Supporting Information (SI): Quantifying transient 3D dynamical phenomena of single mRNA particles in live yeast cell measurements

C.P. Calderon, M.A. Thompson, J.M. Casolari, R.C. Paffenroth, and W.E. Moerner

Contents

1. ADDITIONAL EXPERIMENTAL DATA RESULTS

In the main text, various estimation strategies were used to extract the diffusive and localization noise of experimental trajectories. In this section, we provide additional results to illustrate various technical points (the subsection titles highlight the themes studied). This section focuses exclusively on analyzing experimental data (Sec. 3 provides additional control simulation results).

Corresponding Author(s) E-mail: chris.calderon@numerica.us,wmoerner@stanford.edu

1.1 Directed Motion Results

The results in Figs. S1 and S2 are representative of *ARG*3 mRNA trajectories where the number of rejections implied by the goodness-of-fit tests used to check the consistency of the assumed classic SPT models with experimental data was unusually large (implying rejection or inconsistency of experimental data with fitted models) along a single trajectory (see main text).

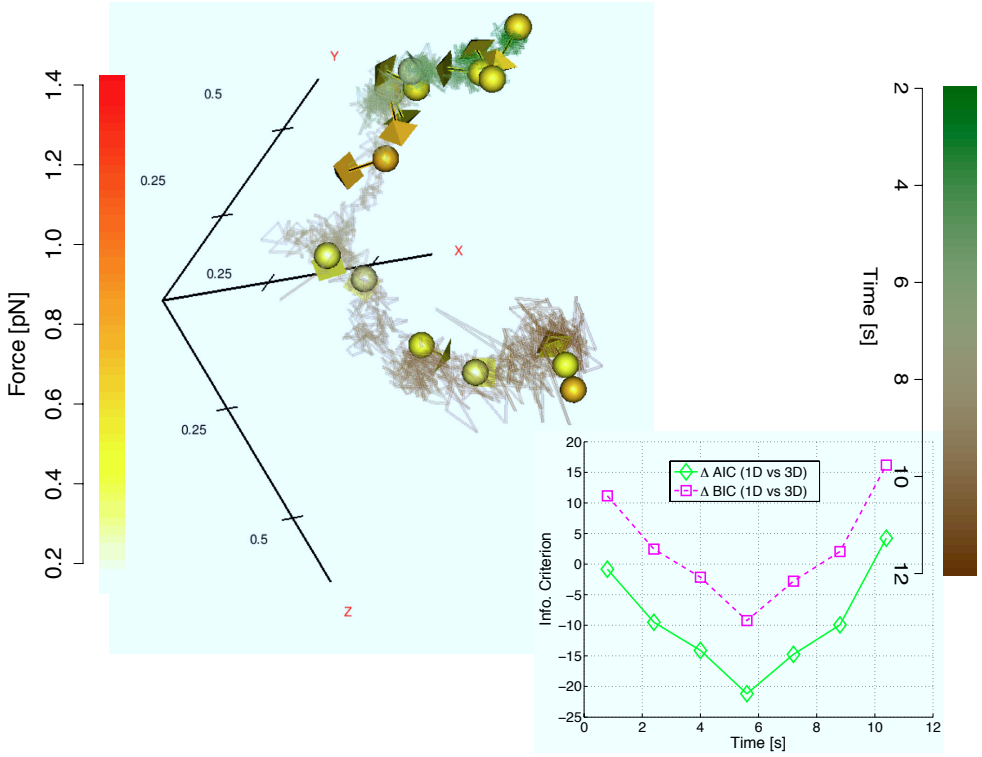

Fig. S 1. An *ARG*3 trajectory exhibiting statistically significant 3D forces. The raw measurements obtained from the microscope are show as a solid curve. The rich geometrical structure in raw data was only noticed after the goodness-of-fit tests flagged this trajectory as not being consistent with classic SPT models. The 3D forces at different times are denoted by colored arrows (color and length scaling dictated by the inferred magnitude). This trajectory also exhibited "tethering" forces for a transient time. The inset plots the difference between the AIC and BIC score obtained for a 3D and 1D model (values below zero suggest that the added complexity associated with estimating a 3D model is justified given the data in a local time window).

1.2 Confined Diffusion Results

As stated in the Methods of the main text, 290/1419 of the trajectories were rejected by the M-test. However, it should be noted that the goodness-of-fit tests used can also be used in different contexts (note that the goodness-of-fit testing techniques employed are discussed in greater detail in SI Sections 2.3-2.5). Figure 3 presents a situation where the tracked mRNA was in the bud of a yeast cell. In this small and highly crowded region of the cell, a researcher may be tempted to claim that the diffusion is "anomalous". The M-test applied to a sequence of three 1D SDE models (the x, y, z components were assume to evolve independently) suggests that there is not enough evidence in the data to reject to simple 1D SDEs. So despite the apparent confined nature, an SDE accounting for 2D or 3D constraint forces is not required to produce a model not rejected by goodness-of-fit testing.

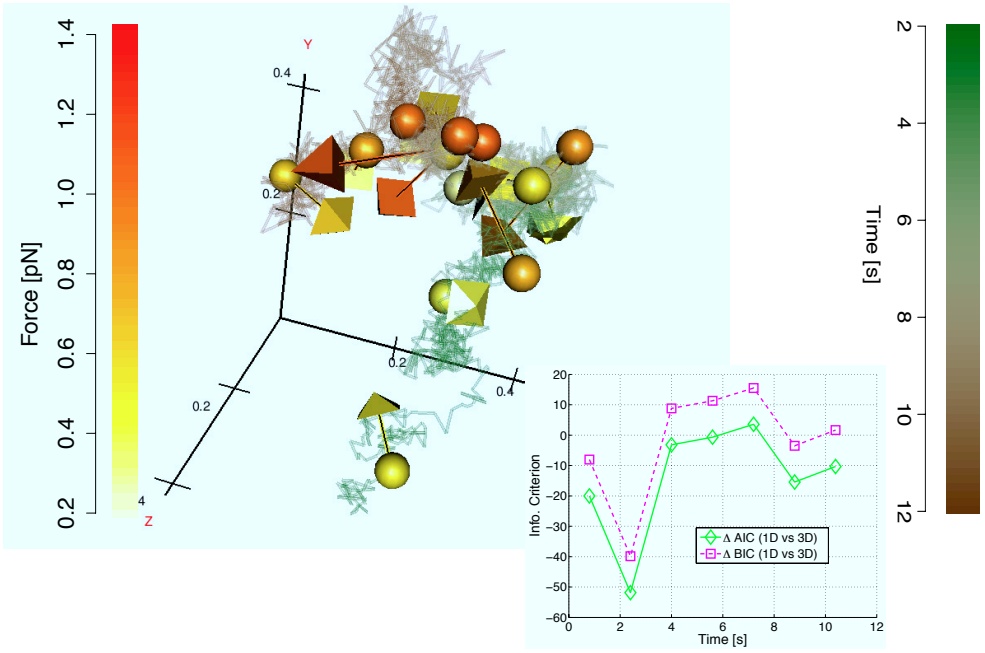

Fig. S 2. Another *ARG*3 trajectory exhibiting transient (but statistically significant) 3D forces. See previous caption for details for additional details.

1.3 Statistically Quantifying Problems with Using Fixed Motion Parameters

Table S1 displays some dangers of not accounting for time-varying parameters in classic SPT analysis methods. If MSD-based[1] and MLE techniques (aiming to achieve optimal estimation[2]) are applied, one can certainly extract some estimate of the diffusion coefficient and measurement noise. The problem is that the statistics of the increments can change over time, so these estimates are of dubious quality. MSD methods are only concerned with the time between observations (so increments observed at vastly different times can get aggregated together in the MSD curve). To apply the hypothesis testing procedure used here,[3] we plugged the parameter estimates obtained from each scheme into the analytically known likelihood used in Ref.[2] to evaluate the corresponding Q-test statistics. This procedure strongly suggested rejection of a model assuming constant (i.e., "time-frozen") parameters; the $p-values$ associated with the test statistics presented are all $< 10^{-6}$.

Table S 1. Diffusion Coefficient and Measurement Noise Estimates (MSD vs. MLE) along with goodness-of-fit for the x,y, and z components of the mRNA trajectory shown in Fig. 1 of the main text. For the MSD analysis, the procedure outlined in Ref.[1] was used. The code associated with Ref.[2] was used to provide the MLE estimates (note: the "motion blur coefficient" was set to zero in the results reported; since this parameter is not statistically identifiable from observed time series data alone). Goodness of fit reported utilizes the exact likelihood and the Q test-statistic[3] (e.g., Q_x denotes goodness-of-fit result for x).

Estimator	$D_x[\mu m^2/s]$	$R_x[nm]$	Q_x	$D_y[\mu m^2/s]$	$R_y[nm]$	Q_y	$D_z[\mu m^2/s]$	$R_z[nm]$	Q_z
MSD	0.015	23.319	135.092	0.009	20.737	104.895	0.003	17.516	31.168
MLE	0.015	12.467	6.044	0.014	11.318	9.118	0.006	13.710	9.398

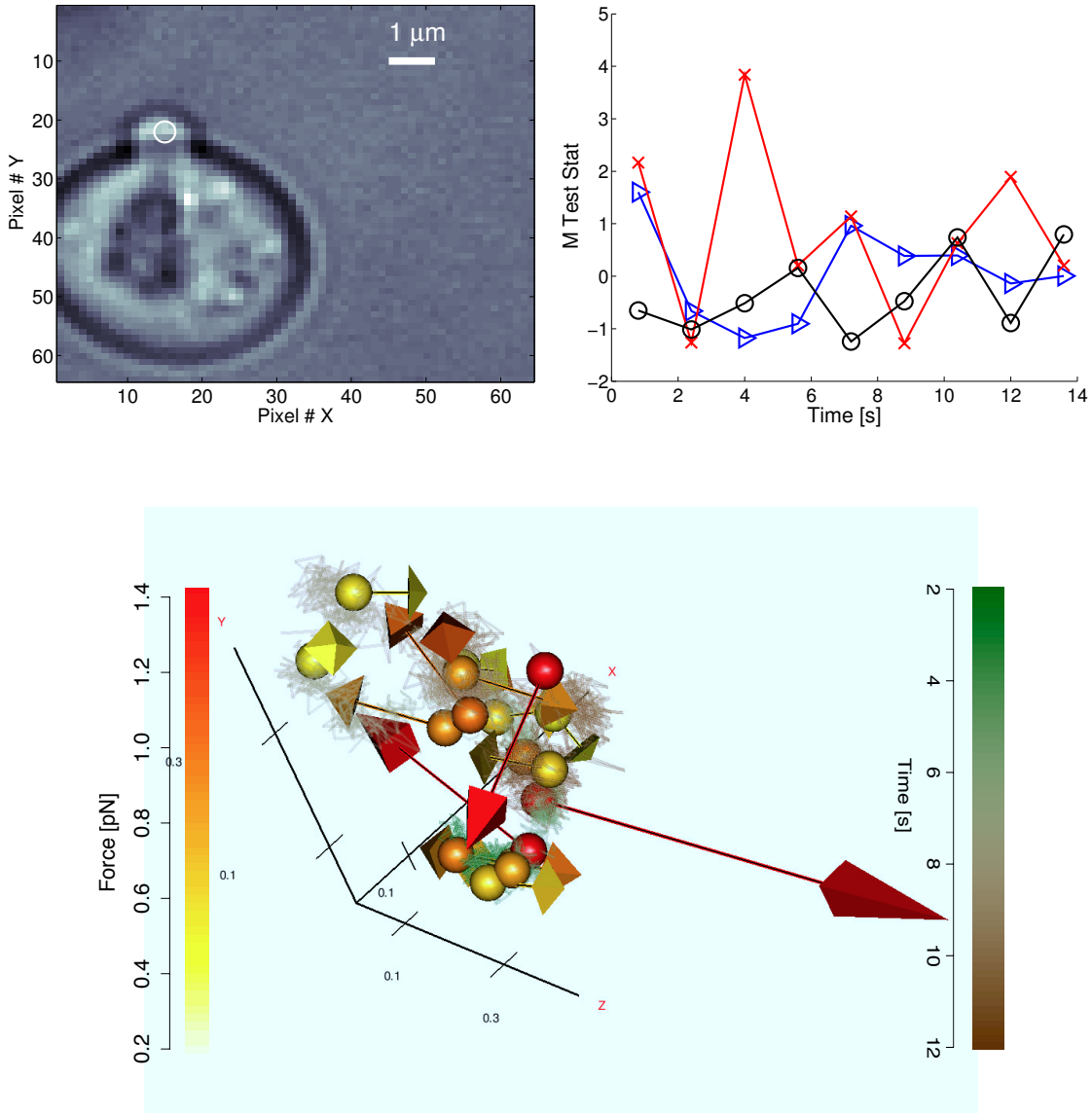

Fig. S 3. Goodness of fit test to a spatially confined mRNA. Top left: white light image of yeast cell (tagged mRNA circled). Top right: M(1,1) test statistic computed over 9 time windows. For this trajectory, three 1D SDE models (i.e. each window of the 3D measurement was used to generate 3 1D SDEs) performed reasonably in terms of the goodness-of-fit tests used (blue denotes x results, red denotes y, and black denotes z). Bottom: The 3D measurements with inferred force vectors superimposed.

1.4 Heterogeneity and Time Changing Effective Kinetic Parameters

1.4.1 Heterogeneity in the Effective Measurement Noise

Figure S4 displays the estimated effective measurement noise associated with z (i.e.,$R_z^{\frac{1}{2}}$) vs. time for seven different trajectories. Note how both the general shape (not just slope and intercept) vary between different trajectories. The non-trivial trend in effective measurement noise is due in part to trajectory specific background

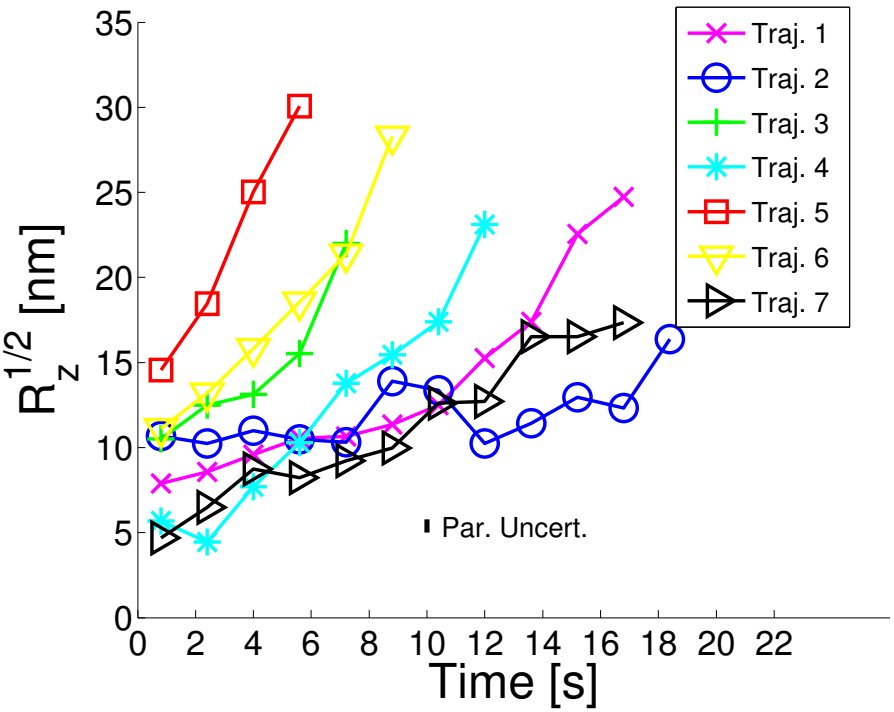

Fig. S 4. The estimated measurement uncertainty in z vs. time for seven different trajectories. The lengths of the trajectories differ due to the different stochastic photobleaching times associated with each measurement. An estimate of the parameter uncertainty is shown as a vertical bar.

fluorescence, differing numbers of active emitters, and other phenomena known to complicate photon counting.[1,4] Although accurately quantifying effective measurement noise may not be of primary interest in SPT studies, it is required for calibrating statistically acceptable models (using the average measurement does not suffice as shown next in experimental results in Table S1 and by simulation results shown in SI Section 3).

1.4.2 Time Changing Kinetic Parameters

In the results reported in Table S1 (discussed in Sec. 1.3), the parameters were assumed frozen for the entire duration of the trajectory. Here we demonstrate that dividing the trajectory into multiple segments, each segment contained 400 observations (the entire trajectory had over 3200 temporal observations), does not mitigate the problem. That is, substantial differences exist between the MLE approach advocated in the main text and MSD based approaches (see Fig. S5). Since multiple GFPs are attached to the mRNA particle,[1] it makes sense that the effective measurement noise increases over time (less photons can be collected per unit time as various GFP dyes randomly photobleach resulting in increased position uncertainty); the MLE approach shows this trend in the localization precision whereas the MSD approach does not. The noisy and biased nature of the estimated effective measurement noise (i.e., localization precision) is due to artifacts associated with suboptimal "time-lag" selection.[2]

1.5 Exploring Sensitivity to Sampling Parameters

In this section, we demonstrate how parameters reported in the main text are changed when various sampling parameters are altered.

1.5.1 Varying Local Window Size W and Sampling Rate Δt

Figure S6 corresponds to in Figs. 1-2 of the main text.

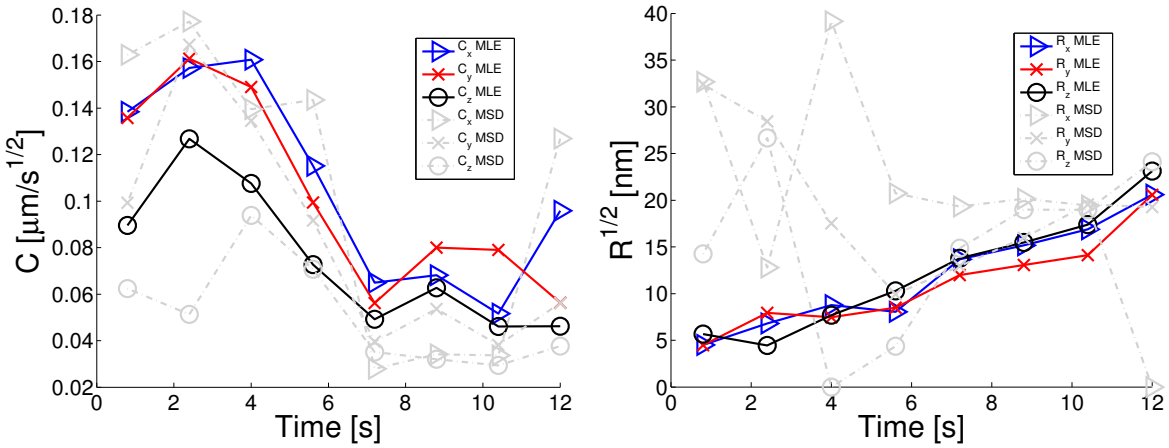

Fig. S 5. Windowed diffusion and measurement noise (for each spatial dimension) vs. time for the MSD estimator discussed in Ref.[1] and the MLE approach used in the main text. The window size $W = 400$ and the underlying trajectory is shown in Fig. 1 of the main text.

1.5.2 Predicted Force Dependence on Window Size W

Figure S7 corresponds to in Figs. 3-4 of the main text. Note that the raw measurements coming from the microscope are corrupted by noise; the force magnitude we plot utilizes the Kalman filter's estimate[6,7] of the unobservable state r (then the magnitude of $F(r)$ is computed).

2. SUPPORTING THEORY AND METHODS

2.1 Mapping Continuous SDEs to Discrete Kalman Filter Models

Here we outline several procedure for extracting the transition density from the assumed continuous time over-damped Langevin equation considered. The Kalman filter equations corresponding to the discretely observed SDE considered throughout this work are as follows:

$$r_{i+1} = A' + B' r_i + q_{i+1} \tag{1}$$
$$\psi_i = r_i + \epsilon_i \tag{2}$$
$$q_{i+1} \sim \mathcal{N}(0, Q_{\Delta t_i}) \tag{3}$$
$$\epsilon_t \sim \mathcal{N}(0, R_i) \tag{4}$$

where we define $\Delta t_i := t_{i+1} - t_i$, $B' := \exp(\Phi B \Delta t_i)$ and $A' := -B^{-1}A + \exp(\Phi B \Delta t_i) B^{-1} A$. The primes are used to indicate the discrete sample analogs that correspond to the continuous time SDE model.[7] The covariance matrix R_i characterizes the mean zero Gaussian noise associated with the measurement ψ_i gathered at discrete time i. The process noise covariance, $Q_{\Delta t_i}$, associated with the discretely sampled state can be computed from the corresponding continuous SDE. The process noise covariance of a discretely observed SDE in the general (non-stationary or stationary) multidimensional linear case is given by:[8]

$$Q_{\Delta t_i} := \int_0^{\Delta t_i} 2 \exp(\Phi B s) C C^{\mathrm{T}} \exp((\Phi B)^{\mathrm{T}} s) ds \tag{5}$$

where the subscript on Q is used to emphasize the dependence on the time between successive observations. Exponentials with matrix arguments imply matrix exponentials. The above integral can be solved exactly by using established tools from control theory.[7,9,10] We have written equations for the case where data is spaced Δt_i time units apart, hence arbitrary time spacing case can be considered (as would be needed with blinking

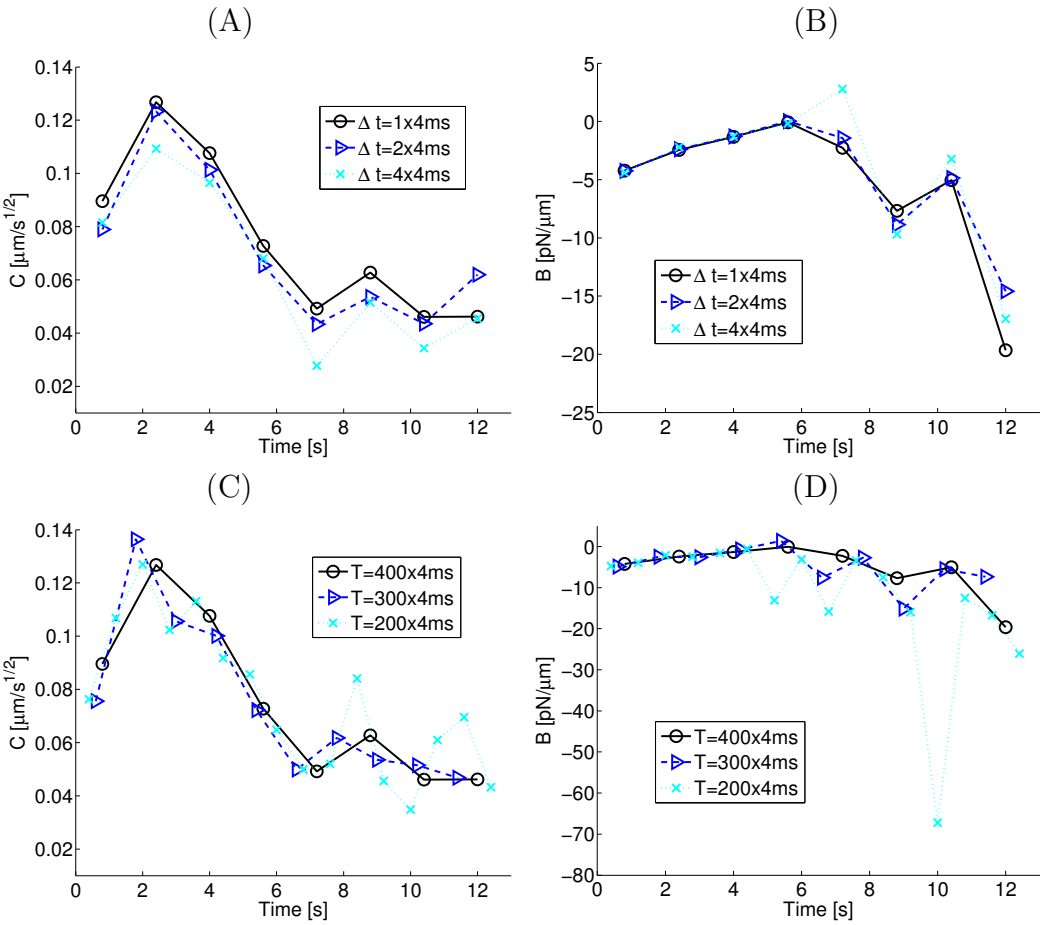

Fig. S 6. Illustration of SDE parameter sensitivity on "time lag" (or "subsampling") and window size. In all cases, the z component of the local diffusion coefficient and the "spring constant" (x, y, and z were plotted in Fig. 2 of the main text; z components were selected because it showed interesting trends and had the largest experimental measurement noise; x and y component results were similar). Panels A-B plot results obtained using 3 different "subsampling" parameters. The term "subsampling" is used here to convey that the microscope generates images every 4ms at the highest sampling frequency of the data, but we can skip every entry and observe images every 8ms, 12ms, etc. (however the net time spanned by a local window used for parameter estimation is constant and independent of the subsampling rate). In Panels C-D the same estimated quantities are plotted, but the time of observation spanned by local windows is varied (the window size ranges from 200-400 observations; however, each trajectory was observed at the finest temporal sampling frequency of 4ms). Since the temporal sampling frequency is fixed in Panels C-D, changing the window size only decreases the number of observations that can be used for estimation and inference (the "spring constant" is the hardest parameter to estimate in terms of asymptotic parameter uncertainty in many situations,[5] so decreasing the number of samples for parameter estimation affects this quantity the most as can be observed in Panel D). In all panels, one can see that the estimated SDE parameters are relatively insensitive to the sampling parameters selected; the changes induced by "heterogeneity" in the cell are large enough in magnitude to dominate the effects of bias and variance induced by or estimation scheme.

quantum dots). In the main text and results reported ,Δt_i was constant. The quantities above can be used for both filtering (inferring the unobservable r_t given ψ_t) or for estimating unknown parameters. One can simply plug the above expression into the equations provided in Chapter 13 of Hamilton for either filtering or estimation.[6]

Note that *if* the linear model above is consistent with the model and the initial conditions are precisely known, then the unobservable state is a multivariate Gaussian process. This idealized process is completely (statistically) characterized by its first two moments.[6,8] So if the measurements are also Gaussian (all assumptions stated here will later statistically tested against data), then computing the likelihood of the observations, ψ_t only requires the mean and covariance of a sequence of Gaussian random variables[6,7] often called the "innovations".[11]

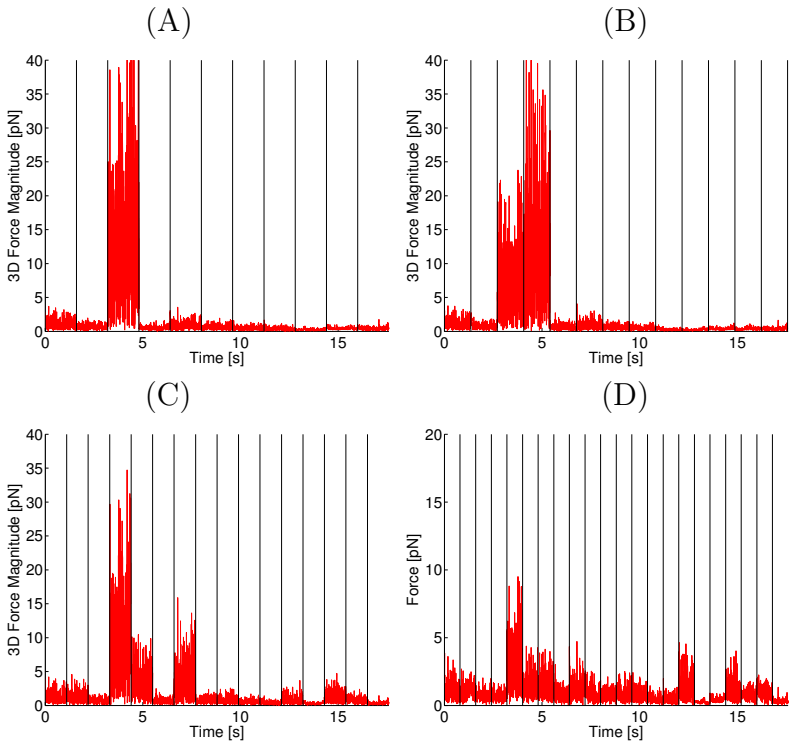

Fig. S 7. 3D force magnitude (implied by the estimated SDE model) vs. time for the trajectory shown in Figs. 3-4 of the main text. Different W window sizes were used for the fit (400,338,275,200); these cases are plotted in Panels A-D, respectively. The different time windows are indicated by solid vertical lines. All models predict a large force around time 3-6 seconds (this is phenomenon relatively independent of W, though the magnitude of the forces does change slightly). The tagged particle would likely be considered static or perhaps confined via traditional analysis in the $\approx$ 3-6 second time window. However, one can easily reject the "static particle" hypothesis *and* one needs to include 3D interactions in order to generate a diffusion model that is not rejected by the data (the SDE models considered include the case where one is only measuring apparatus noise and no dynamics).

Computing the innovations requires iterating the Kalman filter.[6,7,11] After $Q_{\Delta t_i}$ is computed, this is fairly straightforward.[6,7,11]

The above discussion is for general linear equations. However, to illustrate tools familiar to researchers in statistical physics, we focus on a special case where the eigenvalues of B are all real and negative (a case implicitly assumed by common models such as a corralled diffusion[12]). The mean expression is unchanged in this special case, but the covariance computation simplifies if one uses the analytical stationary distribution (or "steady state") formulas from statistical physics.[13] Knowledge of the stationary distribution allows the following compact expression for the conditional (multivariate normal) distribution of the underlying state:[14]

$$r_i \sim \mathcal{N}(\mu_i, Q_{\Delta t_i}) \tag{6}$$

$$\mu_i = -B^{-1}A + \exp(\Phi B \Delta t_i)(r_{i-1} + B^{-1}A) \tag{7}$$

$$Q_{\Delta t_i} = \Lambda - \exp\left(\Phi B \Delta t_i\right)\Lambda \exp\left((\Phi B)^{\mathrm{T}} \Delta t_i\right) \tag{8}$$

$$\Delta t_i = t_i - t_{i-1} \tag{9}$$

where Λ is the covariance characterizing the Gaussian stationary density associated with a given "stable" linear SDE. Expressions for Λ under various models/conditions can be found in textbooks, e.g. Risken;[13] alternatively, the solution for more general linear models can be obtained by solving for the steady state solution of the

Riccati matrix equations where terms involving the measurement noise are set to zero.[7] The utility of the above expression is that the conditional mean and variance of the state can be utilized by the Kalman filter to compute the likelihood function of the innovation sequence. The Kalman filter can provide $p(\psi_{i+1}|\psi_i;\theta)$ using the mapping between continuous SDEs and discrete Gaussian time series shown here; see Hamilton[6] Chapter 13 for details.

2.2 Generating Initial Conditions for a 3D SDE MLE Search

In principle, the MLE parameters can be found by simply optimizing the likelihood cost function stated in the methods. However, it is often both practically and theoretically useful to have a procedure for systematically generating reasonable initial guesses to the parameters before one attempts a non-linear MLE search.[15] One computational complication encountered when estimating an SDE is that one can get stuck in a local minimum of the likelihood function that does not correspond to the MLE. Fortunately, the goodness-of-fit tests discussed in later sections can help identify this problem. Identifying a poor fit is helpful but rejection can occur either because one is stuck in a local minimum or because the assumed model (with the global MLE parameter computed exactly) is inconsistent with the observations. If one has a procedure for generating reasonable initial guesses to the model, and if one rejects the model given the data, then one can be more confident the rejection is due to poor modeling assumptions (and not being stuck in a local minima). The next two subsections outline our procedure for generating initial conditions for the overdamped Langevin SDE discussed here and in the main text. It should be noted, that within each local window, we subtract the time averaged measurement vector from all observation in that window (i.e. the resulting data analyzed has an empirical mean exactly equal to zero). This was done only to simplify interpretation of the parameters. For example, when the mean is subtracted, the parameter A corresponds to the instantaneous force at the empirical mean of the position measurements in the local window; this is the quantity reported in Fig 2 of the main text. MLE results reported unaffected by this alternative formulation (and a trivial mapping can remove the influence of the centering).

2.2.1 Initial Estimates of the Thermal and Measurement Noise

One advantage of assuming and fitting an SDE is that the effects of thermal and measurement noise can be systematically and accurately estimated in the time domain if one has frequently sampled data.[16–18] In the first stage of obtaining initial parameter estimates, we estimate a directed diffusion model of the form $r_{t_{i+1}} = r_{t_i} + (t_{i+1} - t_i)v + \eta$ from noisy measurements $\psi_{t_i} = r_{t_i} + \epsilon_{t_i}$ where v is the constant velocity vector and η is a Gaussian random variable with mean zero and variance σ^2 (note: this notation corresponds to a statistical mechanics diffusion coefficient of $\frac{\sigma^2}{2}$). The classic directed diffusion model allows one to separately estimate a MLE parameter for each component of $r = (x, y, z)$, i.e. the parameters associated with x are $\theta = (v_x, R_x, \sigma_x)$ and these are assumed to be statistically independent of the corresponding y and z parameters. The more general multivariate case (e.g. allowing for correlated diffusive noise) can also be readily handled,[18] but we restrict attention to the scalar case to facilitate exposition. We use Δt to denote the uniform time sampling rate used in our experiments. However the methods presented in this subsection do not require uniform temporal sampling in order to obtain exact likelihood expressions (this can be relevant to other super-resolution or quantum dot applications). If the observations are consistent with the model and the parameters are known, then the transition density corresponding to $\{\psi_{t_i} - \psi_{t_{i-1}} - \Delta t v\}_{i=1}^{n_j} \equiv \{\Delta\psi_{t_i} - \Delta t v\}_{i=1}^{n_j}$ is a multivariate Gaussian with mean zero and covariance:[17,18]

$$\mathcal{C} \equiv \begin{pmatrix} \sigma^2\Delta t + 2R & -R & 0 & \dots & 0 \\ -R & \sigma^2\Delta t + 2R & -R & \ddots & 0 \\ 0 & -R & \sigma^2\Delta t + 2R & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -R \\ 0 & 0 & 0 & \dots & -R \quad \sigma^2\Delta t + 2R \end{pmatrix} \tag{10}$$

The eigenvalues and eigenvectors of the above matrix can be computed analytically if one notes that the matrix above can be written as the sum of the identity matrix multiplied by $\sigma^2\Delta t$ minus the second forward difference operator matrix multiplied by the scalar R. Eigenvalue "i" of the latter matrix is given by $4R\sin^2\left(\frac{\pi i}{2(n_j+1)}\right)$, so

eigenvalue i of the above matrix is simply $\sigma^2\Delta t+4R\sin^2\left(\frac{\pi i}{2(n_j+1)}\right)$. The eigenvectors can also easily be computed with a compact expression (and hence $\mathcal{C}^{-1}$ can be explicitly constructed), but one only needs to evaluate the logarithm of the determinant of $\mathcal{C}$ (which only requires eigenvalues) and $V^T\mathcal{C}^{-1}V$ (V denotes a column vector containing each of the observations $\{\Delta\psi_{t_i}-\Delta tv\}_{i=1}^{n_j}$) in order to evaluate the likelihood function.[17,18] We recommended exploiting the tridiagonal structure and sparse matrix solvers[19] in order to efficiently evaluate the likelihood function instead of forming the exact matrix inverse (which is dense). Doing so can save one substantial time if one is analyzing many long SPT trajectories. A quasi-likelihood analysis[17,18] of the above model has some nice statistical properties . The utility of these properties in regards to SPT applications will be demonstrated in the Simulation section. After one finds the MLE (or quasi-MLE) of the model discussed here for each component, one can form the two diagonal matrices $\tilde{R}$ and $\tilde{C}$ each matrix containing the output of the noise parameters of the directed diffusion model on the diagonals (the velocity parameter along with $\tilde{C}$ can be used to provide an initial guess to the vector $\tilde{A}$). Tildes are used to denote the initial guesses formed using the procedure outlined here. Also note that we did not use θ to emphasize that the estimated parameters above are scalars (the vectors and matrices are formed by merging the x,y,z results).

2.2.2 Initial Estimates of the "B" Matrix

There are a variety of approaches one can entertain for estimating the matrix B associated with the linear SDE considered. Our approach utilizes the Riccati matrix equations[7] to generate initial guesses. Here we utilize some tools associated with the continuous Kalman filter.[7] Recall that in the models we are considering, the force is assumed to come from the gradient of a potential that has existent second derivatives everywhere (hence B, representing the "elasticity" or gradient of the force, is a symmetric matrix). The relevant Riccati equations are:

$$\dot{P} \equiv \Upsilon P + P\Upsilon^{\mathrm{T}} + Q - PR^{-1}P \tag{11}$$

$$0 = \Upsilon P_{ss} + P_{ss}\Upsilon^{\mathrm{T}} + Q - P_{ss}R^{-1}P_{ss} \tag{12}$$

where P is the covariance associated with the Kalman filter,[6,7,11] Υ is a matrix related to the SDE parameter B of interest (namely $\Upsilon \equiv \Phi B \equiv CC^{\mathrm{T}}B/(k_BT)$), Q is the "instantaneous" covariance of the process noise in the uniform time spacing case ($= 2CC^{\mathrm{T}}$), R is the covariance of the measurement noise, and P_{ss} filter covariance reached at "steady state". In order to estimate Q and R, we appeal to the estimator advocated in Refs.[17,18] except that we include the constant directed diffusion parameter A in the model in addition to noise estimates (see previous section for the likelihood expression of this model). Recall that in all cases, we compute the empirical mean of the measurements in window j, denote this quantity by $\hat{\mu^j}$, and subtract this from the measurements in window j. To estimate P_{ss}, we first compute the empirical covariance of the measurements in a local window j. Denote this empirical covariance by $\widehat{\mathrm{Cov}}\left(\{\psi_i\}_{i=0}^{nj}\right)$. In the models considered, this empirical covariance can be used to provide an estimate of P_{ss}, denote the estimate by $\tilde{P_{ss}}$, by the equation $\tilde{P_{ss}} = \widehat{\mathrm{Cov}}\left(\{\psi_i\}_{i=0}^{nj}\right) - \tilde{R}$. One can then utilize established matrix methods by plugging in $\tilde{Q} = 2\tilde{C}\tilde{C}^{\mathrm{T}}$, defining $\tilde{\Upsilon} = \tilde{C}\tilde{C}^{\mathrm{T}}\tilde{B}/(k_BT)$, and solving the the transpose of Eqn. 12 for the symmetric $\tilde{B}$ by replacing the estimates denoted by tildes for their analogs without tildes appearing in Eqn. 12. For discrete observations, a heuristic we have found useful is to compute $\tilde{P_{ss}}$ as before, but solve the Riccati equation by omitting any terms including R (the estimate $\tilde{P_{ss}}$ attempts to remove measurement noise effects and back out the "steady state" covariance). Well established linear algebra tools can be used for this computation;[9,10] such tools exploit the symmetric nature of $\tilde{B}\tilde{B}$. The symmetric case considered can also be used to generate initial conditions for more general models (e.g., those having forces not coming from the gradient of a potential). One can then use these initial estimates (denoted by tilde) of the model parameters to start optimization search for the MLE parameters.

The practical and theoretical utility of assuming a stationary filter covariance is admittedly questionable in some scenarios. For this reason, we suggest using other estimates of B, e.g. setting all quantities in the matrix to zero or finding the linear SDE MLE parameter ignoring measurement noise all together * to generate other candidate initial guesses that do not appeal to this type of assumption.[8] One should also be aware that it in 3D

*In 1D this permits a closed-form expression for the MLEs without requiring a parameter search.[5]

SDE estimation, the mathematical conditions guaranteeing that the parameter vector is uniquely identifiable are nontrivial.[6,7] Even if the observed data follows the assumed model and the parameter is identifiable, one has to be careful in finding the global optimal parameter in a nonlinear MLE search.[6]

2.3 Qualitative Description of Goodness of Fit Tests

Although the Kalman filter has been extensively utilized in control and estimation,[7,11] rigorous statistical methods for analyzing the the goodness-of-fit of the model have been lacking. While it is well-known that if the unobservable state and observed measurement data is consistent with the assumed SDE and measurement model, then the normalized "innovation sequence"[11] † is independent and identically distributed (i.i.d.). Furthermore, it is known that the distribution of the normalized innovation sequence follows a multivariate Gaussian distribution characterized by a mean of zero and an identity matrix as the covariance. These statistical properties can be used for filter tuning (i.e., parameter estimation),[11,20] but hypothesis testing methods which jointly check the independence and the standard normal distribution assumption along a single trajectory have only recently been developed.[3,21]

In this subsection, the basic idea behind the goodness-of-fit approach developed by Hong and Li is presented in a descriptive fashion using a classic model in the SPT literature (a so-called directed diffusion model[12]). The example used to illustrate the basic idea behind method is the following: If a tagged particle is subjected to random thermal forces which can be modeled by a Brownian motion process and measurement noise does *not* contaminate the observational data (an admittedly unrealistic assumption), a directed diffusion model would state $r_{t_{i+1}} = r_{t_i} + (t_{i+1} - t_i)v + \eta$ where again v is the constant velocity vector and η is a Gaussian random variable with mean zero and variance σ^2. If this model is consistent with the data and the parameters are precisely known, one can transform the series $\{\psi_{t_i}\}_{i=0}^N$ (which equals $\{r_{t_i}\}_{i=0}^N$ because measurement noise is assumed to be nonexistent) to $\{\nu_{t_i}\}_{i=0}^N$ where $\nu_{t_i} = \big(r_{t_i} - r_{t_{i-1}} - (t_{i+1} - t_i)v\big)/\sigma$. The advantage of this procedure is that a non-stationary time series can be transformed into a sequence of i.i.d. random variables $\{\nu_{t_i}\}_{i=0}^N$ having a standard normal distribution *if* the assumed model is consistent with the measurements and the directed diffusion model parameters are precisely known. Time series methods[3,21] can be used to systematically *jointly* test both the distribution of the transformed variables and the independence assumption. This is accomplished by transforming the ν series into a suitable test statistic.

Although the transformation shown here is trivial for the directed diffusion model (it only requires evaluating the cumulative distribution function of standard normal random variable evaluated at ν_{t_i}), noisily measured nonlinear Markovian SDEs can also be transformed into a "generalized residual" series[21] by introducing the probability integral transformation (PIT), or Rosenblatt transformation, defined by $Z_i \equiv \int_{-\infty}^{\psi_{t_i}} p(\psi|\psi_{t_{i-1}};\theta)d\psi$.[3] Technical details associated with the more general nonlinear multivariate SDE are discussed in Refs.[3] and.[22] After the generalized residuals are formed, an "ominbus" Q-test[3] can be used to test if all known statistical properties of the PIT sequence hold. In addition, the M-tests introduced in Ref.[3] can be used to for a subset of the statistical properties that should hold in the PIT sequence. The Q-tests are "omnibus" in the sense they check for every possible model misspecification[3] but the price one pays for this is power, i.e., one may fail to reject an incorrect model because one is checking too many features simultaneously.[23] Although the M-test does not check the uniform distribution shape, it is useful in detecting statistical dependence induced by dynamical features in the data not accounted for in the models considered (several examples will be shown later). Both tests can be applied to the innovations of a Kalman filter whose parameters are calibrated from data. Advantages of using the generalized residuals vs. standard tools leveraging innovations to assess the suitability of an assumed model to observational data will be presented in the results and is discussed in length elsewhere.[24] A summary of the equations derived in[3,21] needed to evaluate the Q and M test statistics is provided in the next subsection for the reader's convenience.

2.4 Sources of Variation in Test Statistics

It should be noted that there are several sources of variation in this hypothesis testing procedure. Even if an oracle provides the "truth" parameters, the M and Q test statistic would be random variables possessing a certain

†The normalized innovation is the standard innovation left multiplied by the square root of the inverse of innovation covariance matrix. These quantities can be computed from the data and given filter parameters.[11]

distribution under the null hypothesis (here the null hypothesis is the PIT distribution using the correct model and the fixed oracle-given parameter). For finite samples sizes (again plugging in one oracle-given parameter) the distribution of the tests statistic is different from the asymptotic standard normal distribution.[21,25] Another source of variation is the propagation of parameter uncertainty induced by finite sample sizes. Even if one fixed parameter generates every time series, the MLE estimates for each trajectory will contain parameter sampling noise (though this effect disappears for suitable "$\sqrt{N}$" estimators[3]). Yet another source of variation in the M and Q statistics is induced by varying types of model misspecification.

2.5 Mathematical Details of Goodness-of-Fit Testing and Model Selection Criteria

The PIT transformation of a correctly specified model has a known distribution $U[0,1]$ and each observation is statistically independent of the other observations.[3] This condition does not require one resorting to asymptotic arguments, so hypothesis tests can be established which simultaneously check if the the transformed Z_i's (defined in Section 2.3) are statistically independent and have the $U[0,1]$ shape. Deviance from either condition suggests that there is evidence in the data showing statistical problems with the assumed and fitted model. Hong and Li proposed the so-called "omnibus" Q test statistic (Eqn. 14) which jointly checks both the i.i.d. and $U[0,1]$ shape assumption. The M-test does not check for the $U[0,1]$ shape of the PIT transformation but instead focuses on autocorrelations in the Z_i's. In the case of a Kalman filter, this test is fairly similar to analyzing only the autocorrelation function of the innovations; however, the method analyzes the PIT (or generalized residuals) and can be used with nonlinear and/or non-Gaussian dynamical models where a likelihood function is available (as we will show later, the test can detect such features in data where linear models are assumed whereas classic Kalman filter diagnostics fail). The M-test is also useful as a diagnostic tool[3] because it utilizes all moments of an assumed (nonlinear) dynamical model as opposed to just analyzing the first two moments. Tools of this sort will be helpful in assessing our simplified models given observational data.

We note that for general multidimensional processes, there are open questions and issues associated with marginalizing and formation of the PIT sequence.[3,22] The 1D goodness-of-fit tests were formed as recommended originally in Hong and Li.[3] For the multivariate linear cases, we applied the PIT to the sequence of normalized innovations (formed by applying the matrix square root of the filter covariance associated with the MLE to the innovation vector at each time). For a correctly specified model, the latter sequence is i.i.d. normal if the Kalman filter model is consistent with the observations, so the PIT associated with this transformed sequence can be readily formed as in the scalar case, except that the length of the PIT series now increases by a factor of three for (x, y, z) measurements. In the 3D case, we formed the longer 1D PIT sequence by interweaving the PIT computed using the normalized 3D innovation vector (i.e., at each time, we alternated between recording the innovation in the x,y, and z position). We also explored stacking (i.e., place all PITs computed using the normalized x innovations in the order of the corresponding temporal observation, followed by all of those PITs computed using normalized y innovations, etc.). The test statistic results obtained were not very different in either case.

If the underlying data generating process is outside of the linear Gaussian class, extensions of other recent multivariate goodness-of-fit methods for non-stationary SPT signals show promise.[22,26,27] The interesting trajectories we reported in the main text were discovered when the 1D analog of our SDE models were strongly rejected along several windows of a 3D trajectory. Again, in 1D, PIT based methods have been shown to exhibit favorable power in many situations. Simulation results are presented where the PIT based method is shown to be able to detect 3D interactions with reasonable power using the approach discussed in this Section.

After the parameters of the SDE are estimated, the raw data is used to compute the PIT sequence using the methods stated above. Recall that the i.i.d. $U[0,1]$ property holds at the exact data generating process. In practice, one plugs in an MLE $\hat{\theta}$ and uses the SDE model structure along with the observations to compute the associated $\hat{Z}_i$'s. This sequence is then used to compute the quantities reported in Hong and Li,[3] namely:

$$\hat{M}(m,l) \equiv \Big(\sum_{j=1}^{N-1} w^2(j/p)(N-j)\hat{\rho}^2_{ml}(j) - \sum_{j=1}^{N-1} w^2(j/p) \Big) / \Big(2 \sum_{j=1}^{N-2} w^4(j/p) \Big)^{\frac{1}{2}} \tag{13}$$

$$\hat{Q}(j) \equiv \Big(h(N-j)\hat{I}(j) - hA^0\Big)/V_0^{\frac{1}{2}} \tag{14}$$

$$\hat{I}(j) \equiv \int_0^1 \int_0^1 \Big(\hat{g}_j(\mathcal{Z}_1, \mathcal{Z}_2) - 1\Big)^2 d\mathcal{Z}_1 d\mathcal{Z}_2.$$

$$\hat{g}_j(\mathcal{Z}_1, \mathcal{Z}_2) \equiv \frac{1}{N-j} \sum_{n=j+1}^{N} K_h(\mathcal{Z}_1, \hat{Z}_n) K_h(\mathcal{Z}_2, \hat{Z}_{n-j})$$

A full description of the terms appearing above is given in Ref.[3] where they are first derived, but the essence of the above two test statistics can be understood by noting that h is a bandwidth used along with a specialized nonparametric kernel function $K_h(\cdot,\cdot)$ aiming to estimate a $U[0,1]$ distribution on the square (and also take into account boundary effects) given a finite set of observations and N is the number entries in the observed time series. In the definition of Q, j is a user specified index that selects the spacing between the $\hat{Z}_n$'s (a value of $j=1$ is commonly used.[3] In the definition of $\hat{M}$, $\hat{\rho}_{ml}(j)$ denotes the sample cross correlation between $\hat{Z}_n^m$ and $\hat{Z}_{n-|j|}^l$ and $w(\cdot)$ is the Bartlett kernel;[3] note that in $\hat{M}(m,l)$, the data now selects j but one specifies the integer parameters m and l. The expression for the remaining constants A^0 and V_0 are derived and provided in Ref.[3]

2.6 Model Selection Criteria

We also considered model selection in addition to goodness-of-fit in the main text. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are classic methods for selecting models.[28,29] Each penalizes model complexity in a different fashion, namely AIC$= -2\ell(\hat{\theta}) + 2k$ and BIC$=-2\ell(\hat{\theta}) + k\log(N)$ where k represents the number of components in $\hat{\theta}$ and N is the number of observations used to compute the MLE parameter estimate ($\ell(\cdot)$ denotes the log likelihood). Lower values denote a "better" model (in the main text we compare differences between the AIC and BIC computed using 1D and 3D likelihoods evaluated at the MLE). In the plots presented in the main manuscript and in the SI, we plot ΔAIC and ΔBIC; these quantities represent the difference in information criteria between the 3D and 1D models (a value less than zero suggests that a 3D model is favored).

3. ADDITIONAL SIMULATIONS (METHOD VALIDATION)

Here additional simulation data is used to further illustrate some points made in the primary text. For example, we provide a brief study demonstrating how a maximum likelihood method performs as compared to the traditional MSD analysis with simulated data mimicking some features of experimental *ARG3* data (bias and variance of parameters estimates in a batch of trajectories generated by a fixed SDE are studied). Both estimation and inference results are shown for the analysis of a 1D SDE. In addition, a more detailed discussion illustrating how legacy Kalman filter innovation inspired goodness-of-fit analysis methods would fail to detect some features present in simulation data are presented in detail.

3.1 1D Estimation and Goodness of Fit Results for Simulation Data

The first set of set of results focuses on a simulated 1D SDE data where the measurement noise and/or the diffusion coefficient was allowed to vary over time. Four cases are studied, and in all cases 1,600 observations uniformly spaced 4 ms apart were sampled from an SDE with no drift. The reference case (Case 1) had a constant diffusion coefficient and measurement associated with the $t=0$ value of the functions reported in Fig. S8. Case 2 had the same constant measurement noise, but used the diffusion function shown in Fig. S8; Case 3 used a constant diffusion coefficient evaluated at $t=0$, but used the time dependent measurement noise shown in Fig. S8 whereas Case 4 used both time dependent noise functions in Fig. S8. The diffusion coefficient was computed using both the MSD analysis (attempting to account for measurement noise) reported in[1] and with the MLE approach presented here in Section 2.2. The mean and standard deviation of parameter estimates obtained over repeated simulations for the four cases discussed here are reported in Table S2. Note that because

we use a deterministic function of time for the diffusion coefficient, we can easily compute the time averaged integrated diffusion coefficient[18] (i.e., the effective diffusion coefficient) in all cases considered. The asymptotic mean and variance of the parameter distribution of of the quasi-MLE parameter estimates can be obtained in closed-form.[18] When the measurement noise and diffusion coefficient are frozen (Case 1), both estimators have the same average value, but the standard deviation (or "parameter uncertainty") of the MSD based estimator is substantially larger than that of the MLE based approach. In the cases where the diffusion coefficient is not constant (Case 2 and 4), the estimator advocated in Ref.[18] has a mean consistent with the integrated time averaged diffusion coefficient (reported in the row labeled "Truth") whereas the MSD based estimator has a mean away from the effective diffusion coefficient in addition to being associated with a higher parameter uncertainty.

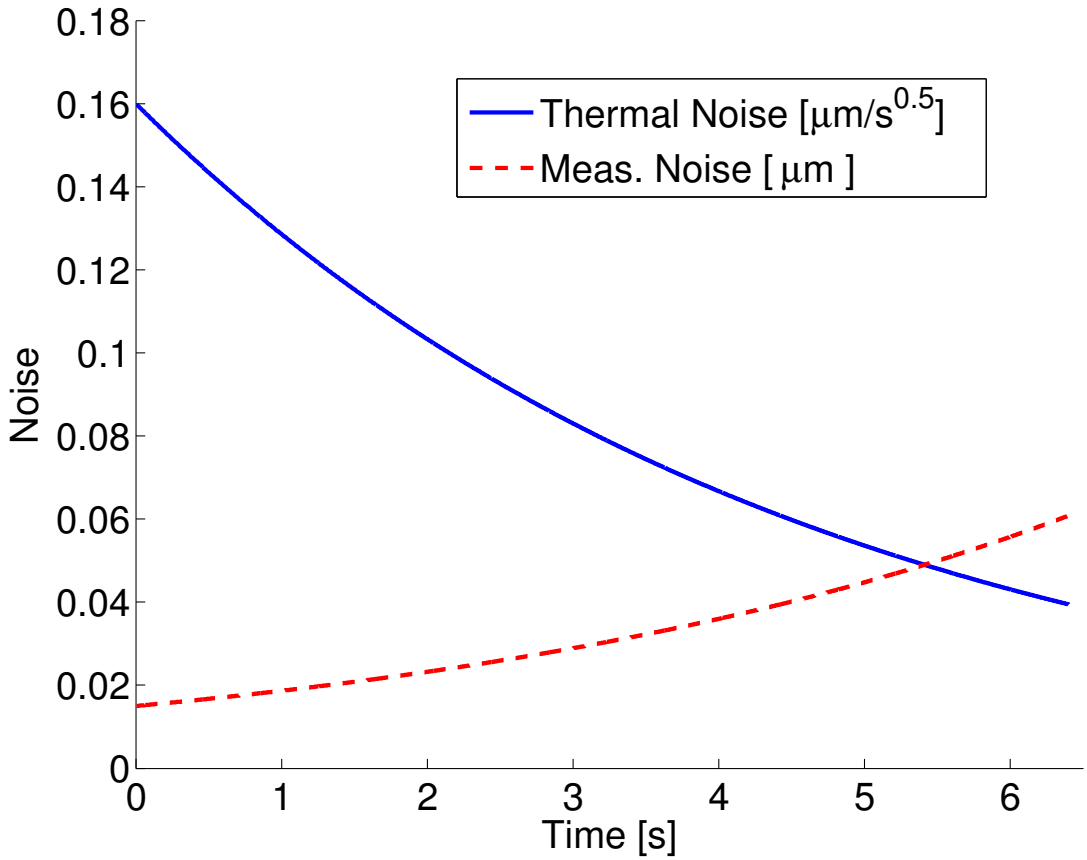

Fig. S 8. Time dependent measurement and thermal noise used in MC simulations.

Table S 2. Diffusion Coefficient (MSD vs. MLE). Both methods account for measurement noise, but the MLE method directly handles correlations induced by differencing measurements.[16]

Case	1	2	3	4
MSD	0.150 (0.123)	0.078 (0.069)	0.151 (0.124)	0.079 (0.069)
MLE	0.160 (0.047)	0.092 (0.032)	0.160 (0.056)	0.092 (0.038)
Truth	0.15	0.093	0.15	0.093

3.2 Demonstration of Problems with Classic Kalman Filter Diagnostics on SPT Data

The previous results just demonstrated that time varying measurement noise and diffusion coefficient affected parameter estimates, but now we focus on using goodness-of-fit tests to detect departures from standard assump-

tions employed in SPT analysis. Namely we assume constant diffusion and measurement noise and we want to see if this incorrect modeling assumption can be detected in the sample sizes under consideration. Table 3 displays the goodness-of-fit tests obtained using the Q statistic. For a fixed parameter, this test statistic asymptotically has a standard normal under a correctly specified model.[3] In finite samples, some simple corrections can be made[21] [‡]. Reliably computing this test statistic requires an accurate means for evaluating the transition density of an assumed model. We consider two cases, one where correlation is correctly accounted for in the transition density (i.e., use the likelihood discussed along with Eqn. 10) and one where correlation induced by differencing measurements is not (this case can be obtained by ignoring the off-diagonal matrix entries in Eqn. 10). The incorrect transition density described in the latter case is commonly used in the SPT literature, e.g. see citations in Berglund.[30] We emphasize that the transition density in the presence of measurement error requires one to account for correlation if one wants to carry out rigorous statistical inference. Carrying out rigorous inference (checking all model assumptions) is important because effective SDE parameters are often changing over time in nanoscale systems.

Using an incorrect transition density would obscure detecting a time changing effective measurement noise or diffusion coefficient (using the wrong transition density may cause excessive rejection even if the modeling assumptions are consistent with the experimental data). Table 3 demonstrates that ignoring statistical dependence causes one to make incorrect inferential decisions even for a correct model (e.g., Case 1 is strongly rejected for all observations). One can also see that time varying measurement noise is easier to statistically detect than time varying diffusive noise in this example.

Table S 3. Goodness-of-Fit Results (Proxy vs. Exact MLE [QMLE]). The case labeled as "proxy" ignores statistical dependence induced by differencing measurements (Ref.[30] cites several publications making this error; the scenario we consider is described in the above paragraphs) whereas the case labeled "exact" correctly accounts for the dependence induced by differencing measurements.

Case	1	2	3	4
Proxy	20.457 (3.339)	34.699 (4.171)	45.595 (5.531)	51.466 (5.800)
Exact	-0.433 (1.129)	-0.133 (1.121)	7.062 (2.471)	6.642 (2.435)

This subsection concludes by discussing and demonstrating the utility of generalized residual based statistics (such as the Q statistic) over classical diagnostic statistics used in stochastic process control theory.[7,11] It is well-known that a correctly specified Kalman filter based model is associated with an innovation sequence which can be normalized to produce an i.i.d. sequence of normal random variables, and several methods exist for separately checking for normality or correlation,[7,11] but rigorous methods for simultaneously checking the shape (i.e., distributional assumptions) and independence assumptions simultaneously were lacking until the development of time series tools like those advocated in Refs.[3,21,26] In order to illustrate the utility of simultaneously checking the shape and independence assumptions, we present an analysis of a normalized innovation sequence (and the corresponding PIT random variable) for one time series realization from data whose population statistics are shown in Tables S2 and S3. Figure S9 displays histograms of the normalized innovations (Panel A) and the generalized residuals (Panel B) as well as two autocorrelation functions derived from these quantities for one trajectory realization from Case 4. Panel C plots the basic sample autocorrelation of each and Panel D plots the sample autocorrelation of of the squared PIT random variable sequence. The mean and variance of the data measured in Panel A was 6.287×10^{-4} and 1.0006, respectively. Using the first two moments of the residuals alone seems to suggest that the observed data was consistent with the model assumptions (despite the fact that the data and assumed model are not consistent in Case 4). Analyzing the full distributional shape alone via the Kolmogorov-Smirnov test produced a p-value of 0.0656 for both Panels A and B (the number was identical because a known one-to-one transformation connects observations in Panel A to those in Panel B in this example). Panel C displays the autocorrelation typically analyzed when doing diagnostic checking on a Kalman filter.[7,11] Again there is some evidence suggesting an inconsistency between model assumptions and the data, but evidence using correlation information in the normalized innovation or generalized residuals alone is relatively weak. However, if one analyzes the sample autocorrelation of the square of the normalized

[‡]The Case 1 correlated case can serve as our "empirical reference null" here.

innovations (this quantity is known to be relevant to detecting departures from Kalman filter type assumption due to noise misspecification induced by "volatility" effects[21]), then misspecification becomes readily apparent. The problem with this approach is that determining which functions of the innovations (here we analyzed the squares because of *a priori* knowledge of the nature of model misspecification) is difficult if one does not know in advance why the assumed model might be inconsistent with the experimental data. Having a test statistic checking against all possible model misspecifications, i.e. and "omnibus" test, can be useful and scientifically insightful. The test introduced in Ref.[3] is omnibus in that it simultaneously checks all distributional assumptions as well as dependence structure. This can dramatically increase power; in this case, the Q statistic computed using the generalized residual for this realization was 7.379. This value far exceeds any threshold typically used in diagnostic checking.

The basic idea of analyzing the sequence of normalized innovation sequence is similar in spirit to the generalized residual or PIT approach discussed here except the latter aims at checking all distributional and dependence assumptions simultaneously (this can be important if there are features in the data that are not Gaussian). Another important difference is that the approach of Hong and Li is applicable to any data (stationary or non-stationary) and stochastic model where a transition density can be accurately approximated whereas the Kalman filter based innovation analysis is restricted to linear models.[11] Note that we only displayed results from a single time series, one does not need an ensemble to do diagnostic testing with even if the data comes from a non-stationary process (the diffusion plus measurement noise model studied here is not a stationary process and non-stationary processes are common in SPT applications). Analyzing the goodness-of-fit of a fitted model against a single realization is useful in studying dynamical heterogeneity and stays closer in spirit to the inspiration behind many single-molecule experiments. Table 3 demonstrates that the high rejection afforded by analyzing the Q statistic (simultaneously checking shape and independence) was not merely a fluke due to a fortuitous realization.

3.3 3D Goodness-of-fit simulation results

Table S 4. Goodness of Fit for 3D Simulation Data Corresponding to Table 1 of the main text.

Case	1	2	3	4
Q	-0.347 (1.200)	0.176 (1.154)	-0.270 (1.173)	0.157 (1.140)
M	-0.484 (0.680)	1.660 (1.642)	-0.799 (0.760)	8.036 (2.975)

In Tab. 4, the average and standard deviation of the M and Q tests (computed for each segment using the computed MLE parameters in that segment) are provided for the four cases studied with the main text Table 1. These tests are asymptotically distributed as a standard normal under a fixed data generating process.[3] We verified that the normal test statistic distribution was reasonable for our sample sizes (i.e., the mean/variance reasonably characterized the test statistic distributions). A one-sided hypothesis test would reject (with Type I error rate, $\alpha, = 0.05$) any test statistic exceeding 1.64 if the test statistic distributions follow a standard normal. There is ample evidence to reject the model neglecting 3D interaction terms using the M statistic for any sample size. The mean of the Monte Carlo simulations actually exceeds the $\alpha = 0.05$ threshold for all 1D cases (note: over half of all incorrectly specified cases were rejected so the empirical mean is not dominated by a few outliers).

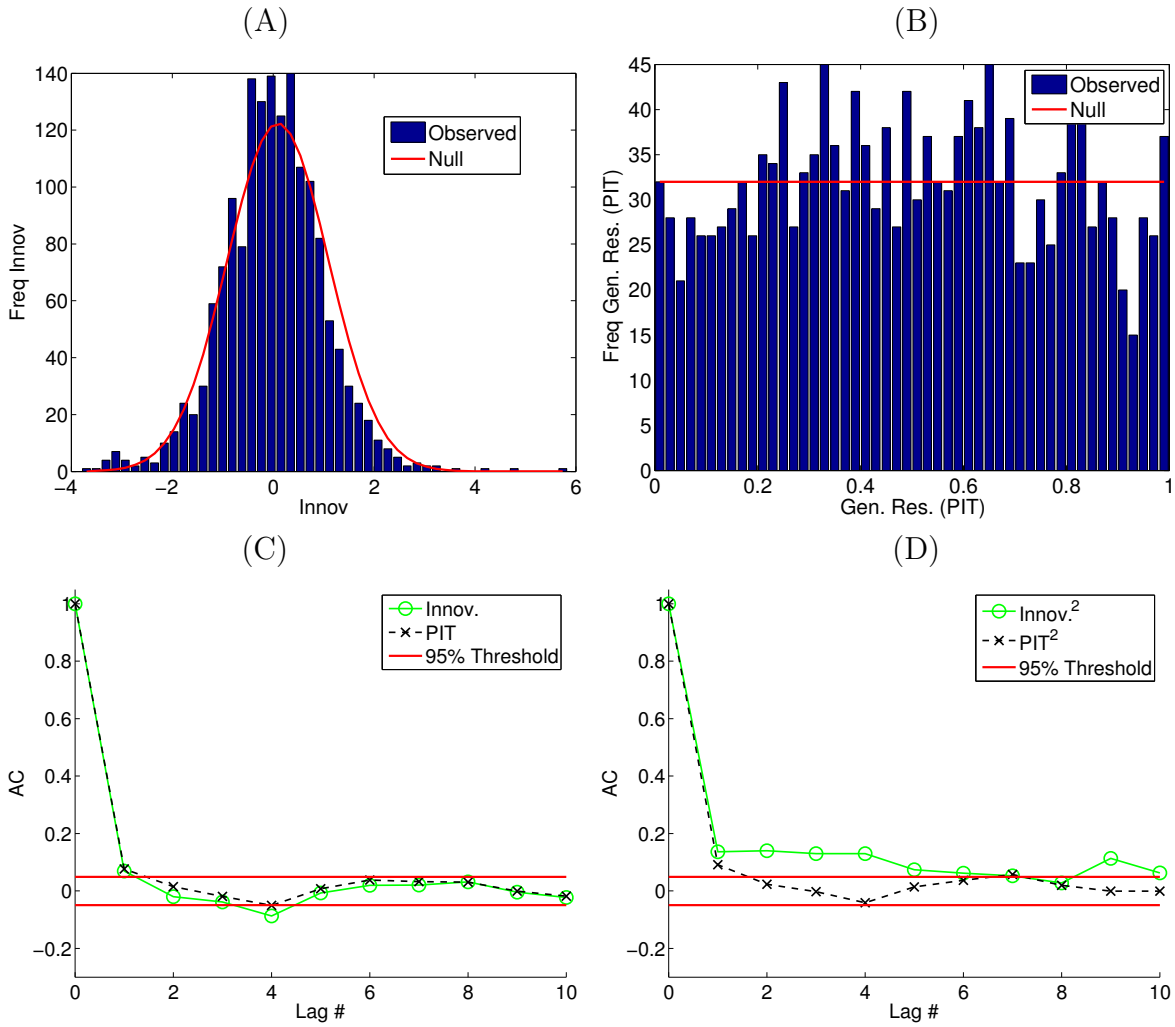

Fig. S 9. Analysis of Innovation and PIT "Residuals". The top two panel show the distribution of the innovations normalized by the matrix square root of the Kalman filter innovation covariance at each time step (Panel A) and the corresponding PIT generalized residuals (Panel B). The bottom two panels show the autocorrelation of these residuals (Panel C) and the square of the residuals (Panel D). The thick solid lines correspond to the 95% confidence band under the null (the null assumes no temporal correlation between the random variables).

REFERENCES

[1] Thompson, M. A., Casolari, J. M., Badieirostami, M., Brown, P. O., and Moerner, W. E. *Proceedings of the National Academy of Sciences of the United States of America* **107**(42), 17864–71 October (2010).

[2] Michalet, X. and Berglund, A. *Physical Review E* **85**(6), 061916 June (2012).

[3] Hong, Y. and Li, H. *Rev. Fin. Studies* **18**, 37–84 (2005).

[4] Thompson, R. E., Larson, D. R., and Webb, W. W. *Biophysical Journal* **82**(5), 2775–83 May (2002).

[5] Tang, C. and Chen, S. *J. Econometrics* **149**, 65–81 (2009).

[6] Hamilton, J. *Time Series Analysis*. Princeton University Press, Princeton, NJ, (1994).

[7] Stengel, R. *Optimal control and estimation*. Dover Publications, Toronto, Ontario, (1994).

[8] Kloeden, P. and Platen, E. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, (1992).

[9] Gardiner, J. D., Wette, M. R., Laub, A. J., Amato, J. J., and Moler, C. B. *ACM Trans. Math. Softw.* , 232–238 (1992).

[10] Hopkins, T. *ACM Transactions on Mathematical Software (TOMS)* **28**(3), 372–375 (2002).
[11] Bar-Shalom, Y., Kirubarajan, T., and Li, X.-R. *Estimation with Applications to Tracking and Navigation.* John Wiley & Sons, Inc., New York, NY, USA, (2002).
[12] Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R. H., and Bertrand, E. *Curr Biol* **13**(2), 161–167 (2003).
[13] Risken, H. *The Fokker-Planck Equation.* Springer-Verlag, (1996).
[14] Blackwell, P. G. *Biometrika* **90**(3), 613–627 September (2003).
[15] Le Cam, L. *International Statistical Review* **58**, 153–171 (1990).
[16] Zhang, L., Mykland, P., and Ait-Sahalia, Y. *Journal of the American Statistical Association* **100**, 1394–1411 (2005).
[17] Xiu, D. *Journal of Econometrics* **159**, 235–250 (2010).
[18] Aït-Sahalia, Y., Fan, J., and Xiu, D. *Journal of the American Statistical Association* **105**(492), 1504–1517 December (2010).
[19] Saad, Y. *Iterative Methods for Sparse Linear Systems (2nd ed.).* Societ of Industrial and Applied Mathematics, Philadelphia, (2003).
[20] Wang, Q. and Moerner, W. E. *ACS nano* **5**(7), 5792–9 July (2011).
[21] Hong, Y., Li, H., and Zhao, F. *Journal of Econometrics* **141**(2), 736–776 December (2007).
[22] Chen, B. and Hong, Y. *Econometric Theory* **26**(04), 1115–1179 November (2010).
[23] Janssen, A. *Annals of Statistics* **28**(1), 239–253 (2000).
[24] Hong, Y. *Journal of the American Statistical Association* **94**(448), 1201 December (1999).
[25] Calderon, C. P. *J Phys Chem B* **114**, 3242–3253 (2010).
[26] Cho, J. S. and White, H. *Journal of Econometrics* **162**(2), 326–344 June (2011).
[27] Remillard, B., Papageorgiou, N. a., and Soustra, F. *Journal of Multivariate Analysis* **110**(1), 30–42 (2012).
[28] McQuarrie, A. and Tsai, C.-L. *Regression and time series model selection.* World Scientific Publishing Company, (1998).
[29] Claeskens, G. and Hjort, N. *Model Selection and Model Averaging.* Cambridge University Press, Cambridge, UK, (2008).
[30] Berglund, A. J. *Physical Review. E* **82**(1), 011917 July (2010).