\ & \times \int_0^{t_3} dt_3' G(t_3-t_3') \int_0^{t_4} dt_4' G(t_4-t_4')\\ & \times \langle F(t_1')F(t_2')F(t_3')F(t_4')\rangle.\end{aligned}$$

The four-point correlation function involving the Gaussian force $F(t)$ can be reduced to two-point correlation functions by using the Wick theorem:

$$\begin{aligned}\langle F(t_1')F(t_2')F(t_3')F(t_4')\rangle = {} & (k_B T)^2[\gamma(|t_1'-t_2'|)\gamma(|t_3'-t_4'|)\\ & +\gamma(|t_1'-t_3'|)\gamma(|t_2'-t_4'|)\\ & +\gamma(|t_1'-t_4'|)\gamma(|t_2'-t_3'|)],\end{aligned} \quad \text{(A2)}$$

from which

$$\begin{aligned}I(t_1,t_2,t_3,t_4) = {} & (k_B T)^2[K(t_1,t_2)K(t_3,t_4)\\ & +K(t_1,t_3)K(t_2,t_4)+K(t_1,t_4)K(t_2,t_3)].\end{aligned}$$

After simplifications, we get

$$\left\langle B_1^2 B_2^2\right\rangle = 2[U(t_1,t_2,t_{01},t_{02})]^2 + U(t_1,t_1,t_{01},t_{01})U(t_2,t_2,t_{02},t_{02}).$$

The second term is equal to $\langle B_1^2\rangle\langle B_2^2\rangle$ and will be canceled from the covariance. In turn, the first term contributes to the covariance of the time-averaged MSD yielding Eq. (19).

1. Normal diffusion

For normal diffusion ($\alpha = 1$), one has

$$\begin{aligned}U(t,t,t_{01},t_{02}) = {} & e^{-(t_{01}+t_{02})/\tau}(1-e^{-t/\tau})^2 - 2e^{-|t_{01}-t_{02}|/\tau}\\ & +e^{-|t_{01}+t-t_{02}|/\tau}+e^{-|t_{02}+t-t_{01}|/\tau}.\end{aligned}$$

Considering separately the cases $t < t_m/2$ and $t > t_m/2$, one obtains the explicit formula for the variance. Instead of providing this formula, we consider several limiting cases:

(i) In the limit $t \to 0$, one obtains

$$\mathrm{var}\{\mathcal{M}(t,t_m)\} \approx \frac{2(k_B T)^2}{\gamma_1^2}\frac{8t^3}{3t_m} + O(t^4).$$

In this limit, $\langle \mathcal{M}(t,t_m)\rangle \approx 2(k_B T/\gamma_1)t$, from which one gets Eq. (59).

(ii) When $\tau \ll t \ll t_m$, one has

$$\mathrm{var}\{\mathcal{M}(t,t_m)\} \approx \frac{2(k_B T)^2}{k^2}\frac{6}{t_m},$$

while $\langle \mathcal{M}(t,t_m)\rangle \approx 2k_B T/k$ so that one gets Eq. (60).

(iii) In the limit $t \to t_m$, one gets

$$\langle \mathcal{M}(t_m,t_m)\rangle = \frac{k_B T}{k}(1-e^{-2t_m/\tau}),$$

so that for large $t_m \gg \tau$, the mean MSD value is approximately twice smaller than the expected limit $2k_B T/k$. For the variance, one gets

$$\mathrm{var}\{\mathcal{M}(t_m,t_m)\} = \frac{2(k_B T)^2}{k^2}(1-e^{-2t_m/\tau})^2,$$

from which the ratio is

$$\frac{\mathrm{std}\{\mathcal{M}(t_m,t_m)\}}{\langle \mathcal{M}(t_m,t_m)\rangle} = \sqrt{2} \approx 1.414\ldots,$$

independently of t_m. This means that for t close to t_m, the standard deviation is larger than the mean value.

(iv) For free normal diffusion ($k = 0$), one finds explicitly

$$\begin{aligned}\mathrm{var}\{\mathcal{M}(t,t_m)\} = {} & \frac{4(k_B T)^2}{3\gamma_1^2(t_m-t)^2}\\ & \times [t^3(4t_m-5t)+\left(\min\{t_m-2t,0\}\right)^4].\end{aligned}$$

The variance is a monotonously increasing function of t.

APPENDIX B: COVARIANCE OF SRMS

The definition (2) of the time-averaged SRMS yields

$$\begin{aligned}\langle \mathcal{R}(t_1,t_m)\mathcal{R}(t_2,t_m)\rangle = {} & \frac{1}{(t_m-t_1)(t_m-t_2)}\int_0^{t_m-t_1} dt_{01}\\ & \times \int_0^{t_m-t_2} dt_{02}\frac{1}{t_1}\int_0^{t_1} dt_1'\frac{1}{t_2}\int_0^{t_2} dt_2'\\ & \times \langle (A_1+B_1)^2(A_2+B_2)^2\rangle,\end{aligned}$$

where functions $A_{1,2}$ and $B_{1,2}$ denote different terms in Eq. (21):

$$A_i = X_0(t_i'+t_{0i}) - \frac{1}{t_i'}\int_0^{t_i'} dt'' X_0(t''+t_{0i}) = \frac{1}{t_i'}J_0(t_i',t_{0i}),$$

$$B_i = \frac{1}{t_i'}\int_0^{t_i'} dt'' t'' \int_0^{t''+t_{0i}} dt''' g(t''+t_{0i}-t''')F(t'''),$$

with the function J_0 defined previously by Eq. (25). We can apply again Eq. (A1) to analyze $\langle (A_1+B_1)^2(A_2+B_2)^2\rangle$ term by term. The first term in Eq. (A1) now represents $\langle r(t_1,t_{01})\rangle\langle r(t_2,t_{02})\rangle$ and will be canceled from the covariance. The second term, which contains the initial conditions, is

explicitly written in terms of

$$\langle B_1 B_2\rangle = \frac{k_B T}{t_1' t_2'} J(t_1',t_2',t_{01},t_{02}),$$

where the function $J(t_1',t_2',t_{01},t_{02})$ is defined by Eq. (30).

We focus now on the last two terms in Eq. (A1) which provide the main contribution to the covariance. Using the Wick theorem (A2), we get

$$\left\langle B_1^2 B_2^2\right\rangle = \frac{(k_B T)^2}{(t_1' t_2')^2}[J(t_1',t_1',t_{01},t_{01})J(t_2',t_2',t_{02},t_{02}) + 2J(t_1',t_2',t_{01},t_{02})^2].$$

The first term in the above equation also appears in $\langle B_1^2\rangle\langle B_2^2\rangle$ that cancels this contribution. The remaining term gives the covariance in Eq. (35).

Although the asymptotic formulas for the covariance would be helpful to avoid time-consuming numerical computation of the quadruple integral, this analysis is beyond the scope of the paper. We consider only the limit $t \to t_{\rm m}$, when the integration variables t_{01} and t_{02} are small so that the function $J(t_1,t_2,t_{01},t_{02})$ can be approximated as $J(t_1,t_2,0,0)$, from which

$$\mathrm{var}\{\mathcal{R}(t_{\rm m},t_{\rm m})\} = \frac{2(k_B T)^2}{t_{\rm m}^2}\int_0^{t_{\rm m}}\frac{dt_1}{t_1^2}\int_0^{t_{\rm m}}\frac{dt_2}{t_2^2}[J(t_1,t_2,0,0)]^2$$

[this is also the variance of $r(t_{\rm m},0)$]. For normal diffusion, the double integral can be computed analytically:

$$\mathrm{var}\{\mathcal{R}(t_{\rm m},t_{\rm m})\} = \frac{2(k_B T)^2}{k^2} V(t_{\rm m}/\tau),$$

where

$$V(z) = \frac{1}{4z^4}[(4z^3 - 21z^2 + 28z + 36) - 16(2z^2 + 3z + 6)e^{-z} + 4(2z^3 + 5z^2 + 8z + 22)e^{-2z} - 16(z+2)e^{-3z} + (z^2 + 4z + 4)e^{-4z}].$$

For large z, this function decays as $1/z$. Since the mean value approaches a constant at long times, one gets the asymptotic formula (62) for the statistical uncertainty of the time-averaged SRMS at long times. Although this result was derived for normal diffusion, it remains qualitatively valid for subdiffusion.

a. Numerical computation

Although the explicit formula (30) was derived for the function $J(t_1,t_2,t_{01},t_{02})$, a standard finite sum approximation of the quadruple integral in Eq. (36) is too time consuming, especially for long sample duration $t_{\rm m}$. A Monte Carlo integration is preferred for a faster numerical evaluation. Rescaling the integration variables, one can approximate the variance as

$$\begin{aligned}\mathrm{var}\{\mathcal{R}(t,t_{\rm m})\} &= \frac{2(k_B T)^2}{t^4}\int_0^1\frac{dx_1}{x_1^2}\int_0^1\frac{dx_2}{x_2^2}\int_0^1 dx_3\int_0^1 dx_4 \\ &\quad\times[J(tx_1,tx_2,(t_{\rm m}-t)x_3,(t_{\rm m}-t)x_4)]^2 \\ &\approx \frac{2(k_B T)^2}{t^4 N}\sum_{n=1}^{N}\left[\frac{J\left(tx_1^{(n)},tx_2^{(n)},(t_{\rm m}-t)x_3^{(n)},(t_{\rm m}-t)x_4^{(n)}\right)}{x_1^{(n)}x_2^{(n)}}\right]^2,\end{aligned}$$

where $\{x_1^{(n)},x_2^{(n)},x_3^{(n)},x_4^{(n)}\}_{n=1}^N$ is a set of independent random variables with a uniform distribution on the unit interval. The accuracy of this approximation increases with N.

b. Free normal diffusion

For the special case of free normal diffusion ($k = m = 0$, $\alpha = 1$), the integrals in Eq. (36) can be computed analytically. Skipping technical details, we give the final result for the variance of the time-averaged SRMS:

$$\mathrm{var}\{\mathcal{R}(t,t_{\rm m})\} = \frac{(k_B T)^2}{\gamma_1^2}\begin{cases}\dfrac{t^3(40t_{\rm m}-49t)}{630(t_{\rm m}-t)^2} & (t < t_{\rm m}/2), \\ \dfrac{5(t_{\rm m}-t)^6 - 16(t_{\rm m}-t)^5 t + 56(t_{\rm m}-t)^3 t^3 - 70(t_{\rm m}-t)^2 t^4 + 56t^6}{630t^4} & (t > t_{\rm m}/2).\end{cases} \tag{B1}$$

In the limit $t \to 0$, one has

$$\mathrm{var}\{\mathcal{R}(t,t_{\rm m})\} \approx \frac{4}{63}\frac{(k_B T)^2}{\gamma_1^2}\frac{t^3}{t_{\rm m}}.$$

In the opposite limit $t \to t_{\rm m}$, one finds

$$\mathrm{var}\{\mathcal{R}(t_{\rm m},t_{\rm m})\} = \frac{4}{45}\frac{(k_B T)^2 t_{\rm m}^2}{\gamma_1^2},$$

from which

$$\frac{\mathrm{std}\{\mathcal{R}(t_{\rm m},t_{\rm m})\}}{\langle\mathcal{R}(t_{\rm m},t_{\rm m})\rangle} \approx \frac{2}{\sqrt{5}} \approx 0.8944,$$

independently of $t_{\rm m}$. For comparison, similar result for the time-averaged MSD was $\sqrt{2}$.

APPENDIX C: SUBDIFFUSION WITHOUT HARMONIC POTENTIAL

When there is no harmonic potential ($k = 0$), the general formula (38) is simplified, yielding Eq. (41) for the kernels $G^{(n)}(t)$. This limiting case can be studied as thoroughly as we did for subdiffusion of massless tracers in harmonic potential. Since derivations are similar to the previous ones, we only formulate the main results.

1. Time-averaged MSD

Substitution of Eqs. (41) into Eqs. (11) and (17) yields

$$\langle \mathcal{M}(t,t_{\rm m})\rangle = \bar{M}(t) - \left(\frac{k_BT}{m} - v_0^2\right)\bar{\tau}^2 \bar{M}_\alpha(t/\bar{\tau}, t_{\rm m}/\bar{\tau}),$$

where $\bar{\tau} = (m/\gamma_\alpha)^{1/(2-\alpha)}$ is the time scale $\tau_{\rm in,fr}$,

$$\bar{M}(t) = \frac{2k_BT}{m} t^2 E_{2-\alpha,3}(-\gamma_\alpha t^{2-\alpha}/m),$$

and

$$\bar{M}_\alpha(z,\zeta) = \frac{1}{\zeta - z}\int_0^{\zeta - z} dx[(x+z)E_{2-\alpha,2}(-(x+z)^{2-\alpha}) - xE_{2-\alpha,2}(-x^{2-\alpha})]^2.$$

The correction $\bar{M}_\alpha(z,\zeta)$ accounts for a finite length $t_{\rm m}$ of the data sample, and vanishes in the limit $t_{\rm m}\to\infty$. The term $\bar{M}(t)$ provides the main contribution and gives the leading terms in the short-time and long-time asymptotic behaviors of the time-averaged MSD:

$$\langle \mathcal{M}(t,t_{\rm m})\rangle \simeq \frac{k_BT}{m}\left[t^2 - \frac{2\gamma_\alpha t^{4-\alpha}}{m\Gamma(5-\alpha)} + \cdots\right] \quad (t \ll \bar{\tau}),$$

$$\langle \mathcal{M}(t,t_{\rm m})\rangle \simeq \frac{2k_BT}{\gamma_\alpha}\frac{t^\alpha}{\Gamma(\alpha+1)} + \cdots \quad (t \gg \bar{\tau}),$$

so that we retrieved the results from [38]. Naturally, the long-time asymptotic behavior for massy tracers without harmonic potential coincides with the short-time asymptotic behavior (51) for massless tracers within harmonic potential.

For normal diffusion ($\alpha = 1$), one gets the explicit solution:

$$\bar{M}(t) = \frac{2k_BT}{m}\bar{\tau}^2(e^{-t/\bar{\tau}} - 1 + t/\bar{\tau}),$$

while $\bar{M}_1(z,\zeta) = M_1(z,\zeta)$, which is given by Eq. (49).

2. Time-averaged SRMS

Substitution of Eqs. (41) into Eqs. (34) yields

$$\langle \mathcal{R}(t,t_{\rm m})\rangle = \bar{R}(t) - \left(\frac{k_BT}{m} - v_0^2\right)\bar{R}_\alpha(t/\bar{\tau}, t_{\rm m}/\bar{\tau}),$$

where

$$\bar{R}(t) = \frac{2k_BT}{m} t^2 E_{2-\alpha,5}(-\gamma_\alpha t^{2-\alpha}/m)$$

and

$$\bar{R}_\alpha(z,\zeta) = \frac{1}{z}\int_{\zeta - z}^{\zeta} dx x^2[E_{2-\alpha,2}(-x^{2-\alpha})]^2 - \frac{1}{z^2(\zeta - z)}\int_0^{\zeta - z} dx[((x+z)^2 E_{2-\alpha,3}[-(x+z)^{2-\alpha}] - x^2 E_{2-\alpha,3}(-x^{2-\alpha}))^2 + xz((x+z)^2\{E_{2-\alpha,2}[-(x+z)^{2-\alpha}]\}^2 - x^2[E_{2-\alpha,2}(-x^{2-\alpha})]^2)]. \quad \text{(C1)}$$

For normal diffusion, the substitution of Eq. (57) and direct integration yield $\bar{R}_1(z,\zeta) = R_1(z,\zeta)$ which is given by Eq. (58). In turn, the main contribution is

$$\bar{R}(t) = \frac{2k_BTt^2}{m}\frac{e^{-t/\bar{\tau}} - 1 + t/\bar{\tau} - \frac{1}{2}(t/\bar{\tau})^2 + \frac{1}{6}(t/\bar{\tau})^3}{(t/\bar{\tau})^4}.$$

APPENDIX D: ASYMPTOTIC ANALYSIS OF CORRECTION TERMS

We investigate the asymptotic properties of the correction terms. Throughout this section, we assume that $\zeta \gg 1$ (i.e., $t_{\rm m} \gg \tau$).

1. Auxiliary integral

We start by studying the asymptotic behavior of the integral,

$$\hat{r}_\alpha(z) \equiv \int_0^z dx[E_{\alpha,1}(-x^\alpha)]^2,$$

which will be used in the next subsections.

For small z, the series representation (39) of the Mittag-Leffler function and term-by-term integration yield

$$\hat{r}_\alpha(z) = \sum_{n=0}^\infty \frac{(-1)^n a_n}{1+\alpha n} z^{1+n\alpha},$$

with the coefficients a_n of the Taylor series of $[E_{\alpha,1}(-x)]^2$,

$$a_n = \sum_{k=0}^n \frac{1}{\Gamma(\alpha k+1)\Gamma[\alpha(n-k)+1]}.$$

Although the above series representation converges for any z, its practical use for numerical computation is limited to relatively small z. In particular, one has

$$\hat{r}_\alpha(z) \simeq z\left(1 - \frac{2z^\alpha}{\Gamma(\alpha+2)} + z^{2\alpha}\left[\frac{2}{\Gamma(2\alpha+2)} + \frac{1}{[\Gamma(\alpha+1)]^2(1+2\alpha)}\right] + O(z^{3\alpha})\right). \quad \text{(D1)}$$

For large z, one can choose a constant z_0 such that $1 \ll z_0 \ll z$ in order to split the integral into two parts,

$$\hat{r}_\alpha(z) = \int_0^{z_0} dx[E_{\alpha,1}(-x^\alpha)]^2 + \int_{z_0}^z dx[E_{\alpha,1}(-x^\alpha)]^2.$$

The first term is a constant, while for the second term the asymptotic series (44) of the Mittag-Leffler function can be used. Multiplying this series by itself, one gets an N-order approximation

$$[E_{\alpha,\beta}(-x)]^2 \simeq \sum_{n=2}^{N} \frac{(-1)^n b_n}{x^n} \qquad (x \to \infty),$$

where

$$b_n = \sum_{k=1}^{n-1} \frac{1}{\Gamma(\beta - \alpha k)\Gamma[\beta - \alpha(n-k)]}.$$

The second term has then the following approximation:

$$\int_{z_0}^{z} dx [E_{\alpha,1}(-x^\alpha)]^2 \simeq \sum_{n=2}^{N} \frac{(-1)^n b_n}{1-\alpha n}\left(z^{1-n\alpha} - z_0^{1-n\alpha}\right).$$

Bringing two terms together, one obtains

$$\hat{r}_\alpha(z) \simeq \left[\int_0^{z_0} dx [E_{\alpha,1}(-x^\alpha)]^2 - \sum_{n=2}^{N} \frac{(-1)^n b_n}{1-\alpha n} z_0^{1-\alpha n}\right] + \sum_{n=2}^{N} \frac{(-1)^n b_n}{1-\alpha n} z^{1-\alpha n} \qquad (z \gg 1). \tag{D2}$$

When $1/2 < \alpha < 1$, the first (constant) term in brackets provides the main contribution, as compared to the last term of the order $z^{1-2\alpha}$, which vanishes in the limit $z \to \infty$. In particular, the integral $\hat{r}_\alpha(\infty)$ converges for $1/2 < \alpha < 1$. At $\alpha = 1/2$, both sums in Eq. (D2) contain an infinite term (for $n = 2$), which comes from the incorrect integration of $1/x$. The logarithmic dependence on z takes place in this case. Similar problems appear for any $\alpha = 1/n$ with $n = 1,2,3,\ldots$. Although accurate formulas can be derived for any $0 < \alpha < 1$, we focus on the range $1/2 < \alpha < 1$.

2. Short-time asymptotics for time-averaged MSD

When $z \ll 1$, the integrand function in Eq. (48) can be decomposed into a Taylor series

$$M_\alpha(z,\zeta) \simeq \frac{1}{\zeta - z}\int_0^{\zeta - z} dx [z f'(x)]^2 + O(z^3),$$

where $f(x) = E_{\alpha,1}(-x^\alpha)$, from which one gets Eq. (50).

3. Long-time asymptotics for time-averaged MSD

In the opposite limit $z \to \zeta$, the term $E_{\alpha,1}[-(x+z)^\alpha] \sim z^{-\alpha}$ can be neglected in comparison to $E_{\alpha,1}(-x^\alpha) \sim 1$ so that

$$M_\alpha(z,\zeta) \simeq \frac{\hat{r}_\alpha(\zeta - z)}{\zeta - z} + O(\zeta^{-\alpha}).$$

Using Eq. (D1), one gets

$$M_\alpha(z,\zeta) \simeq 1 - \frac{2(\zeta - z)^\alpha}{\Gamma(\alpha+2)} + O(\zeta^{-\alpha}) \qquad (z \to \zeta). \tag{D3}$$

In this limit, the "correction" $M_\alpha(z,\zeta)$ is of the same order as the "main" contribution $M(t)$.

4. Short-time asymptotics for time-average SRMS

In the short-time limit ($z \ll 1$), the correction term from Eq. (54) for the time-averaged SRMS can be written as

$$R_\alpha(z,\zeta) = \frac{1}{z}\int_{\zeta - z}^{\zeta} dx h(x) - \frac{1}{z^2(\zeta - z)} \times \int_0^{\zeta - z} dx \{[f(x+z) - f(x)]^2 + xz[h(x+z) - h(x)]\}, \tag{D4}$$

where $f(x) = xE_{\alpha,2}(-x^\alpha)$ and $h(x) = [E_{\alpha,1}(-x^\alpha)]^2$. The integrand function in the second term can be decomposed into a Taylor series up to the fourth order of z. The identity $f'(x) = E_{\alpha,1}(-x^\alpha)$ cancels many terms, from which the second term of Eq. (D4) is

$$h(\zeta) - \frac{z}{2}h'(\zeta) + \frac{z^2}{6}h''(\zeta) - \frac{z^2}{12\zeta}\int_0^{\zeta} dx [f''(x)]^2 + O(z^3).$$

Denoting $H(x)$ the primitive of the function $h(x)$ and expanding $H(\zeta) - H(\zeta - z)$ into a Taylor series, one writes the first integral in Eq. (D4) as

$$h(\zeta) - \frac{z}{2}h'(\zeta) + \frac{z^2}{6}h''(\zeta) + O(z^3).$$

Combining these two results, one obtains

$$R_\alpha(z,\zeta) \simeq \frac{z^2}{12\zeta}\int_0^{\zeta} dx [f''(x)]^2 + O(z^3).$$

The substitution of the function $f(x)$ yields Eq. (55).

The same analysis is applicable for the case of massy particles without harmonic potential. The correction term is given by Eq. (C1) which admits the representation (D4) with $f(x) = x^2 E_{2-\alpha,3}(-x^{2-\alpha})$ and $h(x) = [xE_{2-\alpha,2}(-x^{2-\alpha})]^2$. These functions also satisfy the relations $f'(x) = xE_{2-\alpha,2}(-x^{2-\alpha})$, $[f'(x)]^2 = h(x)$ and $h'(x) = 2f'(x)f''(x)$ so that

$$\bar{R}_\alpha(z,\zeta) \simeq \frac{z^2}{12\zeta}\int_0^{\zeta} dx [E_{2-\alpha,1}(-x^{2-\alpha})]^2 + O(z^3).$$

5. Long-time asymptotics for time-average SRMS

In the opposite limit $z \to \zeta$, the integrand function in the second term of Eq. (54) can be expanded into a series for small x in order to get the largest correction,

$$\lim_{z\to\zeta} R_\alpha(z,\zeta) = r_\alpha(\zeta),$$

where

$$r_\alpha(\zeta) = -[E_{\alpha,2}(-\zeta^\alpha)]^2 + \frac{\hat{r}_\alpha(\zeta)}{\zeta}.$$

Using the asymptotic behavior (D2) of the integral $\hat{r}_\alpha(\zeta)$, one concludes that

$$r_\alpha(\zeta) \sim \zeta^{-1} + O(\zeta^{-2\alpha}) \qquad (\zeta \gg 1), \tag{D5}$$

when $1/2 < \alpha < 1$.

[1] I. M. Tolić-Norrelykke, E.-L. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sorensen, Phys. Rev. Lett. **93**, 078102 (2004).

[2] I. Golding and E. C. Cox, Phys. Rev. Lett. **96**, 098102 (2006).

[3] D. Arcizet, B. Meier, E. Sackmann, J. O. Rädler, and D. Heinrich, Phys. Rev. Lett. **101**, 248103 (2008).

[4] C. Wilhelm, Phys. Rev. Lett. **101**, 028101 (2008).

[5] D. Wirtz, Ann. Rev. Biophys. **38**, 301 (2009).

[6] R. M. L. Evans, M. Tassieri, D. Auhl, and T. A. Waigh, Phys. Rev. E **80**, 012501 (2009).

[7] H. Lee, J. M. Ferrer, F. Nakamura, M. J. Lang, and R. D. Kamm, Acta Biomaterialia **6**, 1207 (2010).

[8] C. Guzman, Appl. Phys. Lett. **93**, 184102 (2008).

[9] M. J. Saxton, Biophys. J. **81**, 2226 (2001).

[10] M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, Biophys. J. **87**, 3518 (2004).

[11] D. S. Grebenkov, Rev. Mod. Phys. **79**, 1077 (2007).

[12] H. Qian, M. P. Sheetz, and E. L. Elson, Biophys. J. **60**, 910 (1991).

[13] M. J. Saxton, Biophys. J. **64**, 1766 (1993).

[14] M. J. Saxton, Biophys. J. **72**, 1744 (1997).

[15] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).

[16] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).

[17] G. M. Zaslavsky, Phys. Rep. **371**, 461 (2002).

[18] R. Metzler and J. Klafter, J. Phys. A **37**, R161 (2004).

[19] M. A. Lomholt, I. M. Zaid, and R. Metzler, Phys. Rev. Lett. **98**, 200603 (2007).

[20] R. Metzler, V. Tejedor, J.-H. Jeon, Y. He, W. H. Deng, S. Burov, and E. Barkai, Acta Phys. Pol. B **40**, 1315 (2009).

[21] J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sorensen, Lene Oddershede, and R. Metzler, Phys. Rev. Lett. **106**, 048103 (2011).

[22] X. Michalet, Phys. Rev. E **82**, 041914 (2010).

[23] T. Savin and P. S. Doyle, Phys. Rev. E **76**, 021501 (2007).

[24] A. J. Berglund, Phys. Rev. E **82**, 011917 (2010).

[25] V. Tejedor, O. Bénichou, R. Voituriez, R. Jungmann, F. Simmel, C. Selhuber-Unkel, L. B. Oddershede, and R. Metzler, Biophys. J. **98**, 1364 (2010).

[26] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II. Nonequilibrium Statistical Mechanics* (Springer-Verlag, Berlin, Heidelberg, 1985).

[27] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).

[28] W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron, *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*, 2nd ed. (World Scientific, Singapore, 2004).

[29] K.-G. Wang and C. W. Lung, Phys. Lett. A **151**, 119 (1990).

[30] K. G. Wang, Phys. Rev. A **45**, 833 (1992).

[31] J. M. Porrà, K.-G. Wang, and J. Masoliver, Phys. Rev. E **53**, 5872 (1996).

[32] K. G. Wang and M. Tokuyama, Physica A **265**, 341 (1999).

[33] E. Lutz, Phys. Rev. E **64**, 051106 (2001).

[34] N. Pottier, Physica A **317**, 371 (2003).

[35] A. D. Vinales and M. A. Desposito, Phys. Rev. E **73**, 016111 (2006).

[36] A. D. Vinales and M. A. Desposito, Phys. Rev. E **75**, 042102 (2007).

[37] A. D. Vinales, K. G. Wang, and M. A. Desposito, Phys. Rev. E **80**, 011101 (2009).

[38] M. A. Desposito and A. D. Vinales, Phys. Rev. E **80**, 021111 (2009).

[39] I. Podlubny, *Fractional Differential Equations* (Academic, London, 1999).

[40] R. B. Davies and D. S. Harte, Biometrika **74**, 95 (1987).

[41] A. T. A. Wood and G. Chan, J. Comput. Graph. Stat. **3**, 409 (1994).

[42] C. R. Dietrich and G. N. Newsam, SIAM J. Sci. Comput. **18**, 1088 (1997).

[43] K. C. Neuman and S. M. Block, Rev. Sci. Instrum. **75**, 2787 (2004).

[44] A. Rohrbach *et al.*, Rev. Sci. Instrum. **75**, 2197 (2004).

[45] A. Yao, M. Tassieri, M. Padgett, and J. Cooper, Lab Chip **9**, 2568 (2009).