Power spectrum analysis with least-squares fitting: Amplitude bias and its elimination, with application to optical tweezers and atomic force microscope cantilevers

Simon F. Nørrelykke and Henrik Flyvbjerg

Power spectrum analysis with least-squares fitting: Amplitude bias and its elimination, with application to optical tweezers and atomic force microscope cantilevers
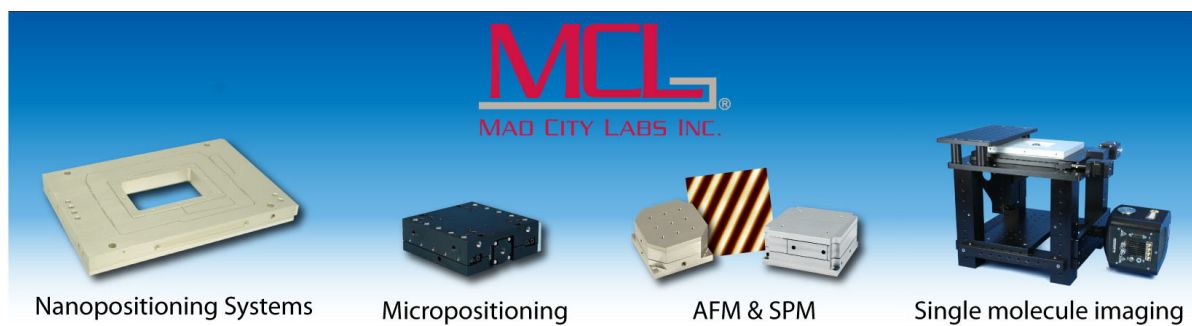
Simon F. Nørrelykke[1] and Henrik Flyvbjerg[2]

[1]*Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany and Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA*
[2]*Department of Micro- and Nanotechnology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark*

Optical tweezers and atomic force microscope (AFM) cantilevers are often calibrated by fitting their experimental power spectra of Brownian motion. We demonstrate here that if this is done with typical weighted least-squares methods, the result is a bias of relative size between $-2/n$ and $+1/n$ on the value of the fitted diffusion coefficient. Here, n is the number of power spectra averaged over, so typical calibrations contain 10%–20% bias. Both the sign and the size of the bias depend on the weighting scheme applied. Hence, so do length-scale calibrations based on the diffusion coefficient. The fitted value for the characteristic frequency is not affected by this bias. For the AFM then, force measurements are not affected provided an independent length-scale calibration is available. For optical tweezers there is no such luck, since the spring constant is found as the ratio of the characteristic frequency and the diffusion coefficient. We give analytical results for the weight-dependent bias for the wide class of systems whose dynamics is described by a linear (integro)differential equation with additive noise, white or colored. Examples are optical tweezers with hydrodynamic self-interaction and aliasing, calibration of Ornstein–Uhlenbeck models in finance, models for cell migration in biology, etc. Because the bias takes the form of a simple multiplicative factor on the fitted amplitude (e.g. the diffusion coefficient), it is straightforward to remove and the user will need minimal modifications to his or her favorite least-squares fitting programs. Results are demonstrated and illustrated using synthetic data, so we can compare fits with known true values. We also fit some commonly occurring power spectra *once-and-for-all* in the sense that we give their parameter values and associated error bars as explicit functions of experimental power-spectral values. © *2010 American Institute of Physics.*
[doi:10.1063/1.3455217]

I. INTRODUCTION

Optical tweezers (OTs) are often calibrated by fitting the power spectrum of a trapped bead's Brownian motion with a Lorentzian.[1–5] Similarly, atomic force microscopes (AFMs) are sometimes calibrated by fitting the power spectrum of the cantilever's Brownian motion with the power spectrum of a damped harmonic oscillator.[6–8] These fits are routinely done by least-squares (LSQ) minimization, the premises of which are rarely satisfied in practice. Here we do maximum likelihood estimation (MLE) of parameters, the premises of which *are* satisfied and give the parameter values, with error bars, as explicit functions of the experimental power spectrum in an excellent approximation. These closed-formula results should be useful for online calibration, since that requires high-speed determination of parameters. They also demonstrate that typical calibrations using least-squares fitting contain 10%–20% systematic errors, also known as *bias*.

We find quite generally that both the sign and the size of the bias depend on the details of how a least-squares fit is implemented through the choice of how the data points are weighted. Analytical expressions for the bias are given and

examples of the behavior of the stochastic fit errors are examined numerically.

These results apply beyond the examples given here, since the same bias occurs in all systems described by a linear (integro)differential equation with additive noise. Thus, using the correction factors given in Eqs. (32) and (41), it is possible to obtain correct unbiased estimates for the fit parameters without performing a computationally expensive MLE. Instead, one simply adjusts the results of an ordinary least-squares (OLS) fit.

We also show that the bias found for least-squares fits appears under fairly general conditions, independent of the details of the function that is fitted, but depending only on the distribution of the error on the data. Analytical correction factors are given for the examples of Gaussian, exponential, and gamma distributed errors. Finally, a general criterion is given for when one may expect a least-squares fit to be biased: when the experimental mean is correlated with the experimental variance. Conversely, if they are uncorrelated the fit is unbiased. The paper is organized as a main text followed by six appendixes that contain proofs and other technical details.

II. DYNAMICS

The equation of motion for a massive particle moving in a harmonic potential under the influence of thermal forces is

$$m\ddot{x}(t) + \gamma\dot{x}(t) + \kappa x(t) = F_{\text{therm}}(t). \quad (1)$$

Here $x(t)$ is the coordinate of the particle as function of time t, m is its inertial mass, γ is its friction coefficient, κ is Hooke's constant, and F_{therm} is the thermal force on the particle. This force is random and assumed to have white-noise statistical properties,

$$\langle F_{\text{therm}}(t)\rangle = 0,$$
$$\langle F_{\text{therm}}(t)F_{\text{therm}}(t')\rangle = 2k_{\text{B}}T\gamma\delta(t-t'), \quad \text{for all } t,t', \quad (2)$$

where δ is Dirac's delta function, $k_{\text{B}}T$ is the Boltzmann energy, and $\langle\cdots\rangle$ is the expectation value with respect to the noise. This theory's power spectrum of thermal motion is derived in Ref. 9.

III. POWER SPECTRA

OTs should be calibrated using the hydrodynamically correct power spectrum given in Refs. 9 and 10, possibly taking into account a number of effects listed there and further detailed in Refs. 11–13. These effects include: (i) the frequency dependence of the friction coefficient due to hydrodynamic self-coupling; (ii) dependence of the friction coefficient on distance to nearby surfaces due to hydrodynamic coupling; (iii) extra $1/f$ power at low frequencies, caused by the laser pointing stability, hydrodynamic self-coupling, etc.; and (iv) optical interference effects caused by a standing wave between the trapped object and the nearby microscope cover-slip surface, when determining the displacement sensitivity (volt to nanometer conversion factor). Generally, the most crucial step, where the largest systematic errors are likely to appear, is the determination of the displacement sensitivity.[10] However, as also described in Ref. 9, there are situations in which an acceptable approximation is achieved by fitting a Lorentzian

$$P_f = \frac{D/(2\pi^2)}{f_c^2 + f^2} \quad (3)$$

to an average $\overline{P}_f^{(\text{ex})}$ of, say, n two-sided experimental power spectra for the Brownian motion of a trapped microsphere [occasionally, confusion arises over extra or missing factors of two in power spectral densities (PSDs)—this is down to the use of one-sided $f > 0$ versus two-sided $f \gtrless 0$ PSDs; as long as the total power is contained in the chosen frequency range, the two approaches are equally correct]. The notation used is essentially the same as in Ref. 9, where the aliased Lorentzian is also derived, i.e., $f_c = \kappa/(2\pi\gamma)$ and $D = k_{\text{B}}T/\gamma$.

Similarly, AFM cantilevers are sometimes calibrated by fitting a frequency interval around the resonance peak[8] using

$$P_f = \frac{D/(2\pi^2)}{\left(\dfrac{2\pi m}{\gamma}\right)^2 (f_0^2 - f^2)^2 + f^2} \quad (4)$$

$$= \frac{D/(2\pi^2)}{\left(\dfrac{Q}{f_0}\right)^2 (f_0^2 - f^2)^2 + f^2}, \quad (5)$$

to a similar average of experimental power spectra for a cantilever's Brownian motion. Here, the characteristic frequency $f_0^2 = \kappa/(4\pi^2 m)$ and the quality factor $Q = \sqrt{m\kappa}/\gamma$. Describing the AFM cantilever as a simple harmonic oscillator is rigorously correct for high quality factors $Q \gg 1$, e.g., in air where dissipative forces are small.[8] In water, this description also works, at least for short stiff cantilevers, but now the drag coefficient is frequency dependent.[7] It is not our aim here to review the expansive literature on AFM calibration—the only point we seek to make is that if the amplitude of the power spectrum is involved, systematic fitting errors are typically present as shown below.

Both these theoretical power spectra follow from the Einstein–Ornstein–Uhlenbeck theory for the Brownian motion of a damped harmonic oscillator in one dimension; see Eq. (1) and Ref. 9, where the parameters in Eqs. (3) and (4) are also defined. In Appendix B we give the results for the aliased AFM PSD.

Least-squares fits of power spectra are biased, as shown below in Sec. V, irrespective of whether one applies the above simplified theory or a more complete one that takes into account aliasing, hydrodynamics, electronic filters, etc. This bias will be on the amplitude *only* and will not affect shape parameters, i.e., D will be biased, whereas f_c, f_0, and Q will not.

IV. STATISTICAL PROPERTIES OF EXPERIMENTAL POWER SPECTRUM

Here, and throughout the rest of the paper, we differentiate between theoretical *expectation* values, indicated by brackets $\langle\cdots\rangle$, and experimental *averages*, indicated by a bar $\overline{\cdot}$. Following the notation of Ref. 9, we write the thermal force in Eq. (2) in terms of a normalized white-noise process $\eta(t)$,

$$F_{\text{therm}}(t) = \sqrt{2k_{\text{B}}T\gamma}\,\eta(t), \quad (6)$$

whose Fourier transform obeys

$$\langle\tilde{\eta}_k\rangle = 0; \langle\tilde{\eta}_k^*\tilde{\eta}_\ell\rangle = t_{\text{msr}}\delta_{k,\ell}, \quad (7)$$

where $k,\ell = -N/2+1, \ldots, N/2$ are integers. Moreover, since $\eta(t)$ is a white-noise process, hence temporally uncorrelated, Re $\tilde{\eta}_k$ and Im $\tilde{\eta}_k$ are two mutually independent random variables, uncorrelated for different $k > 0$, with Gaussian distributions by virtue of the central limit theorem, or, equivalently, by virtue of $\eta(t)$ being the first derivative of a Wiener process with respect to time. Consequently, the sum of their squares $|\tilde{\eta}_k|^2$ is a non-negative random variable, uncorrelated for different $k > 0$, with *exponential* distribution.[14] Hence, so are the experimental values $P_f^{(\text{ex})}$ for the power spectrum for $f = k/t_{\text{msr}} > 0$.

Thus the dynamical theory defined in Eqs. (1) and (2) predicts not just the expectation value P_f for the experimental spectrum $P_f^{(ex)}$, as given in Eq. (4). It predicts also the *distribution* from which the experimental power spectrum is "drawn." The power-spectral value $P_f^{(ex)}$ at each frequency f is an independent random number, drawn from an exponential distribution with expectation value P_f

$$p(P_f^{(ex)}; P_f) = \frac{1}{P_f} \exp(-P_f^{(ex)}/P_f). \quad (8)$$

Consequently,

$$\langle P_f^{(ex)} \rangle = P_f, \quad (9)$$

$$\sigma(P_f^{(ex)}) = \langle (P_f^{(ex)} - P_f)^2 \rangle^{1/2} = P_f, \quad (10)$$

and the signal-to-noise ratio $\langle P_f^{(ex)} \rangle / \sigma(P_f^{(ex)})$ equals one. This is why we average over n experimental spectra

$$\bar{P}_f^{(ex)} \equiv \frac{1}{n} \sum_{i=1}^{n} P_{f,i}^{(ex)}, \quad (11)$$

before plotting and fitting: to reduce noise.

If the n spectra are statistically independent—as is the case if they are computed from data taken in nonoverlapping time intervals—then we have, unchanged, that

$$\langle \bar{P}_f^{(ex)} \rangle = P_f, \quad (12)$$

but

$$\sigma(\bar{P}_f^{(ex)}) = \sigma(P_f^{(ex)})/\sqrt{n} = P_f/\sqrt{n}, \quad (13)$$

and $\bar{P}_f^{(ex)}$ is distributed according to a distribution that is the convolution of n identical exponential distributions, viz., the gamma distribution

$$p_n(\bar{P}_f^{(ex)}; P_f) = \bar{P}_f^{(ex)n-1} \frac{(n/P_f)^n}{\Gamma(n)} \exp(-n\bar{P}_f^{(ex)}/P_f), \quad (14)$$

where n is the shape parameter, P_f/n is the scale parameter, and $\Gamma(n) = (n-1)!$. The mean P_f is the product of the shape and the scale parameters, the mode is $P_f(n-1)/n$, the variance P_f^2/n is the product of the mean and the scale, and the skewness is $2/\sqrt{n}$. From the known distribution, it is easy to show that the qth moment of $\bar{P}^{(ex)}$ is

$$\langle \bar{P}^{(ex)q} \rangle = \frac{\Gamma(n+q)}{n^q \Gamma(n)} P_f^q, \quad (15)$$

a result that we shall need later.

In the limit $n \to \infty$, this distribution approaches a Gaussian by courtesy of the central limit theorem, but it is a slow approach because the starting point, the exponential distribution, is highly skewed. For moderate values of n, $p_n(\bar{P}_f^{(ex)}; P_f)$ is far from Gaussian (see Fig. 1). It is also quite skewed, but this is *not* the source of the bias. The bias is caused by the correlation between the experimental mean $\bar{P}_f^{(ex)}$ and the experimental variance
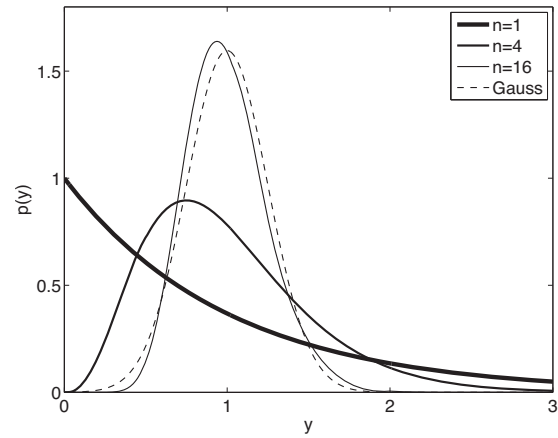

FIG. 1. The gamma distribution [Eq. (14)] as function of $\bar{P}_f^{(ex)}/P_f$ for various values of the shape parameter n. Thick full line, $n=1$; full line, $n=2$; and thin full line $n=16$. Also shown is the Gaussian limit value, plotted here with a variance of $1/n = 1/16$ (dashed line).

$$s_n^2(\bar{P}_f^{(ex)}) \equiv \frac{1}{n-1} \sum_{i=1}^{n} (P_{f,i}^{(ex)} - \bar{P}_f^{(ex)})^2, \quad (16)$$

as discussed in greater detail below. The latter is easily shown to satisfy $\langle s_n^2(\bar{P}_f^{(ex)}) \rangle = \sigma^2(\bar{P}_f^{(ex)})$.

A. Ubiquity of the gamma distribution

As a matter of fact, all systems described by a linear (integro)differential equation with additive noise, be it white or colored, have power spectral values that are described by the statistics in Eq. (14). To see that white and colored noise lead to the same statistics, all we need to realize is that the PSD of a colored noise is the product of the PSD of a white noise and some function $F(f)$ that describes the color of the noise (a filter function). Quite generally, if the dynamics of x is given by a polynomial Q in d/dt with constant coefficients

$$Q\left(\frac{d}{dt}\right) x(t) = \psi(t), \quad (17)$$

where ψ is a colored noise, then the power spectrum

$$P_f^{(ex)}(x) = \frac{|\tilde{\psi}_f|^2}{|Q(-i2\pi f)|^2} = \frac{|\tilde{\eta}_f|^2 F(f)}{|Q(-i2\pi f)|^2} \quad (18)$$

is seen to be the product of the power spectrum of a white noise η and some function of f. Thus, for a given frequency f, the statistics of the power spectral value is determined by the statistics of white noise.

V. LEAST-SQUARES FITTING

As alluded to in the previous section, one effect of the exponential distribution of $P_{f,i}^{(ex)}$ is that weighted least-squares (WLS) fitting to $\bar{P}_f^{(ex)}$ does not follow the textbook behavior and returns a result that is systematically wrong or biased. In the following, we remind the reader of the rather narrow set of circumstances under which least-squares fitting works. We then relax the requirements and see what effect it has on the fit results.

WLS fitting minimizes

$$\chi^2(\theta) = \sum_{i=1}^{N} [y_i - f_i(\theta)]^2 w_i^2, \tag{19}$$

with respect to fit parameters θ. Here, $y_i = f_i + e_i$ are the experimental data (dependent variable), e_i denotes errors, f_i denotes the fit function evaluated at i (the independent variable), and w_i are weights. It is assumed that the experimental data are uncorrelated and that the weights are uncorrelated with the experimental data. If, in addition, the weights are also independent of i, and thus constant, WLS reduce to OLS. When the theoretical model f_i is a linear function of the fit parameters θ, both OLS and WLS minimizations are known to return parameter estimates that are best linear unbiased estimate. The former was first shown by Gauss in 1829 whereas the latter was shown by Aitken[15] in 1935. Specifically, in WLS the errors do not have to be independent, just uncorrelated, and in OLS they do not have to come from the same distribution, they are only assumed to have the same variance (homoscedasticity).

Things turn particularly simple when the experimental data are Gaussian distributed. We can then choose y_i, e.g., as the average of several measurements, and the weights w_i, as the reciprocal of the experimental standard deviation. With this choice, y_i and w_i are independent because the sample mean and sample variance calculated from independent, identically Gaussian distributed random variables are statistically independent. In this Gaussian scenario, it is possible to assign meaningful confidence intervals to the fit parameters and calculate the goodness of the fit, also known as its *support* or p-value, and this is why one is sometimes tempted to assume that data are Gaussian even when it is not quite the case. Finally, with Gaussian data, MLE and WLS are mathematically identical as it easily seen by inserting a Gaussian in place of p_n in Eq. (50) and deriving the corresponding cost function, which is simply χ^2.

We now ask what the effect is on the estimate of the fit parameters if the data are not Gaussian, the weights are not independent of the experimental value, and the model is a nonlinear function of the fit parameters. All of these circumstances are realized when fitting power spectra with the statistic given in the previous section.

A. Some analytical results for bias in least-squares fitting

An estimator $\hat{\theta}$ is biased if its expectation value differs from the true value $\langle \hat{\theta} \rangle \neq \theta^*$ of the quantity it is an estimator for. Below, we show that some common least-squares estimators for power spectra are biased.

At the minimum of the χ^2 function given in Eq. (19), the first derivative with respect to the fit parameters is zero, and the stationarity conditions thus read

$$\sum_{i=1}^{N} [y_i - f_i] w_i^2 \partial_\theta f_i = \sum_{i=1}^{N} [y_i - f_i]^2 w_i \partial_\theta w_i. \tag{20}$$

Notice that in order to be completely general, we allowed the weights to depend on the fit parameters. We now treat a few specific, but universal, scenarios one at a time.

1. Experimental standard deviations as weights

Proceeding as usual with classic WLS, we pick the weights to be inversely proportional to the experimental estimate for the standard deviation. This estimate could be the known uncertainty from the experimental apparatus, but often it is estimated simply as

$$s_{n,i} = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (y_{i,j} - \bar{y}_i)^2}, \tag{21}$$

where n is the number of times the experiment is repeated with the independent variable set to its ith value. The experimental average (sample mean)

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{i,j} \tag{22}$$

is our estimate for the expectation value of y_i. Doing this for each i,

$$\chi^2(\theta) = \sum_{i=1}^{N} [\bar{y}_i - f_i(\theta)]^2 s_{n,i}^{-2}. \tag{23}$$

The stationarity conditions then reduce to

$$\sum_{i=1}^{N} \frac{\bar{y}_i}{s_{n,i}^2} \partial_\theta f_i = \sum_{i=1}^{N} \frac{1}{s_{n,i}^2} f_i(\theta) \partial_\theta f_i, \tag{24}$$

the solution of which gives us our estimate $\hat{\theta}[y]$ for θ for a given data set $y = (y_{i,j})_{i=1,\ldots,N;\ j=1,\ldots,n}$. Note that because N and n are finite, the estimate for θ resulting from Eq. (24) is a stochastic quantity and will vary from one experimental realization to another.

To calculate the bias of the estimator for given n, we need to find its expectation value for that n. We do this in two steps: first, we let the number of data points $N \to \infty$, while keeping n and the number of fit parameters fixed, as in an infinite experiment. In this limit, the fit returns an estimate θ_n^* of θ that is no longer a fluctuating quantity, but may still depend on n. So, for $N = \infty$, Eq. (24) reads

$$\sum_{i=1}^{\infty} \frac{\bar{y}_i}{s_{n,i}^2} \partial_\theta f_i(\theta_n^*) = \sum_{i=1}^{\infty} \frac{1}{s_{n,i}^2} f_i(\theta_n^*) \partial_\theta f_i(\theta_n^*). \tag{25}$$

Note that the experimental values $\bar{y}_i$ and $s_{n,i}$ are still fluctuating quantities described by the same statistic as before, whereas $f_i(\theta_n^*)$ and $\partial_\theta f_i(\theta_n^*)$ are not fluctuating because they are functions of nonfluctuating variables. So, when in the next step we take the expectation value of Eq. (25), only $\bar{y}_i$ and $s_{n,i}$ are affected,

$$\sum_{i=1}^{\infty} \left\langle \frac{\bar{y}_i}{s_{n,i}^2} \right\rangle \partial_\theta f_i(\theta_n^*) = \sum_{i=1}^{\infty} \left\langle \frac{1}{s_{n,i}^2} \right\rangle f_i(\theta_n^*) \partial_\theta f_i(\theta_n^*), \tag{26}$$

which is solved by

$$f_i(\theta_n^*) = \frac{\langle \bar{y}_i / s_{n,i}^2 \rangle}{\langle 1/s_{n,i}^2 \rangle}, \quad \text{for all } i = 1, \ldots, \infty, \tag{27}$$

if a parameter set θ_n^* exists, which solves all these many equations simultaneously. Provided that we know the distribution of y_i, we can calculate the expectation values at the right-hand-side of this equation and thus determine whether

the fit is biased. For a start, note that if the sample mean and the reciprocal sample variance are uncorrelated, then the numerator on the right-hand-side factorized to $\langle \bar{y}_i \rangle \langle 1/s_{n,i}^2 \rangle$. Thus the right-hand-side equals $\langle \bar{y}_i \rangle$, i.e., the fit is unbiased. So it is sufficient that mean and reciprocal variances are uncorrelated to ensure an unbiased fit.

It turns out that it is also a necessary condition if there is no redundancy in the parameterization of f_i by θ, i.e., if their relationship is one-to-one. This is the only sensible way to parametrize a function and easily achieved by elimination of a possible redundancy through reparametrization. Assuming this one-on-one relationship, we use that $f_i(\theta^*) = \langle \bar{y}_i \rangle$ and Eq. (27) to write

$$f_i(\theta^*) = \left[\langle \bar{y}_i \rangle \frac{\langle 1/s_{n,i}^2 \rangle}{\langle \bar{y}_i/s_{n,i}^2 \rangle} \right] f_i(\theta_n^*), \quad \text{for all } i = 1, \ldots, \infty. \quad (28)$$

We now see that the fit is unbiased if and only if the term in square brackets in Eq. (28) equals unity.

In summary, *a necessary and sufficient criterion for least-squares fitting to be unbiased is that the sample mean and reciprocal sample variance are uncorrelated.* Note that "uncorrelated" is a weaker requirement than "independent" and although the latter implies the former, the reverse is not generally true. Naturally, when the sample mean and sample variance are independent, the sample mean and reciprocal sample variance are also uncorrelated.

Notice that we nowhere used what $s_{n,i}$ is, and in fact a more general version of Eq. (28) is

$$f_i(\theta^*) = \left[\langle y_i \rangle \frac{\langle w_i^2 \rangle}{\langle y_i w_i^2 \rangle} \right] f_i(\theta_n^*), \quad (29)$$

where w_i is assumed independent of θ but is otherwise unconstrained. The criterion for an unbiased estimator is now simply that the data y_i and the squared weights w_i^2 are uncorrelated.

Under some circumstances, it is possible to determine the bias of the fit parameters from the bias of the fit function given in Eq. (29). If the function is invertible this is trivially the case. However, it is also possible if the amplitude of the fit function is determined by just one of the fit parameters and the term in the square brackets is independent of i. In this case, bias can be removed from the parameter estimate $\hat{\theta}$ by simply multiplying the amplitude-parameter by the square bracket and leaving the other parameters unchanged. In Sec. V A 2, we give an example of this.

2. Experimental averages as weights

If the sample standard deviation is proportional to the sample mean, $s_{n,i} \propto \bar{y}_i$ for each i, the mean and variance are clearly not uncorrelated and we expect the parameter estimate to be biased. From Eq. (28) we find

$$f_i(\theta^*) = \left[\langle \bar{y}_i \rangle \frac{\langle \bar{y}_i^{-2} \rangle}{\langle \bar{y}_i^{-1} \rangle} \right] f_i(\theta_n^*). \quad (30)$$

As a real-world example, consider gamma-distributed experimental data. We already mentioned that the gamma distribution describes all power spectra resulting from linear dynamical equations driven by an additive noise, white or colored. Carrying on as before, we use the standard deviation to weight the data, $w_f = 1/s_n(\bar{P}_f^{(ex)})$, where f is the frequency. Since the standard deviation and the mean are proportional [see Eq. (13)], we have $w_f \propto 1/\bar{P}_f^{(ex)}$. Using the known gamma distribution of $\bar{P}_f^{(ex)}$ [see Eq. (14)], we can compute $\langle (\bar{P}_f^{(ex)})^{-q} \rangle$ for $q = 1, 2, \ldots$ and find the value of the bracketed bias term in Eq. (30),

$$\left[\langle \bar{P}_f^{(ex)} \rangle \frac{\langle (\bar{P}_f^{(ex)})^{-2} \rangle}{\langle (\bar{P}_f^{(ex)})^{-1} \rangle} \right] = n/(n-2). \quad (31)$$

That is, *WLS fitting of this wide class of power spectra has a built-in, frequency independent, multiplicative bias of $(n-2)/n$ for $n > 2$.* The true power spectral form can then be obtained from the least-squares fit as

$$P_f(\theta^*) = \frac{n}{n-2} P_f(\theta_n^*). \quad (32)$$

In this scenario, the theoretical $N \to \infty$ limit that we took in going from Eq. (24) to Eq. (25) corresponds to letting the measurement time (or sampling frequency) become infinite while keeping the sampling frequency (or measurement time) and all other experimental factors fixed.

In the simple case of a Lorentzian, Eq. (3), we have $\theta^* = (D, f_c)$ and

$$P_f(\theta^*) \equiv \frac{D/(2\pi^2)}{f_c^2 + f^2} = \frac{n}{n-2} \frac{D_n^*/(2\pi^2)}{(f_{c,n}^*)^2 + f^2}, \quad (33)$$

from which we see that $f_{c,n}^*$ is unbiased. In other words, the true value of the fit parameters can be obtained from the WLS estimates as

$$D = \frac{n}{n-2} D_n^* \approx \frac{n}{n-2} D^{(lsq)}, \quad (34)$$

$$f_c = f_{c,n}^* \approx f_c^{(lsq)}, \quad (35)$$

where $D^{(lsq)}$ and $f_c^{(lsq)}$ are the stochastically fluctuating values returned by a least-squares fit to a finite N data set. This is a general feature of the power spectra. They can be written as the product of an overall multiplicative scale factor and a shape function depending on f, where *only the scale factor is influenced by the bias.*

3. Theoretical values as weights

If the sample standard deviation is known to be proportional to $f_i(\theta^*)$, it seems reasonable to weight the data by the theoretical value $1/f_i$. After inserting $w_i \propto 1/f_i$ in Eq. (19), the quantity to minimize is

$$\chi^2(\theta) = \sum_{i=1}^{N} [y_i - f_i(\theta)]^2 f_i^{-2}(\theta), \quad (36)$$

and the stationarity equations (20) become

$$\sum_{i=1}^{N} y_i f_i^{-2} \partial_\theta f_i = \sum_{i=1}^{N} y_i^2 f_i^{-3} \partial_\theta f_i. \quad (37)$$

Repeating the arguments from Sec. V A 1, the expectation value of the stationarity conditions in the limit $N \to \infty$ is

$$\sum_{i=1}^{\infty} \langle y_i \rangle f_i^{-2}(\theta_n^*) \partial_\theta f_i(\theta_n^*) = \sum_{i=1}^{\infty} \langle y_i^2 \rangle f_i^{-3}(\theta_n^*) \partial_\theta f_i(\theta_n^*), \qquad (38)$$

which is solved by

$$f_i(\theta_n^*) = \frac{\langle y_i^2 \rangle}{\langle y_i \rangle}, \quad \text{for all } i = 1, \dots, \infty, \qquad (39)$$

if a parameter set θ_n^* exists, which solves all these many equations simultaneously. The true value $f_i(\theta^*)$ can in that case be written as

$$f_i(\theta^*) = [\langle y_i \rangle^2 / \langle y_i^2 \rangle] f_i(\theta_n^*), \qquad (40)$$

and θ^* can be determined from θ_n^* if $\langle y_i \rangle^2 / \langle y_i^2 \rangle$ is independent of i, constant, and can be absorbed in θ_n^* in a simple manner.

If we again use the gamma-distributed power spectral data as example, we see that

$$P_f(\theta^*) = \frac{n}{n+1} P_f(\theta_n^*), \qquad (41)$$

i.e., we once more have a bias that scales with the number n of spectra averaged over, although this time the bias is half the size and positive.

Closed-form expressions for $f_{c,n}^*$ and D_n^* are straightforward to obtain when P_f is a Lorentzian[9] (Ref. 9 contains a typographical error: DT_{msr} should simply read D). The expression for $f_{c,n}^*$ given in Ref. 9 is unbiased, whereas the expression for D_n^* has its bias removed by multiplying by $n/(n+1)$. Note that apart from the $n/(n+1)$ factor on D, the results derived in Ref. 9 are mathematically identical to those derived using an MLE approach in Sec. V A 4.

4. Constant values as weights ("unweighted")

When all weights are assumed to be equal, the function to minimize

$$\chi^2(\theta) = \sum_{i=1}^{N} [y_i - f_i(\theta)]^2 \qquad (42)$$

has stationarity equations which, for $N \to \infty$, gives

$$\sum_{i=1}^{\infty} f_i(\theta_n^*) \partial_\theta f_i(\theta_n^*) = \sum_{i=1}^{\infty} y_i \partial_\theta f_i(\theta_n^*). \qquad (43)$$

Taking the expectation value on both sides, we have

$$\sum_{i=1}^{\infty} f_i(\theta_n^*) \partial_\theta f_i(\theta_n^*) = \sum_{i=1}^{\infty} \langle y_i \rangle \partial_\theta f_i(\theta_n^*). \qquad (44)$$

This equation may have the solution

$$f_i(\theta_n^*) = \langle y_i \rangle = f_i(\theta^*). \qquad (45)$$

In other words, this estimator is unbiased. However, it is not a precise estimator. The stochastic errors on values it returns tend to be larger than those obtained with a weighted fit. Figures 2 and 3 illustrate this for the case of fitting an aliased Lorentzian to power spectra from OTs.
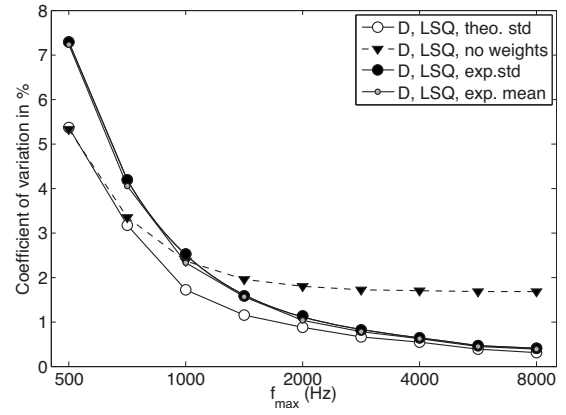

FIG. 2. Stochastic error on fit parameters as a function of cutoff frequency f_{max} for least-squares fitting of aliased Lorentzians using various weights [Eq. (19)]. For all fits, the errors are substantial for low values of f_{max}. Notice the logarithmic frequency scale. For large f_{max}, WLS returns a variance nearly an order of magnitude smaller than for OLS (fitting with constant weights). The total number of acquired data points was held fixed at $N = 262\,144$, and the fit was done to the average of $n = 16$ power spectra generated with $f_c = 500$ Hz, $D = 0.46$ μm/s^2, and $f_{sample} = 16\,384$ Hz. Symbols show the coefficient of variation $s_n(D)/\bar{D}$, with D measured over 100 independent stochastic simulations. The coefficient of variation for f_c shows similar trends and is roughly two times larger at these settings (data not shown). For OLS [triangles, Eq. (42)] the error is independent of f_{max} for large frequencies because the information there is de-emphasized by the fitting algorithm. Using the experimental standard deviation or mean as weights [filled circles, Eq. (23)] leads to *stochastic* errors nearly as small as when using theoretical weights [empty circles, Eq. (36)]—the *systematic* errors (biases) are twice as big and of the opposite sign however, see Fig. 4.
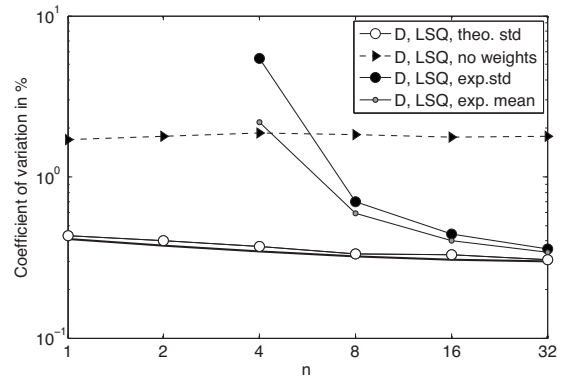

FIG. 3. Stochastic error on fit parameters as a function of number of power spectra averaged over n for least-squares fitting of aliased Lorentzians using various weights, Eq. (19). Constant weights (triangles), experimental standard deviation (black circles), experimental average (gray circles), and theoretical weights (white circles). The thick black line through white circles shows the theoretical prediction [Eq. (E2)]. The number of acquired data points was held fixed at $N = 262\,144$, and the fit was done to all the data $f_{max} = f_{Nyq}$ for power spectra generated with $f_c = 500$ Hz and $D = 0.46$ μm/s^2. Symbols show the coefficient of variation $s_n(D)/\bar{D}$, with D measured over 1000 independent stochastic simulations. Using theoretical weights clearly outperforms all the other weighing schemes for $n < 32$; for $n > 32$, the experimental weighing scheme performs comparably well—the use of the experimental standard deviation as weight leads to slightly larger variation than use of the experimental average, as expected, and both show more variation than results obtained with theoretical weights. Results obtained with constant weights—sometimes referred to as "no weights" or "unweighted"—show roughly ten times larger variation than does results obtained with theoretical weights, for all n.

5. Iterated values as weights

When weights are determined iteratively, a fit is performed and the value returned for θ is used in the weight of a new fit and so on, until some convergence criterion has been met. The function to minimize with respect to θ for given $\theta^{(\text{iter})}$ reads

$$\chi^2(\theta) = \sum_{i=1}^{N} [y_i - f_i(\theta)]^2 w_i^2(\theta^{(\text{iter})}). \tag{46}$$

It has the stationarity equations

$$\sum_{i=1}^{N} f_i w_i^2(\theta^{(\text{iter})}) \partial_\theta f_i = \sum_{i=1}^{N} y_i w_i^2(\theta^{(\text{iter})}) \partial_\theta f_i. \tag{47}$$

We now use the same arguments as above together with the observation that at the fixed-point solution to this iterative scheme, $\theta^{(\text{iter})} = \theta_n$ and both equal θ_n^* in the limit $N \to \infty$. Taking the expectation value of the equation after taking the limit $N \to \infty$, we see that θ_n^* may satisfy

$$f_i(\theta_n^*) = \langle y_i \rangle = f_i(\theta^*), \tag{48}$$

in which case this estimator is unbiased. The estimation scheme just described is also known as "iteratively re-weighted least squares" and convergence is not guaranteed.

6. Error bars

Error bars on the parameter estimates $\hat{\theta}$ for θ^* can be calculated from the error bars on the estimator $\hat{\theta}_n$ for θ_n^* and the simple relationship in terms of n between θ^* and θ_n^*, by propagation of errors [see Eq. (73)], provided the latter exists. The error bars on $\hat{\theta}_n$ can be those returned by a least-squares fitting routine used to determine $\hat{\theta}_n$, or they may be known theoretically. In that connection, note that the error bars given in Ref. 9 for WLS with theoretical weights are correct only in the limit $n = \infty$, but by replacing D with $n/(n+1)D$, they are correct for all n.

When $\hat{\theta}$ is found from a $\hat{\theta}_n$, which itself is found by minimizing χ^2, the precision with which χ^2 is known determines the precision of the estimate $\hat{\theta}$. Because the various weights in χ^2 have different stochastic fluctuations, we expect that theoretical weights [Eq. (36)] will give smaller variation in $\hat{\theta}$ than experimental weights [Eq. (23)]. A small calculation shows that the variance of the stationarity equations [Eq. (24)] around zero includes a term $\langle \bar{P}_f^{(\text{ex})-4} \rangle \propto \{(n-1)(n-2)(n-3)(n-4)\}^{-1}$, i.e., the error bars are expected to be large for $n \sim 4$, if experimental weights are used. When theoretical weights are used, the error bars are small and well defined for all n. These results are verified by example (see Fig. 3).

B. Results for power spectra

We verified the above theoretical results by Monte Carlo simulations of the OT and AFM power spectra and found perfect agreement. Below, we list the specific results.

For the diffusion coefficient D, we found that the outcome of least-squares fitting $D^{(\text{lsq})}$ depends on the number, n, of power spectra averaged over in the following manner for
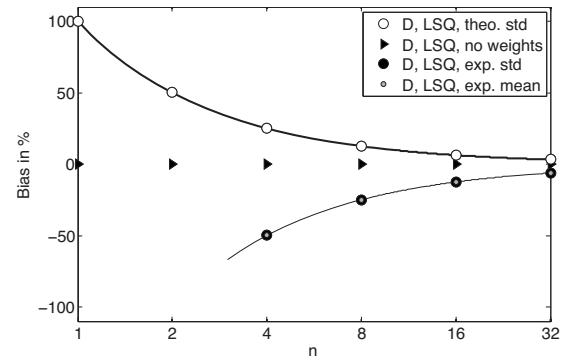

FIG. 4. Systematic error (bias) on fit parameters as a function of number of power spectra averaged over n for least squares fitting of aliased Lorentzians using various weights, Eq. (19). The corner frequency is unbiased (data not shown), independent of n, whereas the diffusion coefficient shows a strong weighing-scheme dependent bias. The sampling frequency was held fixed at $f_{\text{sample}} = 16\,384$ Hz, and the fit was done to the averaged power spectra generated with $N = 262\,144$, $f_c = 500$ Hz, and $D = 0.46$ μm/s^2. Symbols show the signed error on the mean of D measured over 1000 independent stochastic simulations. For least squares fitting with constant weights (triangles) the bias on D is zero. Least squares fitting with experimental weights (black circles: standard deviation; gray circles: average value) has a systematic error on D with $-2/n$ dependence (thin black line) for $n > 2$. Theoretical weights (white circles) lead to an $1/n$-dependence (thick black line) in the systematic error on D.

the aliased (see Fig. 4) and nonaliased Lorentzians, as well as the aliased and nonaliased AFM expressions (data not shown):

(1) $D^{(\text{lsq})} = D(n-2)/n$, $n > 2$. If data points are weighted with their experimental standard deviation or mean, $w_f = 1/s_n(\bar{P}_f^{(\text{ex})})$ or $w_f = \sqrt{n}/\bar{P}_f^{(\text{ex})}$. This is the weighting choice with the largest bias and with large stochastic errors for $n \leq 4$. Here, D is underestimated.

(2) $D^{(\text{lsq})} = D(n+1)/n$. If data points are weighted using the theoretical expectation value for the standard deviation, $w_f = \sqrt{n}/P_f$. This is the least-squares fitting of Ref. 9; a hybrid of MLE and LSQ, analytically tractable and thus very robust. It also has the smallest stochastic errors. Here, $D^{(\text{lsq})}$ overestimates D.

(3) $D^{(\text{lsq})} = D$. If the weights are updated iteratively to the fitted value, $w_f = 1/P_f^{\text{iter}}$. This is a practically correct but computationally intensive and somewhat unstable approach; the initial guess must be close to the true value, or the algorithm will not converge.

(4) $D^{(\text{lsq})} = D$. If all data points are given the same weight, $w_f = \text{const}$. This method de-emphasizes information available at higher frequencies because power spectral values have constant *relative* error, hence rapidly decreasing *absolute* error beyond f_c or f_0, yet are treated as having *constant* absolute error. Consequently, this estimator has low precision, typically with two to ten times larger stochastic errors than the cases (1)–(3) above. This is OLS fitting applied far outside its range of validity.

Figures 2–4 show examples of the size of these stochastic errors and biases, respectively, for fits of the aliased Lorentzian to data.

We found that the expectation values of the fitted values

for f_c, f_0, and Q were independent of weighing scheme and unbiased, whereas the standard deviations varied by up to an order of magnitude depending on weighing scheme applied. Furthermore, fitting of nonaliased PSDs to aliased data introduces large systematic errors for the trivial reason that they fail to capture the shape of the PSD near the Nyquist frequency (data not shown).

VI. MLE OF FIT PARAMETERS

We now proceed to determine the fit parameters by the method of MLE. We also derive closed-form expressions for the fit parameters in terms of the experimental data values. They should be useful for speedy online calibration.

Both power spectra, given in Eqs. (3) and (4), are consequences of a linear, time-invariant dynamics driven by a white noise [Eq. (1)]. They are consequently of the form

$$P_f = \frac{1}{a + bf^2 + cf^4}, \tag{49}$$

with $c=0$ in the case of the Lorentzian, and (a,b) or (a,b,c) parameters to be fitted. This simple form, combined with the simple statistical properties of the experimental power spectral values, makes rigorous MLE of the parameters a straightforward numerical optimization problem as shown below.

When fitting, we fit only to the positive-frequency part of the power spectrum or to a subset of it. So here we considered only that part of the spectrum. Since $\bar{P}_f^{(ex)}$ and $\bar{P}_{f'}^{(ex)}$ are uncorrelated for $f \neq f'$, the probability density for the experimental spectrum $\bar{P}_f^{(ex)}$, given its expectation value P_f, is

$$p(\bar{P}^{(ex)}|P) = \prod_f p_n(\bar{P}_f^{(ex)}; P_f), \tag{50}$$

where p_n is given in Eq. (14). Thus MLE of the theory's parameter values consists in choosing these parameters so they maximize $p(\bar{P}^{(ex)}|P)$ for given $\bar{P}^{(ex)}$, or, equivalently, minimize the negative logarithm

$$\mathcal{F}' \equiv n \sum_f (\bar{P}_f^{(ex)}/P_f + \ln P_f) + \text{const}, \tag{51}$$

where

$$\text{const} = \ln \Gamma(n) - n \ln n - (n-1)\ln \bar{P}_f^{(ex)} \tag{52}$$

is a constant with respect to (a,b,c) so it can be ignored when minimizing. We are then left with the task of finding the values of (a,b,c) that minimize the cost function

$$\mathcal{F}(a,b,c) \equiv \sum_f (\bar{P}_f^{(ex)}/P_f + \ln P_f), \tag{53}$$

which is an uncomplicated optimization problem that can be solved numerically with standard programs. Good starting values are given in Eqs. (63) and (64).

A. From nonlinear to linear stationarity equations via a simple trick

Although we gave the solution to the full MLE problem implicitly above, as the minimum of $\mathcal{F}$, Eq. (53), we can speed up the fitting process substantially (numerical optimization can be slow when the data sets are large) and gain more insight into the problem by taking a few more analytical steps before turning to numerics.

1. Results for OTs

For P_f given in Eq. (3), written as in Eq. (49) with $c=0$, $\mathcal{F}$ in Eq. (53) reads

$$\mathcal{F}(a,b) = \sum_f [(a + bf^2)\bar{P}_f^{(ex)} - \ln(a + bf^2)]. \tag{54}$$

It is minimized with respect to a and b when these parameters satisfy the stationarity condition

$$\sum_f \bar{P}_f^{(ex)} = \sum_f P_f = \sum_f \frac{1}{a + bf^2},$$

$$\sum_f f^2 \bar{P}_f^{(ex)} = \sum_f f^2 P_f = \sum_f \frac{f^2}{a + bf^2}. \tag{55}$$

These are nonlinear equations for a and b. However, from Eq. (15) we know that we can write

$$P_f = \frac{n\langle \bar{P}_f^{(ex)2}\rangle}{(n+1)P_f} = \frac{n\langle \bar{P}_f^{(ex)2}\rangle}{(n+1)}(a + bf^2). \tag{56}$$

Substituting for P_f in Eq. (55) we find

$$\sum_f \bar{P}_f^{(ex)} = \frac{n}{n+1}\sum_f \langle \bar{P}^{(ex)2}\rangle(a + bf^2),$$

$$\sum_f f^2 \bar{P}_f^{(ex)} = \frac{n}{n+1}\sum_f f^2\langle \bar{P}^{(ex)2}\rangle(a + bf^2), \tag{57}$$

which is linear in a and b. Before solving for a and b we introduce the statistic

$$S_{p,q} \equiv \frac{1}{K}\sum_f f^{2p}\bar{P}_f^{(ex)q}, \tag{58}$$

with K as the number of terms in the sum, and the statistic obeys $\lim_{K\to\infty} S_{p,q} = \langle S_{p,q}\rangle$.

The sums should only include those frequencies the user deems relevant, i.e., frequencies corresponding to mechanical or electronic resonances can be excluded, and high/low frequency cutoffs can be applied. That is, the statistics can be trimmed iteratively if required. Power spectral values too far from a fit can be identified and excluded from the sums, after which a new fit is found, etc., until a steady state is reached and all power spectral values satisfy the user-defined acceptance criterion. When no frequencies are excluded, $nK=N = t_{msr}f_{sample}$, the total number of data acquired. In this latter case, one fits the power spectrum all the way out to the Nyquist frequency and aliasing should be taken into account (see the appendices); unless aliasing is eliminated by over-sampling data acquisition electronics.

With this notation, Eq. (57) can be written in matrix form

$$\begin{pmatrix} \langle S_{0,2} \rangle & \langle S_{1,2} \rangle \\ \langle S_{1,2} \rangle & \langle S_{2,2} \rangle \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \frac{n+1}{n} \begin{pmatrix} S_{0,1} \\ S_{1,1} \end{pmatrix}, \qquad (59)$$

with solution

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1+1/n}{\langle S_{0,2} \rangle \langle S_{2,2} \rangle - \langle S_{1,2} \rangle^2} \begin{pmatrix} S_{0,1} \langle S_{2,2} \rangle - S_{1,1} \langle S_{1,2} \rangle \\ S_{1,1} \langle S_{0,2} \rangle - S_{0,1} \langle S_{1,2} \rangle \end{pmatrix}. \quad (60)$$

In this expression, we know $S_{p,q}$ from the experiment, but not $\langle S_{p,q} \rangle$. However, we can always write $S_{p,q} = \langle S_{p,q} \rangle + \delta S_{p,q}$ and substitute for $\langle S_{p,q} \rangle$ in Eq. (60). About $\delta S_{p,q}$ it is easy to show that $\langle \delta S_{p,q} \rangle = 0$ and

$$\langle (\delta S_{p,q})^2 \rangle = \frac{1}{K^2} \left\{ \frac{\Gamma(n+2q)}{n^{2q} \Gamma(n)} - \left[\frac{\Gamma(n+q)}{n^q \Gamma(n)} \right]^2 \right\} \sum_f f^{4p} P_f^{2q}$$

$$= \frac{1}{K} [\text{weak } n\text{-dep.}] \widetilde{S}_{2p,2q}, \qquad (61)$$

where $\widetilde{S}$ is defined as in Eq. (58) except for $P_f = \langle \bar{P}_f^{(ex)} \rangle$ replacing $\bar{P}_f^{(ex)}$. Above, $q=2$ in the relevant expressions and therefore $\widetilde{S}_{p,q}$ scales as $1/K$ for $p=0,1$ and is independent of K for $p=2$. Thus, apart from some n-dependence, the variance of $S_{p,q}$ scales as $1/K^2$ or $1/K$ both of which are very small numbers in a typical experiment, where $K=N/n$ is of order $10^4 - 10^6$. It is therefore an excellent approximation to replace $\langle S_{p,q} \rangle$ with $S_{p,q}$ in Eq. (60),

$$\begin{pmatrix} a \\ b \end{pmatrix} \approx \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \equiv \frac{1+1/n}{S_{0,2} S_{2,2} - S_{1,2}^2} \begin{pmatrix} S_{0,1} S_{2,2} - S_{1,1} S_{1,2} \\ S_{1,1} S_{0,2} - S_{0,1} S_{1,2} \end{pmatrix}, \quad (62)$$

where now $(\hat{a}, \hat{b})$ are our estimates of (a,b). By comparing Eqs. (3) and (49) we then have

$$f_c = \left(\frac{a}{b} \right)^{1/2} \approx \left(\frac{\hat{a}}{\hat{b}} \right)^{1/2} = \left(\frac{S_{0,1} S_{2,2} - S_{1,1} S_{1,2}}{S_{1,1} S_{0,2} - S_{0,1} S_{1,2}} \right)^{1/2}, \qquad (63)$$

$$D = \frac{2\pi^2}{b} \approx \frac{2\pi^2}{\hat{b}} = \frac{2\pi^2 n}{n+1} \frac{S_{0,2} S_{2,2} - S_{1,2}^2}{S_{1,1} S_{0,2} - S_{0,1} S_{1,2}}. \qquad (64)$$

2. Results for AFM cantilevers

Identical reasoning leads, in the case of $c \neq 0$, to the three coupled linear equations for a, b, and c,

$$\underbrace{\begin{pmatrix} S_{0,2} & S_{1,2} & S_{2,2} \\ S_{1,2} & S_{2,2} & S_{3,2} \\ S_{2,2} & S_{3,2} & S_{4,2} \end{pmatrix}}_{\mathbf{S}} \underbrace{\begin{pmatrix} a \\ b \\ c \end{pmatrix}}_{\vec{v}} \approx (1+1/n) \underbrace{\begin{pmatrix} S_{0,1} \\ S_{1,1} \\ S_{2,1} \end{pmatrix}}_{\vec{s}}, \qquad (65)$$

which are inverted to give

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \approx \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} \equiv \frac{n+1}{n} \mathbf{S}^{-1} \begin{pmatrix} S_{0,1} \\ S_{1,1} \\ S_{2,1} \end{pmatrix}. \qquad (66)$$

Comparing Eqs. (4) and (49) then gives

$$f_0 = \left(\frac{a}{c} \right)^{1/4} \approx \left(\frac{\hat{a}}{\hat{c}} \right)^{1/4}, \qquad (67)$$

$$D = \frac{2\pi^2}{b + 2(ac)^{1/2}} \approx \frac{2\pi^2}{\hat{b} + 2(\hat{a}\hat{c})^{1/2}}, \qquad (68)$$

$$\left(\frac{2\pi m}{\gamma} \right)^2 = \frac{c}{b + 2(ac)^{1/2}} \approx \frac{\hat{c}}{\hat{b} + 2(\hat{a}\hat{c})^{1/2}}. \qquad (69)$$

For completeness and ease of implementation we give the inversion formulas $\mathbf{S}^{-1} = \mathbf{C} / \mathcal{D}$,

$$\mathcal{D} = S_{0,2} S_{2,2} S_{4,2} - S_{0,2} S_{3,2}^2 - S_{1,2}^2 S_{4,2} + 2 S_{1,2} S_{2,2} S_{3,2} - S_{2,2}^3, \qquad (70)$$

$$\mathbf{C} = \begin{pmatrix} C_0 & C_1 & C_2 \\ C_1 & C_5 & C_3 \\ C_2 & C_3 & C_4 \end{pmatrix}, \qquad (71)$$

and

$$\begin{aligned} C_0 &= S_{2,2} S_{4,2} - S_{3,2}^2, \\[4pt] C_1 &= S_{2,2} S_{3,2} - S_{1,2} S_{4,2}, \\[4pt] C_2 &= S_{1,2} S_{3,2} - S_{2,2}^2, \\[4pt] C_3 &= S_{1,2} S_{2,2} - S_{0,2} S_{3,2}, \\[4pt] C_4 &= S_{0,2} S_{2,2} - S_{1,2}^2, \\[4pt] C_5 &= S_{0,2} S_{4,2} - S_{2,2}^2. \end{aligned} \qquad (72)$$

Of course, it is also possible to numerically invert Eq. (65) and insert the resulting values for $\hat{a}, \hat{b}, \hat{c}$ in Eqs. (63), (64), and (67)–(69). The above results, however, should resolve potential issues with numerical stability arising from the inversion of near-singular matrices.

B. Examples of fitting the power spectra

Figure 5(a) shows the average of $n=16$ power spectra for the Brownian motion of a microsphere in an optical trap. The data are synthetic and computer generated (see Ref. 9). Thus, the values for f_c and D are known exactly and can be used as benchmarks for results obtained by fitting, with no worry about any of the complicating circumstances that can affect data in the real world. The red line is the expectation value of the aliased Lorentzian, taking as input the exactly known values of f_c and D. The yellow line is the aliased Lorentzian corresponding to the stochastically realized parameter values for f_c and D, determined by rigorous MLE of (a,b) with no use of our simplifying trick. The dashed blue line is the result of MLE, with the use of our simplifying trick. All of these three lines plot virtually on top of each other. Finally, the black lines show the result of least-squares fitting of an aliased Lorentzian to the data, using as weights: (i) the standard deviations on the $n=16$ spectra with a resulting $-2/n=-12.5\%$ systematic error on D [see Eq. (32)] and (ii) theoretical weights, resulting in a $1/n=6.25\%$ systematic error on D [see Eq. (41)].
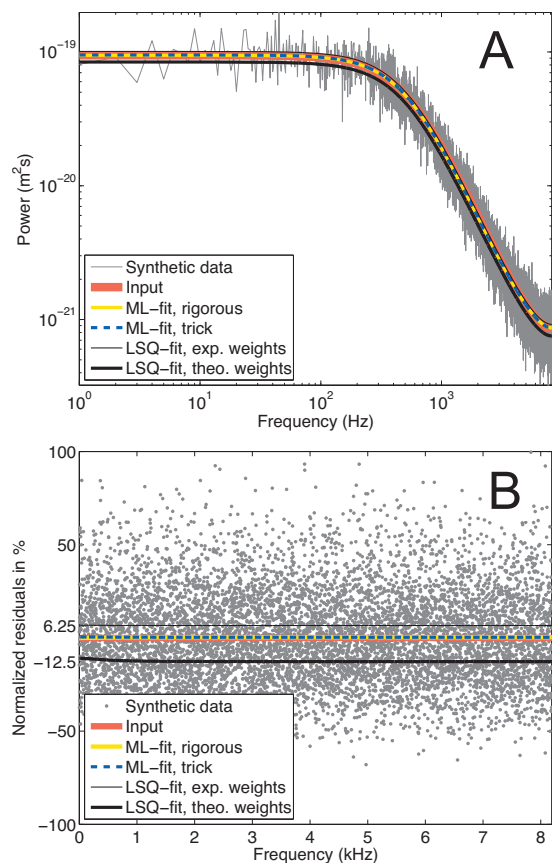
FIG. 5. (Color) Various fits of aliased Lorentzian to synthetic power spectrum for microsphere in optical trap. Data are artificial, with known parameter values $f_c = 500$ Hz and $D = 0.46$ $\mu m^2/s$, $f_{sample} = 16\,384$ Hz, and $t_{msr} = 16$ s. (a) Data and aliased Lorentzians. PSD values (gray line) are averages over 16 spectra. Note how the three Lorentzians corresponding to, respectively, the exactly known values of f_c and D (red line), rigorous numerical ML-fit (yellow line), and analytical ML-fit using simplifying trick (dashed blue line) all plot on top of each other. In contrast, the Lorentzian from numerical least-squares fits with experimental or theoretical weights (thick and thin black lines, respectively) are offset by -12.5% and $+6.25\%$ due to their systematic errors on D. (b) Residual plots, i.e., same data and fits as shown in panel (a), but divided by the known true expectation value, then unity subtracted. The agreement between data and the ML-fits is practically perfect whereas the least-squares fits clearly underestimate/overestimate the PSD.

Note that for clarity of presentation we discuss the nonaliased cases in the main text and give closed-form aliased results in Appendixes B and C. Numerical tests of the theory are done using the aliased theory in order to utilize all the available data. When fitting a nonaliased expression to aliased data it is necessary to introduce a cutoff frequency $f_{max} \ll f_{Nyq}$ and discard all data above it.

Figure 5(b) shows the same data and fits in a normalized *residual plot*, i.e., data or fits minus the true expected value (the residue) divided by the true expected value. Thus, deviation from zero in this plot shows by how much data and fits differ from the true expected value, measured in units of the true expected value. The data should scatter about 0 with standard deviation $n^{-1/2} = 16^{-1/2} = 25\%$, whereas the fits should simply trace zero for all frequencies. Here, the true expected value is known because we use synthetic data; we normalize by it to show the results of all fits in a single figure. In an experiment, the true expected value is not
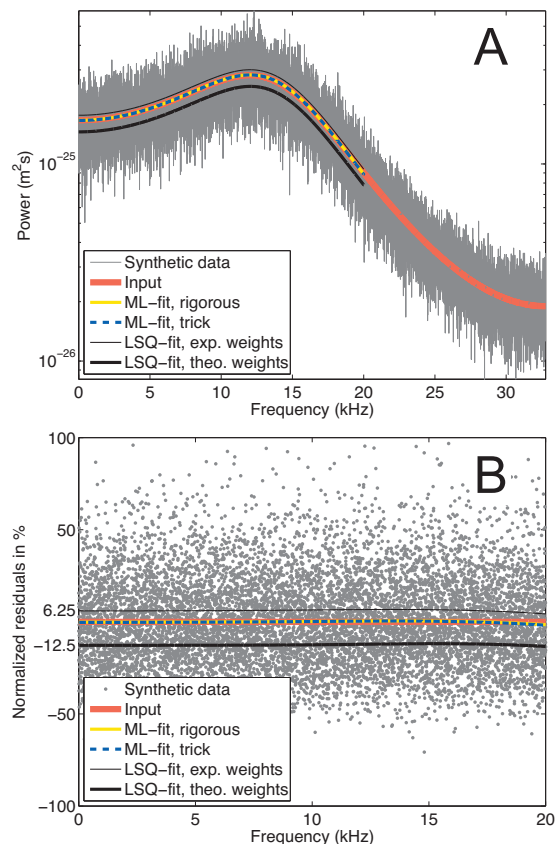

FIG. 6. (Color) Various fits of Eq. (4) to the average of 16 power spectra for an AFM cantilever. Data are synthetic with known parameter values $f_0 = 15$ kHz, $D = 0.0010$ $\mu m^2/s$, and $(2\pi m/\gamma)^2 = 6169$ μs^2, $f_{sample} = 65\,536$ Hz, and $t_{msr} = 8$ s. (a) Data and fits. Note how the three curves corresponding to, respectively, the exactly known parameters (red line), rigorous numerical ML-fit (yellow line), and analytical ML-fit using simplifying trick (dashed blue line) all plot on top of each other. In contrast, the result of a numerical least-squares fits with experimental or theoretical weights (thick and thin black lines, respectively) are offset by -12.5% and $+6.25\%$ due to their systematic errors on D. In all the fits, only frequencies below 20 kHz were included. (b) Residual plots, i.e., same as shown in panel (a), but divided by the exactly known expectation value, then unity subtracted. At the highest frequencies, where the nonaliased fits start to deviate from the aliased PSD, a slight departure from a straight line is seen in all the fits.

known and the fitted value is used instead—when a bias is present the residues will consequently differ from zero in a systematic manner.

This bias always hides in the scatter of the data in a figure like Fig. 5(a) because $n^{-1} \leq n^{-1/2}$ for $n = 1, 2, \ldots$. However, the bias is substantial for small to intermediate values of n and will reveal itself in a plot like Fig. 5(b).

Figure 6 is similar to Fig. 5, except its power spectrum describes the confined Brownian motion of a massive particle, e.g., an AFM cantilever in air, and the fits do not take aliasing into account. The data are synthetic and computer generated (see Appendix A).

C. Error bars on fit parameters

Because we derived closed-form expressions for the fit parameters, we can also calculate expressions for the expected error bars on the fit parameters. We do that and compare to the theoretical limit for how small the error bars can be.

Quite generally, irrespective of whether we study the aliased or nonaliased Lorentzian or the AFM PSD, the error bars on the fit parameters are found by propagating the errors on a, b, and c using the generic formula for a function z (calculating the differential Δz and squaring it),

$$\sigma^2(z[a,b,c]) = \langle (\Delta z)^2 \rangle$$
$$= (\partial_a z)^2 \langle (\Delta a)^2 \rangle + (\partial_b z)^2 \langle (\Delta b)^2 \rangle$$
$$+ (\partial_c z)^2 \langle (\Delta c)^2 \rangle + 2\partial_a z \partial_b z \langle \Delta a \Delta b \rangle$$
$$+ 2\partial_a z \partial_c z \langle \Delta a \Delta c \rangle + 2\partial_b z \partial_c z \langle \Delta b \Delta c \rangle, \quad (73)$$

with $\partial_a z = \partial z / \partial a$, and similarly for b and c; $z = f_c, f_0, D, \ldots$, and $\langle \Delta a \Delta b \rangle$, etc., are elements of the covariance matrix

$$\mathrm{cov}(a,b,c) \equiv \begin{pmatrix} \langle (\Delta a)^2 \rangle & \langle \Delta a \Delta b \rangle & \langle \Delta a \Delta c \rangle \\ \langle \Delta a \Delta b \rangle & \langle (\Delta b)^2 \rangle & \langle \Delta b \Delta c \rangle \\ \langle \Delta a \Delta c \rangle & \langle \Delta b \Delta c \rangle & \langle (\Delta c)^2 \rangle \end{pmatrix}. \quad (74)$$

This covariance matrix can be written in terms of the expectation values for the statistics (see Appendix D),

$$\mathrm{cov}(a,b,c) \approx \mathrm{cov}(\hat{a},\hat{b},\hat{c}) = \frac{1}{N}\frac{n+3}{n}\langle \mathbf{S} \rangle^{-1}, \quad (75)$$

with

$$\langle \mathbf{S} \rangle = \frac{n+1}{n}\widetilde{\mathbf{S}}, \quad (76)$$

where $\widetilde{\mathbf{S}}$ is a matrix of the same form as $\mathbf{S}$ in Eq. (65), but with expectation values $P_f = \langle \bar{P}^{(\mathrm{ex})} \rangle$ replacing the experimental values $\bar{P}^{(\mathrm{ex})}$ in the statistics [Eq. (58); see Eq. (F3)].

To find out how much of the available information we are putting to use, we can express the covariance matrix in terms of the Fisher information matrix $\mathcal{I} = N\widetilde{\mathbf{S}}$ for the full (not the trick) MLE problem,

$$\mathrm{cov}(a,b,c) \approx \mathrm{cov}(\hat{a},\hat{b},\hat{c}) = \frac{n+3}{n+1}\mathcal{I}^{-1}, \quad (77)$$

from which we see that as n increases, we approach the Cramér–Rao bound[16] on the variance for an unbiased estimator, which is just $\mathcal{I}^{-1}$. The main thing to notice, however, is that all the covariances in Eq. (75) are very small because N is very large—the weak n dependence (see Fig. 7) is of no consequence in comparison. In an experiment, simply insert the fitted value for the power spectrum as a best estimate for P_f in $\widetilde{\mathbf{S}}$.

1. Results for OTs

For the nonaliased Lorentzian in Eq. (3) we then have, using the generic equation (73),

$$\sigma^2(f_c) = \frac{f_c^2}{4}\left[\frac{\langle (\Delta a)^2 \rangle}{a^2} + \frac{\langle (\Delta b)^2 \rangle}{b^2} - 2\frac{\langle \Delta a \Delta b \rangle}{ab} \right] \quad (78)$$

and

$$\sigma^2(D) = D^2 \frac{\langle (\Delta b)^2 \rangle}{b^2}, \quad (79)$$

irrespective of the method used to estimate a, b, and c. Actual numbers are found by inserting the estimates $(\hat{a},\hat{b},\hat{c})$.
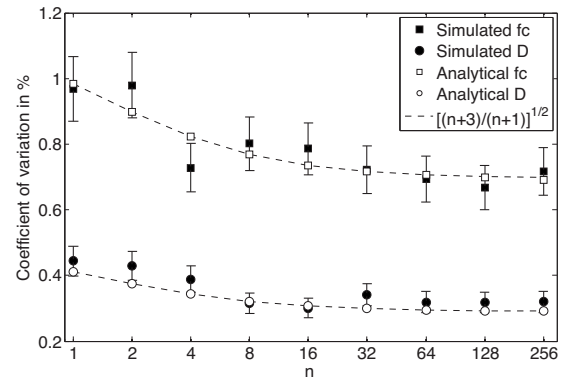

FIG. 7. Stochastic error on fit parameters as a function of number n of PSDs averaged over, for ML-fits of aliased Lorentzians using simplifying trick. Compared to the dependence on the measurement time (Fig. 8), the error is only weakly dependent on n. The sampling frequency was held fixed at $f_{\mathrm{sample}} = 16\,384$ Hz, and the fit was done to the averaged power spectra generated with $N = 262\,144$ ($t_{\mathrm{msr}} = 16$ s), $f_c = 500$ Hz, and $D = 0.46$ $\mu\mathrm{m/s}^2$. Filled symbols show the coefficient of variation $\sigma(X)/\langle X \rangle$, with $X = f_c, D$ determined using Eqs. (C2) and (C3), from 100 independent stochastic simulations. Empty symbols show the theoretical expectation values from Eqs. (E1) and (E2). Dashed lines show the $\sqrt{(n+3)/(n+1)} \in [1:\sqrt{2}]$ scaling predicted from Eqs. (77) and (E5).

The results for the aliased Lorentzian are given in Appendix E.

2. Results for AFM cantilevers

For the AFM, the variance of the three fit parameters are

$$\sigma^2(f_0) = \frac{f_0^2}{16}\left[\frac{\langle (\Delta a)^2 \rangle}{a^2} + \frac{\langle (\Delta c)^2 \rangle}{c^2} - 2\frac{\langle \Delta a \Delta c \rangle}{ac} \right], \quad (80)$$

$$\sigma^2(D) = \frac{D^4}{4\pi^4}[\langle (\Delta a)^2 \rangle f_0^{-4} + \langle (\Delta b)^2 \rangle + \langle (\Delta c)^2 \rangle f_0^4$$
$$+ 2\langle \Delta a \Delta b \rangle f_0^{-2} + 2\langle \Delta a \Delta c \rangle + 2\langle \Delta b \Delta c \rangle f_0^2], \quad (81)$$

$$\sigma^2(G) = \frac{G^4}{c^2}[\langle (\Delta a)^2 \rangle f_0^{-4} + \langle (\Delta b)^2 \rangle + \langle (\Delta c)^2 \rangle (G^{-1} - f_0^2)^2$$
$$+ 2\langle \Delta a \Delta b \rangle f_0^{-2} + 2\langle \Delta a \Delta c \rangle (1 - G^{-1}f_0^{-2})$$
$$+ 2\langle \Delta b \Delta c \rangle (G^{-1} - f_0^2)], \quad (82)$$

where

$$G = (2\pi m/\gamma)^2. \quad (83)$$

3. How good are the fits?

The *goodness of fit*, i.e., the support for the hypothesis that the fitted theory is correct, is[17] the probability that a repetition of the experiment yields a data set with a smaller value for p. A calculation shows that for $K \gg 1$ the support, or backing, is

$$B(s_0) = \mathrm{erfc}(|s_0 - K|\sqrt{n/(2K)}), \quad (84)$$

where erfc is the complementary error function, $s_0 \equiv \Sigma_f \bar{P}_f^{(\mathrm{ex})}/P_f$, and K is the number of terms in the sum s_0. We assumed above that this number is much larger than the number of parameters fitted, hence equal to the number of degrees of freedom. It is of order 10^4–10^6 in our case, while

two to three parameters are fitted. When the sum $s_0 = K$, the expectation value for s_0, the backing is one but it rapidly drops to zero as s_0 becomes larger or smaller than K. The backing calculated for the fits shown in Fig. 5 were, respectively, 0 (LSQ fits; zero within the numerical precision of MATLAB), 0.87 (ML-fit with simplifying trick), and 1.00 (rigorous numerical ML-fit; the first deviation from unity is in the seventh decimal place). These numbers are stochastically varying because the PSD values are. The backing for the rigorous numerical ML-fit will always be close to one because the stationarity conditions [see Eq. (53)] for which the fit parameters are determined are virtually the same as $s_0 = K$.

VII. SUMMARY AND CONCLUSION

(1)   A time series of N coordinate values for an optically trapped microsphere or an AFM cantilever doing Brownian motion gives rise to N power spectral values ($N/2 + 1$ distinct values) with signal-to-noise ratio 1 (Sec. IV). Parameters characterizing the power spectrum can consequently be determined with stochastic errors of order $1/\sqrt{N}$ in the aliased case (Sec. VI C and Appendix E). For the nonaliased case, see also Fig. 2.

(2)   For the purpose of displaying the experimental power spectrum, its signal-to-noise value is reduced by a factor $\sqrt{n}$ by dividing the original data, the time series of N coordinates, into n equally long, nonoverlapping subseries (Sec. IV). From these, n experimental power spectra are calculated and averaged over. This noise-reduced power spectrum covers the same frequency interval, but the separation Δf between consecutive points has increased by a factor n.

(3)   Noise reduction trades resolution on the frequency axis for resolution on the power axis in a manner that loses no information about the parameters characterizing the power spectrum.

(4)   The $1/\sqrt{n}$ scatter of experimental power spectral values should not be confused with the expected stochastic error on parameter values characterizing the power spectrum; the latter are of order $1/\sqrt{N}$ with only a weak n-dependence, as shown in Eq. (75) and Figs. 7 and 8.

(5)   Formulas given above, based on maximum likelihood estimates, eliminate this issue and all further fitting for the cases of the nonaliased and aliased Lorentzian power spectrum and for the nonaliased damped harmonic oscillator as model for an AFM cantilever. Fitting has been done once-and-for-all and the results, including error bars and a goodness-of-fit measure, are given by the formulas in Sec. VI and Appendix C.

(6)   For other systems described by a linear Langevin equation driven by white or colored noise (Sec. IV A), one can determine parameters of the theoretical power spectrum by fitting it to an experimental spectrum, using WLS fitting and experimental or theoretical error bars as weights, which is computationally faster and more robust than MLE. The resulting
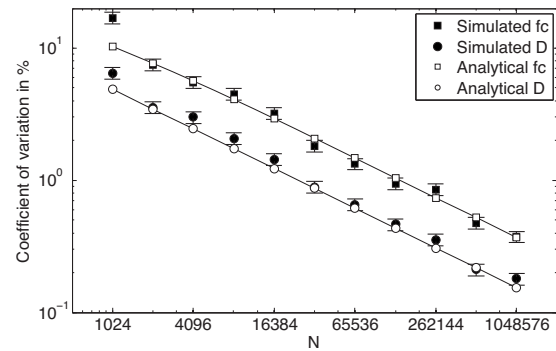

FIG. 8. $1/\sqrt{N}$-dependence of stochastic error on fit parameters for ML-fits of aliased Lorentzians using simplifying trick. An expected $1/\sqrt{N}$ behavior is easily made out on the double-logarithmic scale. The sampling frequency was held fixed at $f_{sample} = 16\,384$ Hz, and the fit was done to the average of $n = 16$ power spectra generated with $f_c = 500$ Hz and $D = 0.46$ $\mu m/s^2$. Filled symbols show the coefficient of variation $\sigma(X)/\langle X \rangle$, with $X = f_c, D$ determined using Eqs. (C2) and (C3), from 100 independent stochastic simulations. Empty symbols connected by lines show the theoretical expectation values from Eqs. (E1) and (E2).

value for the diffusion coefficient D should be corrected as described in Eqs. (32) and (41), respectively.

(7)   Quite generally, beyond optical traps and AFM cantilevers, if the dependent variable (the data) and the squared weights are independent, then WLS yields unbiased results. If not, the least-squares fit is typically biased, but this bias can often be removed altogether by a simple rescaling of the fit results (Sec. V).

(8)   To minimize the stochastic error on fit parameters, we suggest setting $n = 8$ or larger if using ML-fits with our simplifying trick or WLS with theoretical weights [see Eqs. (77) and (E5) and Fig. 7]. If WLS with experimental weights is used, we suggest $n = 16$ or larger (see Fig. 3). Also, for given f_c, n, and N, $f_{sample} = 8f_c$ will minimize the stochastic error (see Fig. 9). Typically, however, t_{msr} and f_c are fixed and f_{sample} should simply be set as high as meaningfully possible as N is the most important factor in increasing the precision (see Fig. 8).

(9)   For OTs, the highest precision is obtained by fitting over as much of the frequency range as can be captured by theory, including modification to the PSD due to hydrodynamics, optics, and instrumentation.[9,10,13] If a certain level of precision is required it then becomes a matter of measuring long enough with the sampling frequency as high as experimentally possible (the minimum in Fig. 9 is for *fixed N*).

(10)   When doing least-squares fitting, we recommend the use of theoretical weights [Eq. (36)] if the standard deviation is known to be proportional to the expected value because this minimizes the stochastic errors on the fit parameters (see Figs. 2 and 3). If any bias is present, that will also be smaller than the bias resulting from experimental weights, and can often be eliminated with the help of Eq. (40) or Eq. (41).
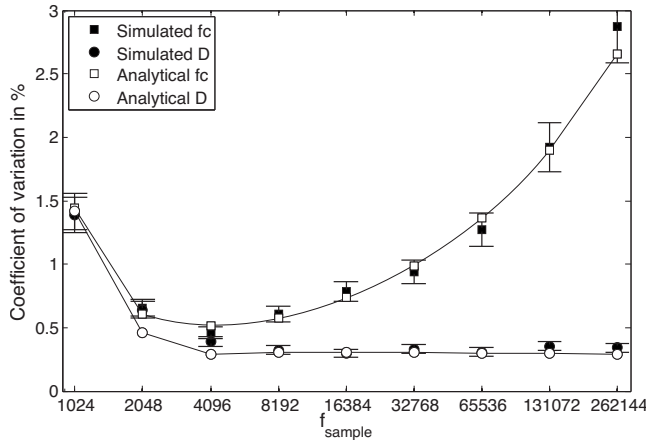
FIG. 9. Stochastic error on fit parameters as a function of sampling frequency f_{sample} for ML-fits of aliased Lorentzians using simplifying trick. A nonmonotonic behavior is seen with minimum around $f_{sample} = 8f_c$. The number of acquired data points was held fixed at $N = 262\,144$, and the fit was done to the average of $n = 16$ power spectra generated with $f_c = 500$ Hz and $D = 0.46$ $\mu m/s^2$. Filled symbols show the coefficient of variation $\sigma(X)/\langle X \rangle$, with $X = f_c, D$ determined using Eqs. (C2) and (C3), from 100 independent stochastic simulations. Empty symbols connected by lines show the theoretical expectation values from Eqs. (E1) and (E2).

ACKNOWLEDGMENTS

APPENDIX A: MONTE CARLO SIMULATION OF THE EINSTEIN–ORNSTEIN–UHLENBECK THEORY OF BROWNIAN MOTION IN A HARMONIC POTENTIAL

In the case of non-negligible mass m, Eq. (1) is rewritten as two coupled first-order equations,

$$\frac{d}{dt}\begin{pmatrix} x(t) \\ v(t) \end{pmatrix} = -\mathbf{M}\begin{pmatrix} x(t) \\ v(t) \end{pmatrix} + (2D)^{1/2}\frac{\gamma}{m}\begin{pmatrix} 0 \\ \eta(t) \end{pmatrix}, \quad (A1)$$

where we introduced the 2×2 matrix

$$\mathbf{M} = \begin{pmatrix} 0 & -1 \\ \dfrac{\kappa}{m} & \dfrac{\gamma}{m} \end{pmatrix}. \quad (A2)$$

Equation (A1) has the solution

$$\begin{pmatrix} x(t) \\ v(t) \end{pmatrix} = (2D)^{1/2}\frac{\gamma}{m}\int_{-\infty}^{t} dt' e^{-\mathbf{M}(t-t')}\begin{pmatrix} 0 \\ \eta(t') \end{pmatrix}, \quad (A3)$$

from which follows that

$$\begin{pmatrix} x_{j+1} \\ v_{j+1} \end{pmatrix} = e^{-\mathbf{M}\Delta t}\begin{pmatrix} x_j \\ v_j \end{pmatrix} + \begin{pmatrix} \Delta x_j \\ \Delta v_j \end{pmatrix}. \quad (A4)$$

Here

$$e^{-\mathbf{M}\Delta t} = \frac{1}{\lambda_+ - \lambda_-}\begin{pmatrix} -\lambda_- c_+ + \lambda_+ c_- & -c_+ + c_- \\ \lambda_+\lambda_-(c_+ - c_-) & \lambda_+ c_+ - \lambda_- c_- \end{pmatrix}, \quad (A5)$$

where

$$\lambda_{\pm} \equiv \frac{\gamma}{2m} \pm \sqrt{\frac{\gamma^2}{4m^2} - \frac{\kappa}{m}}, \quad \mathbf{u}_{\pm} \equiv \begin{pmatrix} \mp 1 \\ \pm\lambda_{\pm} \end{pmatrix} \quad (A6)$$

are the two eigenvalues and corresponding eigenvectors of $\mathbf{M}$,

$$c_{\pm} = \exp(-\lambda_{\pm}\Delta t), \quad (A7)$$

and we introduced the notation

$$\begin{pmatrix} \Delta x_j \\ \Delta v_j \end{pmatrix} \equiv \Delta x_{+,j}\mathbf{u}_+ + \Delta x_{-,j}\mathbf{u}_-, \quad (A8)$$

with

$$\Delta x_{\pm,j} \equiv (2D)^{1/2}\frac{\lambda_+ + \lambda_-}{\lambda_+ - \lambda_-}\int_{t_j}^{t_{j+1}} dt' e^{-\lambda_{\pm}(t_{j+1}-t')}\eta(t'). \quad (A9)$$

From Eq. (2) follows that $\Delta x_{\pm,j}$ are two random lengths drawn from Gaussian distributions with vanishing expectation value and known variances:

$$\langle \Delta x_{\pm,j}\Delta x_{\pm,k} \rangle = \sigma_{\pm}^2 \delta_{j,k},$$

$$\sigma_{\pm}^2 \equiv 2D\left(\frac{\lambda_+ + \lambda_-}{\lambda_+ - \lambda_-}\right)^2 \frac{1 - c_{\pm}^2}{2\lambda_{\pm}}. \quad (A10)$$

It also follows that $\Delta x_{+,j}$ and $\Delta x_{-,j}$ are correlated with each other, but uncorrelated with all $\Delta x_{\pm,k}$ for $j \neq k$

$$\langle \Delta x_{+,j}\Delta x_{-,k} \rangle = \sigma_{+-}^2 \delta_{j,k},$$

$$\sigma_{+-}^2 \equiv 2D\left(\frac{\lambda_+ + \lambda_-}{\lambda_+ - \lambda_-}\right)^2 \frac{1 - c_+ c_-}{\lambda_+ + \lambda_-}. \quad (A11)$$

From their known correlation follows, after some calculation, that they can be expressed in terms of two *uncorrelated* Gaussian distributed random numbers with unit variance, $\eta_j^{(a)}$ and $\eta_j^{(b)}$, as

$$\begin{pmatrix} \Delta x_j \\ \Delta v_j \end{pmatrix} = \left[A_+\begin{pmatrix} -1 \\ \lambda_+ \end{pmatrix} + A_-\begin{pmatrix} 1 \\ -\lambda_- \end{pmatrix} \right](1+\alpha)^{1/2}\eta_j^{(a)}$$
$$+ \left[A_+\begin{pmatrix} -1 \\ \lambda_+ \end{pmatrix} - A_-\begin{pmatrix} 1 \\ -\lambda_- \end{pmatrix} \right](1-\alpha)^{1/2}\eta_j^{(b)}, \quad (A12)$$

where we introduced the notation

$$A_{\pm} = \frac{\lambda_+ + \lambda_-}{\lambda_+ - \lambda_-}\sqrt{\frac{(1 - c_{\pm}^2)D}{2\lambda_{\pm}}} \quad (A13)$$

and

$$\alpha = 2\frac{\sqrt{\lambda_+\lambda_-}}{\lambda_+ + \lambda_-}\frac{1 - c_+ c_-}{\sqrt{(1 - c_+^2)(1 - c_-^2)}}. \quad (A14)$$

So iteration of Eq. (A4) with use of Eq. (A12) generates a time series of positions x_j, which is sampled equidistantly in time with separation Δt from the continuous-time solution to Eq. (1). Since we use the exact analytical solution of Eq. (1) in the generation of this series, the finite value of Δt

causes no discretization error. The only numerical errors associated with our solution are associated with the representation of real numbers on a computer, and, rather hypothetical, with the use of pseudorandom numbers.

APPENDIX B: ALIASED AFM POWER SPECTRUM

For the AFM, using the results from Appendix A, we get for the aliased power spectrum

$$\langle P_k^{(\mathrm{ex})} \rangle = \langle |\hat{x}_k^2| \rangle / t_{\mathrm{msr}} = \frac{\sigma_+^2 \alpha_- + \sigma_-^2 \alpha_+ - 2\sigma_{+-}^2 \alpha_{+-}}{\alpha_+ \alpha_-} \Delta t, \qquad (B1)$$

where

$$\alpha_+ = 1 + c_+^2 - 2c_+ \cos(2\pi k/N), \qquad (B2)$$

$$\alpha_- = 1 + c_-^2 - 2c_- \cos(2\pi k/N), \qquad (B3)$$

$$\alpha_{+-} = 1 + c_+ c_- - (c_+ + c_-)\cos(2\pi k/N), \qquad (B4)$$

and we used that the discrete Fourier transform of $\eta_j^{(a)}$ and $\eta_j^{(b)}$ have the following characteristics:

$$\langle \hat{\eta}_k^{(a)*} \hat{\eta}_l^{(b)} \rangle = 0, \qquad (B5)$$

$$\langle \hat{\eta}_k^{(a)*} \hat{\eta}_l^{(a)} \rangle = \langle \hat{\eta}_k^{(b)*} \hat{\eta}_l^{(b)} \rangle = t_{\mathrm{msr}} \Delta t \delta_{k,l}. \qquad (B6)$$

For completeness, we note that the expression in Eq. (B1) has as limiting expression Eq. (C1) when the mass vanishes,

$$\lim_{m \to 0} \langle P_k^{(\mathrm{ex})} \rangle = \frac{\sigma_-^2}{\alpha_-} \Delta t, \qquad (B7)$$

as is seen by inspection.

APPENDIX C: MLE FOR ALIASED POWER SPECTRA

In an OT experiment, the time series of bead positions $x(t)$ is obtained by sampling the continuous output from the photodiode at discrete times $t_j = j\Delta t$ and $\Delta t = 1/t_{\mathrm{msr}}$. Applying the discrete Fourier transform to $x(t)$ we find[9] that the expectation value for the aliased power spectrum can be written in the form

$$P_k^{\mathrm{alias}} = \frac{1}{A + B \cos(2\pi k/N)}, \qquad (C1)$$

where A and B are related to f_c and D through

$$f_c = \frac{f_{\mathrm{sample}}}{2\pi} u, \qquad (C2)$$

$$D = \frac{f_{\mathrm{sample}}^2}{A \tanh(u)} u, \qquad (C3)$$

$$u = \cosh^{-1}(-A/B). \qquad (C4)$$

By inserting Eq. (C1) in Eq. (53), the stationarity conditions ($\partial_A \mathcal{F} = \partial_B \mathcal{F} = 0$) are seen to be

$$\sum_f \bar{P}_f^{(\mathrm{ex})} = \sum_f P_f, \qquad (C5)$$

$$\sum_f \cos(2\pi k/N) \bar{P}_f^{(\mathrm{ex})} = \sum_f \cos(2\pi k/N) P_f. \qquad (C6)$$

We now repeat the trick introduced in Sec. VI A to turn Eqs. (C5) and (C6) into expressions linear in A and B,

$$\underbrace{\begin{pmatrix} R_{0,2} & R_{1,2} \\ R_{1,2} & R_{2,2} \end{pmatrix}}_{\mathbf{R}} \underbrace{\begin{pmatrix} A \\ B \end{pmatrix}}_{\vec{u}} \approx (1 + 1/n) \underbrace{\begin{pmatrix} R_{0,1} \\ R_{1,1} \end{pmatrix}}_{\vec{r}}, \qquad (C7)$$

which are solved to give

$$A \approx \hat{A} \equiv \frac{n+1}{n} \frac{R_{0,1} R_{2,2} - R_{1,1} R_{1,2}}{R_{0,2} R_{2,2} - R_{1,2}^2}, \qquad (C8)$$

$$B \approx \hat{B} \equiv \frac{n+1}{n} \frac{R_{0,2} R_{1,1} - R_{0,1} R_{1,2}}{R_{0,2} R_{2,2} - R_{1,2}^2}, \qquad (C9)$$

where we introduced the (aliased) statistics

$$R_{p,q} = \frac{1}{K} \sum_k \cos^p(2\pi k/N) \bar{P}_k^{(\mathrm{ex})q}. \qquad (C10)$$

We do not attempt here to give the aliased results for f_0, D, and G from the AFM case. To avoid the aliasing of high frequency noise to the lower frequencies of interest, a high sampling frequency is often used when acquiring AFM data. However, only the region around f_0 is well captured by the dampened harmonic oscillator theory and therefore no more than this region is fitted. Since the aliased expressions only deviate substantially from the nonaliased ones at high frequencies, and because the nonaliased expressions are much simpler, we only treated the nonaliased MLE for the AFM here.

APPENDIX D: COVARIANCE MATRIX

This appendix derives the results that are used in Sec. VI C and Appendix E. To calculate the covariance matrix we look at the response of the estimated fit parameters to fluctuations in the statistics $S_{p,q}$. The calculations go through unchanged for the aliased Lorentzian (see above) with statistics $R_{p,q}$. First, we note that we can write each term in Eq. (65),

$$\mathbf{S}\vec{v} \approx \mathbf{S}\hat{\vec{v}} \equiv \mathbf{S} \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} \equiv (1 + 1/n)\vec{s}, \qquad (D1)$$

as the sum of its "true underlying" value and a fluctuation ($\Delta \mathbf{S}$ and $\Delta \vec{s}$) or a response ($\Delta \vec{v}$) to fluctuations

$$\mathbf{S} = \langle \mathbf{S} \rangle + \Delta \mathbf{S}, \qquad (D2)$$

$$\hat{\vec{v}} = \vec{v} + \Delta \vec{v}, \qquad (D3)$$

$$\vec{s} = \langle \vec{s} \rangle + \Delta \vec{s}, \qquad (D4)$$

where the elements in $\Delta \mathbf{S}$ are

$$\Delta S_{p,q} = \frac{1}{K} \sum_f f^{2p} [\bar{P}^{(\mathrm{ex})q} - \langle \bar{P}^{(\mathrm{ex})q} \rangle]. \qquad (D5)$$

To first order in the fluctuations, we thus have

$$\Delta\vec{v} = \begin{pmatrix} \Delta a \\ \Delta b \\ \Delta c \end{pmatrix} = \langle\mathbf{S}\rangle^{-1}\left(\frac{n+1}{n}\Delta\vec{s} - \Delta\mathbf{S}\vec{v}\right), \tag{D6}$$

where we notice that to first order $\langle\Delta\vec{v}\rangle = 0$ which, however, does not mean that this is an unbiased estimator as shown below. The covariance matrix $\langle\Delta\vec{v}\otimes\Delta\vec{v}\rangle$ as given in Eq. (74) then follows after some calculation using Eqs. (49) and (15).

We emphasize here that the above calculations were to first order in $1/\sqrt{K}$, with K as the number of terms in the statistics $S_{p,q}$. Whereas the relative size of an individual fluctuation in the power spectral value $\Delta P^{(\text{ex})}$ is independent of K, the relative sizes of the overall fluctuations in the sums $\Delta S_{p,q}$ go to zero as $1/\sqrt{K}$. Since K is typically of the order of 10^4–10^6, this first-order approximation is very good.

APPENDIX E: ERROR BARS FOR THE ALIASED LORENTZIAN

The error bars on f_c and D are calculated as before, using Eq. (73) for the variance,

$$\sigma^2(f_c) = \frac{f_{\text{sample}}^2 A^2}{4\pi^2(A^2 - B^2)}\left[\frac{\langle(\Delta A)^2\rangle}{A^2} + \frac{\langle(\Delta B)^2\rangle}{B^2} - 2\frac{\langle\Delta A\Delta B\rangle}{AB}\right] \tag{E1}$$

and

$$\sigma^2(D) = (\partial_A D)^2\langle(\Delta A)^2\rangle + (\partial_B D)^2\langle(\Delta B)^2\rangle$$
$$+ 2\partial_A D\partial_B D\langle\Delta A\Delta B\rangle, \tag{E2}$$

where

$$\partial_A D = \frac{D}{A}\left(\frac{B^2}{A^2-B^2} - 1 - \frac{A}{u\sqrt{A^2-B^2}}\right), \tag{E3}$$

$$\partial_B D = \frac{D}{B}\left(\frac{A}{u\sqrt{A^2-B^2}} - \frac{B^2}{A^2-B^2}\right), \tag{E4}$$

and u is given in Eq. (C4). The covariance matrix is calculated as before, giving

$$\text{cov}(A,B) \equiv \begin{pmatrix} \langle(\Delta A)^2\rangle & \langle\Delta A\Delta B\rangle \\ \langle\Delta A\Delta B\rangle & \langle(\Delta B)^2\rangle \end{pmatrix} \approx \frac{1}{N}\frac{n+3}{n+1}\widetilde{\mathbf{R}}^{-1}, \tag{E5}$$

where $\widetilde{\mathbf{R}}$ is a matrix with the same structure as $\mathbf{R}$ [see Eq. (C7)] but with the experimental values $\bar{P}^{(\text{ex})}$ replaced by the theoretical (in practice the fitted) values P_f in all the statistics [Eq. (C10)].

We tested the above analytical results by comparing to the results of simulations: Multiple independent position time series for a massless particle diffusing in a harmonic potential were created using the methods given in Ref. 9. The resulting power spectra were fitted using Eqs. (C2)–(C10). Since we simulate a stochastic process there is scatter in the fitted parameters and it is this scatter that we compare to the results given in Eqs. (E1) and (E2). The results are shown in Figs. 7 and 8. Figure 8 shows the expected $1/\sqrt{N}$ scaling, whereas Fig. 7 shows a $\sqrt{n+3}/\sqrt{n+1}$ scaling.

Figure 9 shows that the optimal tradeoff between precision and amount of data acquired seems to manifest itself at a sampling frequency roughly eight times the corner fre-

quency; any slower than this leads to large errors in both parameters because the PSD essentially reduces to the ratio of D to f_c, i.e., two parameters are used to fit a single constant. Sampling much faster than f_c, but keeping the number of acquired data points fixed, has no effect on the error on D but is detrimental for f_c since there is progressively less information about f_c at larger frequencies. The effect on the precision of increasing n is small but positive; compared to the effect of N and f_{sample} it can be ignored [after including it as described in Eqs. (C8) and (C9)].

For comparison, the error bars on the fit parameters, as a function of cutoff frequency f_{max}, from a least-squares fit are shown in Fig. 2. The average of $n=16$ synthetic PSDs was fitted by minimizing Eq. (23) with the data points weighted by the standard deviation of the n PSDs. Also shown is a LSQ fit where the weights are kept constant; this is the kind of fitting performed by primitive LSQ routines. For a detailed discussion of how the stochastic error depends on the fitting range $[f_{\text{min}}:f_{\text{max}}]$, the reader is referred to Sec. VIII in Ref. 9.

APPENDIX F: BIAS OF THE MLES

Obviously, we must be paying a price somewhere for turning a nonlinear problem into a linear one with our little trick, or else we would have turned a nonlinear problem into an exactly solvable mathematical problem. That is sometimes done, but not here. The trick works through an approximation, and the resulting approximate estimator is biased, which means that its expectation value is different from the true value of the quantity it estimates. So on average it misses the correct result. Bias is systematic error on averages. That is the *nature* of the price we pay. Fortunately, it is negligible in size, as we demonstrate now.

For the nonaliased Lorentzian and AFM we find, by expanding Eq. (D1) to first order in $\Delta\vec{v}$ and *second* order in $\Delta S_{p,q}$,

$$\langle\Delta\vec{v}\rangle = \begin{pmatrix} \langle\Delta a\rangle \\ \langle\Delta b\rangle \\ \langle\Delta c\rangle \end{pmatrix}$$
$$= \langle\mathbf{S}\rangle^{-1}\left[\langle\Delta\mathbf{S}\langle\mathbf{S}\rangle^{-1}\Delta\mathbf{S}\rangle\vec{v} - \frac{n+1}{n}\langle\Delta\mathbf{S}\langle\mathbf{S}\rangle^{-1}\Delta\vec{s}\rangle\right]. \tag{F1}$$

For the nonaliased Lorentzian, this expression can be reduced to

$$\begin{pmatrix} \langle\Delta a\rangle \\ \langle\Delta b\rangle \end{pmatrix} = \frac{2}{N}\frac{n+2}{n+1}(\widetilde{S}_{0,2}\widetilde{S}_{2,2} - \widetilde{S}_{1,2}^2)^{-2}\begin{pmatrix} \widetilde{S}_{2,2} & -\widetilde{S}_{1,2} \\ -\widetilde{S}_{1,2} & \widetilde{S}_{0,2} \end{pmatrix}$$
$$\times\begin{pmatrix} \widetilde{S}_{2,3} & -2\widetilde{S}_{1,3} & \widetilde{S}_{0,3} \\ \widetilde{S}_{3,3} & -2\widetilde{S}_{2,3} & \widetilde{S}_{1,3} \end{pmatrix}\begin{pmatrix} \widetilde{S}_{0,2} \\ \widetilde{S}_{1,2} \\ \widetilde{S}_{2,2} \end{pmatrix}, \tag{F2}$$
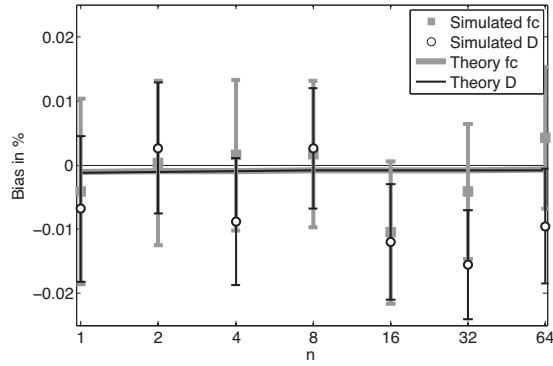
where

FIG. 10. Bias of the parameter estimates for f_c (gray squares) and D (white circles), from ML-fits with our trick, of the aliased Lorentzian. Error bars are standard errors on the mean. The point of this figure is only to show that the bias is indeed very small and can be completely ignored. Theoretical expectation value for the bias as given in Eq. (F6) (gray line) and Eq. (F7) (black line) are less than 1/1000 of a percent for these settings. Bias was measured as the difference between the average of 1000 determinations of the fit parameters and the known input values $f_c=500$ Hz and $D=0.46 \ \mu\text{m/s}^2$. Simulations were run with $N=2^{20}$ and $f_{\text{sample}}=2^{12}$ Hz—values chosen to minimize the stochastic errors. Data were treated with n nonoverlapping Hann windows before calculation of the PSD.

$$\widetilde{S}_{p,q} = \frac{1}{K} \sum_f f^{2p} \langle \overline{P}^{(\text{ex})} \rangle^q = \frac{1}{K} \sum_f f^{2p} P_f^q. \tag{F3}$$

That is, the bias is proportional to $1/N$ which is a very small number and displays a weak n-dependence. From these expressions we find the bias on f_c and D to be

$$\langle \Delta f_c \rangle = \frac{f_c}{2} \left(\frac{\langle \Delta a \rangle}{a} - \frac{\langle \Delta b \rangle}{b} \right), \tag{F4}$$

$$\langle \Delta D \rangle = -D \frac{\langle \Delta b \rangle}{b}. \tag{F5}$$

We only know the true values of f_c and D in simulations, in an experiment the best estimate of the true value would be the fitted value.

For the aliased Lorentzian we find in an analogous manner

$$\langle \Delta f_c \rangle = \frac{f_{\text{sample}}}{2\pi} \frac{A}{\sqrt{A^2 - B^2}} \left[\frac{\langle \Delta B \rangle}{B} - \frac{\langle \Delta A \rangle}{A} \right], \tag{F6}$$

$$\langle \Delta D \rangle = -D \left[\left(\frac{B^2}{A^2 - B^2} + \frac{A}{\sqrt{A^2 - B^2}} \right) \left(\frac{\langle \Delta A \rangle}{A} + \frac{\langle \Delta B \rangle}{B} \right) - \frac{\langle \Delta A \rangle}{A} \right], \tag{F7}$$

where $\langle \Delta A \rangle$ and $\langle \Delta B \rangle$ are found from Eq. (F1) by everywhere replacing $\widetilde{S}$ with $\widetilde{R}$. The result of a numerical test of the above relations is shown in Fig. 10.

[1] K. C. Neuman and S. M. Block, Rev. Sci. Instrum. **75**, 2787 (2004).
[2] K. C. Neuman, T. Lionnet, and J.-F. Allemand, Annu. Rev. Mater. Res. **37**, 33 (2007).
[3] K. C. Neuman and A. Nagy, Nat. Methods **5**, 491 (2008).
[4] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante, Annu. Rev. Biochem. **77**, 205 (2008).
[5] T. T. Perkins, Laser Photonics Rev. **3**, 203 (2009).
[6] J. L. Hutter and J. Bechhoefer, Rev. Sci. Instrum. **64**, 1868 (1993).
[7] D. Walters, J. P. Cleveland, N. H. Thomson, P. K. Hansma, M. A. Wendman, G. Gurley, and V. Elings, Rev. Sci. Instrum. **67**, 3583 (1996).
[8] J. Sader, J. Appl. Phys. **84**, 64 (1998).
[9] K. Berg-Sørensen and H. Flyvbjerg, Rev. Sci. Instrum. **75**, 594 (2004).
[10] S. F. Tolić-Nørrelykke, E. Schäffer, J. Howard, F. S. Pavone, F. Jülicher, and H. Flyvbjerg, Rev. Sci. Instrum. **77**, 103101 (2006).
[11] K. Berg-Sørensen, L. Oddershede, E.-L. Florin, and H. Flyvbjerg, J. Appl. Phys. **93**, 3167 (2003).
[12] K. Berg-Sørensen, E. J. G. Peterman, T. Weber, C. F. Schmidt, and H. Flyvbjerg, Rev. Sci. Instrum. **77**, 063106 (2006).
[13] E. Schäffer, S. F. Nørrelykke, and J. Howard, Langmuir **23**, 3654 (2007).
[14] Here and below, we used the following results that are easy to verify by direct calculation or by consulting a statistics text book. If two independent stochastic variables are normally distributed $X, Y \sim \mathcal{N}(0, \sigma^2/2)$, then their individual squares are gamma distributed $X^2, Y^2 \sim \Gamma[(1/2), (1/\sigma^2)]$, and the sum of their squares is exponentially distributed $Z=X^2+Y^2 \sim \Gamma[1, (1/\sigma^2)] = E[(1/\sigma^2)]$. Finally, averaging over n independent, exponentially distributed variables returns a gamma distributed variable $(1/n)\Sigma_{i=1}^n Z_i \sim \Gamma[n, (n/\sigma^2)]$.
[15] A. C. Aitken, Proc. R. Soc. Edinburgh **55**, 42 (1935).
[16] C. R. Rao, *Linear Statistical Inference and Its Applications* (Wiley, New York, 1973).
[17] N. C. Barford, *Experimental Measurements: Precision, Error and Truth*, 2nd ed. (Wiley, New York, 1990).