

Intrinsic Cell-to-Cell Variance from Experimental Single-Cell Motility Data

Anton Klimek ¹, Johannes C. J. Heyn ², Debasmitta Mondal ^{3,4}, Sophia Schwartz ⁵, Joachim O. Rädler,²
Perna Sharma ^{3,6}, Stephan Block ⁵ and Roland R. Netz ^{1,*}

¹Fachbereich Physik, *Freie Universität Berlin*, 14195 Berlin, Germany

²Faculty of Physics and Center for NanoScience, *Ludwig Maximilians Universität*, 80539 München, Germany

³Department of Physics, *Indian Institute of Science*, Bangalore 560012, India

⁴James Franck Institute, *University of Chicago*, Chicago, Illinois 60637, USA

⁵Fachbereich Chemie und Biochemie, *Freie Universität Berlin*, 14195 Berlin, Germany

⁶Department of Bioengineering, *Indian Institute of Science*, Bangalore 560012, India



(Received 18 October 2024; accepted 9 May 2025; published 17 June 2025)

When analyzing the individual positional dynamics of an ensemble of moving objects, the extracted parameters that characterize the motion of individual objects, such as the mean-squared instantaneous velocity or the diffusivity, exhibit a spread that is due to the convolution of three different effects: (i) Motion stochasticity, caused by the fluctuating environment and enhanced by limited observation time, (ii) measurement errors that depend on details of the detection technique, and (iii) the intrinsic parameter variance that characterizes actual differences between individual objects, which is the quantity of ultimate interest. We develop the theoretical framework to separate these three effects based on the generalized Langevin equation, which constitutes the most general description of active and passive dynamics, as it derives from the underlying general many-body Hamiltonian for the studied system without approximations. We apply our methodology to determine intrinsic cell-to-cell differences of living and actively moving human breast-cancer cells, algae cells, and, as a benchmark, size differences of passively moving polystyrene beads in water. We find algae and human breast-cancer cells to exhibit significant individual differences, reflected by the spread of the intrinsic mean squared instantaneous velocity over two orders of magnitude, which is remarkable in light of the genetic homogeneity of the investigated breast-cancer cells and highlights their phenotypical diversity. Quantification of the intrinsic variance of single-cell properties is relevant for infection biology, ecology, and medicine, and it opens up new possibilities to estimate population heterogeneity on the single-organism level in a nondestructive manner. Our framework is not limited to motility properties but can be readily applied to other experimental time-series data.

DOI: [10.1103/77t1-2qv](https://doi.org/10.1103/77t1-2qv)

I. INTRODUCTION

Cells of a population typically exhibit largely different genotypes and phenotypes, which creates optimized fitness in reaction to external stimuli. Oftentimes, it is important to know how heterogeneous cells or organisms actually are, e.g., in order to estimate survival probabilities in reaction to environmental changes [1–3] or in order to determine the likelihood of an infection by pathogens [4,5]. The heterogeneity of a population of organisms is typically quantified by hand-picked parameters, for instance the cell size [6], specific binding coefficients [7,8], positional speed of the organism [9–11], etc. One experimentally relatively easily obtainable observable is the position of an organism measured over time, i.e., its positional trajectory, which requires minimal perturbation of the organism during the measurement and has been

used to distinguish different classes of organisms from each other [12]. To quantify and compare trajectories, one usually assumes a model. The most general equation of motion for a general observable that can be derived exactly from the underlying Hamiltonian dynamics is the generalized Langevin equation (GLE), which was successfully used to model binary reaction dynamics in water [13], water vibrational IR line shapes [14], butane dihedral dynamics [15], single-cell motion [12,16], and protein folding [17,18]. The GLE includes non-Markovian effects, i.e., the memory of the trajectory of its past, and it is valid even for nonequilibrium processes, such as cell motion [12,19,20]. Important limiting cases of the GLE are the persistent random walk and Brownian diffusion models as well as active walk models such as the active Ornstein-Uhlenbeck model [16,21,22], which are explained in more detail further below. The motion of any organism and of any microscopic object, be it active or passive, contains intrinsic stochasticity due to the fluctuating environment [23], which is fully captured by the description with the GLE.

Every trajectory that is observed over a finite time span inevitably leads to some uncertainty in the extracted parameters. Therefore, every experimental determination of model parameters contains a spread that originates from the stochasticity of the motion, entangled with experimental measurement errors

*Contact author: rnetz@physik.fu-berlin.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

and with the actual spread due to differences between the individual moving objects. There are many methods available to estimate parameters of stochastic processes [16,24] and to estimate the inevitable variance of such parameters [25,26]. One standard way to assess parameter uncertainties for stochastic processes is to simulate the stochastic model and to compare the simulation outcome with experimental data [25,27]. However, the field of motility analysis still lacks methods to disentangle the spread of intrinsic properties between individual objects from the noise of the environment and other experimental uncertainties.

In this paper, we introduce an approach to tackle this problem: By (i) using previously introduced methods for extracting single-cell GLE parameters from experimental positional trajectories [12,16,28,29], (ii) generating synthetic trajectories from the GLE over the experimental finite observation times with Markovian embedding techniques, and (iii) employing a statistical variance analysis, we estimate the intrinsic spread of the parameters characterizing the individual moving objects. As validation of our GLE-based approach, we demonstrate that the active and passive movements of individual objects of three very different types, chosen to cover a wide range of different patterns of motion, are all perfectly described by the GLE. Specifically, we consider the passive diffusion of polystyrene beads in water, the active motion of human breast-cancer cells on a substrate, and the flagella-propelled active motion of algae cells in two-dimensional confinement. The polystyrene-bead system constitutes a benchmark with experimentally known bead-to-bead size differences, for which we confirm that our model estimates the correct intrinsic parameter variance.

In our previous papers, where we introduced the methods to extract GLE parameters from positional trajectories, we had observed a significant spread of the extracted single-cell parameters, but we could not interpret or analyze this spread since it was not clear whether it is an intrinsic property of the cells or due to noise [12,16]. Here we show that for the algae and the human breast-cancer cell systems, the large variance of the GLE motion parameters over up to two orders of magnitude is in fact largely intrinsic and not due to noise or finite trajectory length.

This finding is very unexpected for the cancer cells, as they derive from a single cell line and thus share a very similar genetic code. It is less surprising for the algae cells, as they stem from multiple colonies and presumably are genetically rather heterogeneous. Our results thus showcase a significant phenotypic cell variance, manifested in the differences of the motion of individual cells, that does not reflect genetic diversity. Our framework is applicable to ensembles of time-series data of varying type and from different systems. It generally allows us to separate stochastic environmental influences from intrinsic variations between individual ensemble members.

II. RESULTS

A. Trajectory model

We analyze the trajectories of three different moving objects: first, the purely passive motion of polystyrene beads in bulk water with a known size distribution; second, the mo-

tion of human breast-cancer cells of line MDA-MB-231 on a one-dimensional micro lane [16,30,31]; and finally, the motion of micro algae cells *Chlamydomonas reinhardtii* confined between two glass plates [12,32], which resembles their natural environment in soil. Exemplary microscopy images of the three different moving objects considered in the experiments are shown in Figs. 1(a)–1(c). We choose these three systems as they cover important limiting scenarios, namely passive Brownian diffusion (polystyrene beads), active persistent random motion (cancer cells), and strongly non-Markovian active motion (algae cells). Thus, our datasets include Markovian motion and non-Markovian motion in equilibrium as well as out of equilibrium. The experiments yield two-dimensional trajectories $x(t)$, $y(t)$ for the center position of the polystyrene beads and the algae cells, examples of which are shown in Figs. 1(d) and 1(f), respectively, and one-dimensional trajectories $x(t)$ for the breast-cancer cells, an example of which is shown in Fig. 1(e). We describe all trajectories by the one-dimensional GLE [33–35]

$$\ddot{x}(t) = - \int_{t_0}^t \Gamma_v(t-t') \dot{x}(t') dt' + F_R(t) \quad (1)$$

with an identical equation for $y(t)$ in the case of two-dimensional motion. For Gaussian motion processes, which perfectly describe the experiments as explained further below, the equation of motion is linear and there is no coupling between the motion in the x and y directions, therefore we average all two-dimensional trajectory data over the two directions. In the GLE Eq. (1), $\ddot{x}(t) = \dot{v}(t)$ denotes the acceleration of the position, $\Gamma_v(t)$ is the memory kernel that describes how the acceleration at time t depends on the velocity $\dot{x}(t') = v(t')$ at previous times and therefore accounts for non-Markovian friction effects, and $F_R(t)$ is a random force that describes interactions with the surrounding and the interior of the moving object. The GLE contains no deterministic force term because the experimental systems are isotropic and homogeneous in space. We keep the inertial term proportional to the trajectory acceleration $\ddot{x}(t)$ in order to regularize the differential equation, which otherwise can become singular for certain functional forms of the memory kernel [22,36]. In fact, the GLE in Eq. (1) is the most general equation of motion for Gaussian unconfined motion processes, and it can be derived by projection from the underlying many-body Hamiltonian, even in the presence of nonequilibrium effects, which obviously are present for living organisms [19,20]. We show further below that all different types of active and passive motion we consider are perfectly described by the GLE Eq. (1), which illustrates its universal applicability. An important special case of the GLE Eq. (1) is the Markovian limit with $\Gamma(t) = 2\delta(t)/\tau_m$, which leads to the well-known Langevin equation and describes persistent motion. In the so-called overdamped limit, in which case the persistence time τ_m goes to zero, one recovers the even simpler Brownian diffusion model.

If the motion can be described as a Gaussian process, which for polystyrene beads, cancer cells, and for algae cells is suggested by the fact that the single-individual velocity distributions are Gaussian, as shown in Figs. 2(b), 2(e), and 2(h), the random force is Gaussian as well [20], with correlations

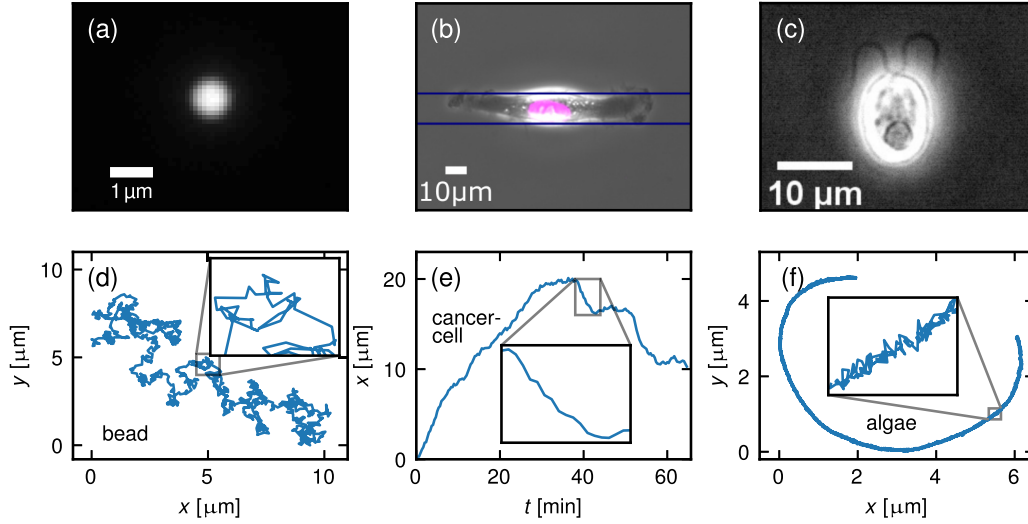


FIG. 1. Microscopy image of (a) a fluorescently labeled polystyrene bead with radius $r = 0.5 \mu\text{m}$ in a three-dimensional aqueous solution, (b) a living breast-cancer cell (MDA-MB-231), where the pink area is the fluorescently labeled nucleus of the cell and the horizontal lines indicate the fibronectin-covered lane on which the cell moves, and (c) a *Chlamydomonas reinhardtii* algae cell confined between two glass plates at a separation of about $10 \mu\text{m}$. In (d)–(f) we show exemplary trajectories of the different moving objects shown in (a)–(c). The trajectory lengths are (d) $L = 23.7 \text{ s}$, (e) $L = 64.7 \text{ min}$, and (f) $L = 12.4 \text{ s}$, where the insets show enlarged sections of the discrete trajectories.

given by

$$\langle F_R(t)F_R(0) \rangle = \Gamma_R(t), \quad (2)$$

where $\Gamma_R(t)$ denotes the symmetric random-force kernel. Averaging cell velocities over the cell population can result in non-Gaussian velocity distributions, even though distributions are Gaussian on the individual trajectory level [12,16], as we show in the Supplemental Material (SM) [37]. Therefore, we check the Gaussianity on the single-trajectory level in Fig. 2.

The GLE Eq. (1) does not contain parameters depending on time, and therefore it describes a stationary process. Stochastic models with time-dependent parameters are well known [38]. However, for the polystyrene beads, the cancer cells, as well as for the algae cells, the velocity distributions do not change in time, as shown in Figs. 2(c), 2(f), and 2(i), which suggests that the GLE Eq. (1) can be used to model the objects' motion. This is confirmed by the fact that the GLE model with time-independent parameters describes the single-cell trajectories very well over the entire trajectory length, which is shown later.

For an equilibrium system, such as the passively moving polystyrene beads in water, the fluctuation-dissipation theorem (FDT) predicts for the GLE in Eq. (1) $\Gamma_R(t) = B\Gamma_v(|t|)$, where $B = \langle v^2 \rangle$ is the mean-squared velocity given by $B = k_B T/m$ according to the equipartition theorem, with m the moving object's mass and $k_B T$ the thermal energy [33,34]. For living organisms, the FDT and the equipartition theorem generally do not apply, so $\Gamma_R(t)$ and $B\Gamma_v(|t|)$ are not necessarily equal, and a system temperature or an effective mass that characterizes the trajectory cannot be defined [39,40]. However, for nonequilibrium systems with Gaussian statistics, an effective kernel $\Gamma(|t|) = \Gamma_R(t)/B = \Gamma_v(|t|)$ can be constructed, which is unique and exactly reproduces the velocity autocorrelation function (VACF) of the nonequilibrium GLE with $\Gamma_R(t) \neq B\Gamma_v(|t|)$ [12,20]. The dynamics of a Gaussian observable are fully determined by its VACF [41], therefore

the effective kernel $\Gamma(t)$ that reproduces the experimental VACF describes the entire observable dynamics. In fact, starting from explicitly time-dependent Hamiltonians, it has been shown recently that for Gaussian observables, a GLE of the form of Eq. (1) with $\Gamma_R(t) = B\Gamma_v(|t|)$ follows [19,20], in line with our approach to describe active cell motion by an effective kernel $\Gamma(t)$.

The kernel $\Gamma(t)$ can be extracted from experimental trajectories via the VACF and is consistent with different combinations of $\Gamma_R(t)$ and $\Gamma_v(t)$ in the nonequilibrium GLE Eq. (1) [12,16] (as shown in the SM [37]). While $\Gamma(t)$ accurately describes the dynamics of the nonequilibrium trajectories, the underlying nonequilibrium model, defined by the functions $\Gamma_v(t)$ and $\Gamma_R(t)$, cannot be determined without additional measurements or assumptions [41]. Thus, caution is required when interpreting the effective model, defined by $\Gamma(t)$, in terms of the underlying motion mechanisms. For instance, the effective kernel $\Gamma(t)$ does not describe the physical time-dependent friction acting on the moving object. A heuristic way of decomposing the effective kernel $\Gamma(t)$ into $\Gamma_v(t)$ and $\Gamma_R(t)$ consists of assuming $\Gamma_v(t)$ to be infinitely short-ranged and proportional to the Stokes friction coefficient [16], which, however, is not pursued further in this paper.

It turns out that for infinitely short-ranged $\Gamma_v(t)$ and for random force correlations $\Gamma_R(t)$ with a negative exponential decay, which corresponds to the active Ornstein-Uhlenbeck model [21], the motion exhibits persistence over times that are substantially longer than the decay time of the effective memory $\Gamma(t)$ [22]. For experimental data where the discretization time is larger than the decay time of the effective memory $\Gamma(t)$, one would then obtain an effective Markovian model with $\Gamma(t) = 2\delta(t)/\tau_m$ and a persistence time τ_m that, due to the nonequilibrium character of the underlying GLE model, cannot be interpreted in terms of the physical mass and friction characterizing the moving object. This turns out to be the case for cancer cells, as we show below.

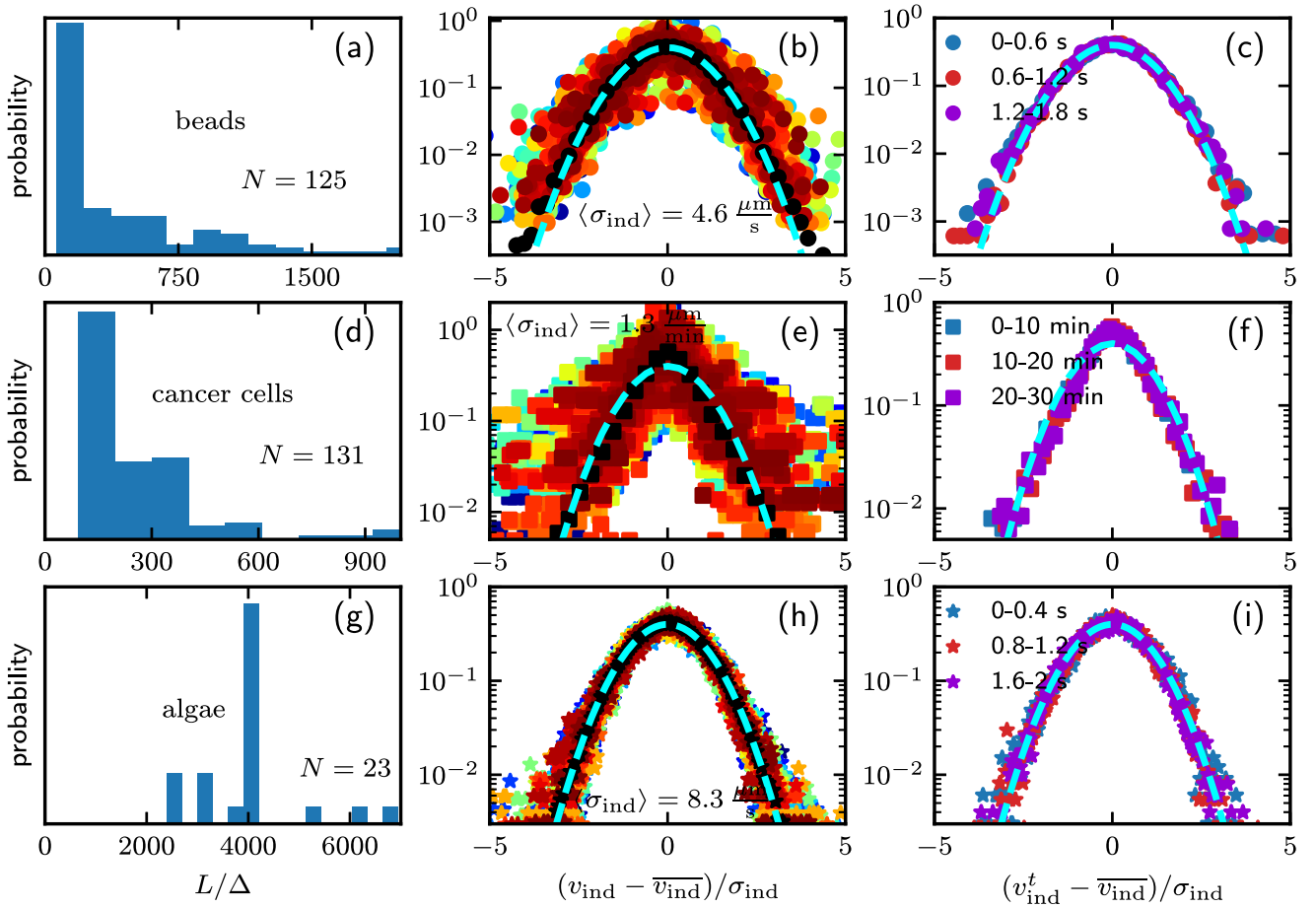


FIG. 2. Distributions of trajectory lengths L in units of the experimental discretization time Δ for (a) polystyrene beads with $\Delta = 0.02$ s, (d) breast-cancer cells with $\Delta = 20$ s, and (g) algae cells with $\Delta = 0.002$ s. Velocity distributions of individual moving objects rescaled by subtracting their individual mean velocity $\bar{v}_{\text{ind}}$ and dividing by their individual standard deviation σ_{ind} for (b) polystyrene beads, (e) breast-cancer cells, and (h) algae cells, where individual cells are distinguished by color, and population mean values $\langle \sigma_{\text{ind}} \rangle$ are given in the plots. The black symbols in (b), (e), and (h) represent the ensemble average over all moving objects, which agrees, except for the cancer cells in (e), very well with the standard normal distribution shown as a dashed blue line. An alternative fit of the cancer-cell data to a Laplace distribution is presented in the SM [37]. The average over individually rescaled velocity distributions is identical for three different time windows for (c) polystyrene beads, (f) breast-cancer cells, and (i) algae cells, which suggests stationarity of the motion.

Since we show below that the GLE is appropriate to model our experimental trajectory data, we can use it to produce synthetic trajectories with the same length distribution as the experimental trajectories and compare them to the experiments. The comparison between synthetic and experimental data with the same trajectory length distributions, which are shown in Figs. 2(a), 2(d), and 2(g), is key in determining the intrinsic variance of parameters characterizing the individual objects.

B. Trajectory analysis

A suitable characterization of stochastic trajectories employs the VACF or the mean-squared displacement (MSD), defined by

$$C_{vv}(t) = \langle v(0)v(t) \rangle, \quad (3)$$

$$C_{\text{MSD}}(t) = \langle [x(0) - x(t)]^2 \rangle, \quad (4)$$

respectively. Even though the VACF is just the curvature of the MSD, $C_{vv}(t) = \frac{1}{2} \frac{d^2}{dt^2} C_{\text{MSD}}(t)$, these correlation functions accentuate different aspects of trajectories, which is important for experimental data, which always contain noise and measurement errors. The correlation functions in Eqs. (3) and (4) for single trajectories are obtained by time-averaging. For the passively moving polystyrene beads, the MSD in Fig. 3(a) demonstrates the purely diffusive nature of the motion over the observation time range, where the ensemble average over the time-averaged MSDs of all beads (solid black line) agrees well with the Brownian prediction $C_{\text{MSD}}(t) = 2Dt$ for a sphere of radius $r = 0.5 \mu\text{m}$ shown as a dashed green line. For this prediction, we use the Einstein relation between the friction of a moving object with its environment γ at temperature T and the diffusivity D , where the friction of a sphere with radius r depends on the solution viscosity η according to

$$D = \frac{k_B T}{\gamma} = \frac{k_B T}{6\pi\eta r}. \quad (5)$$

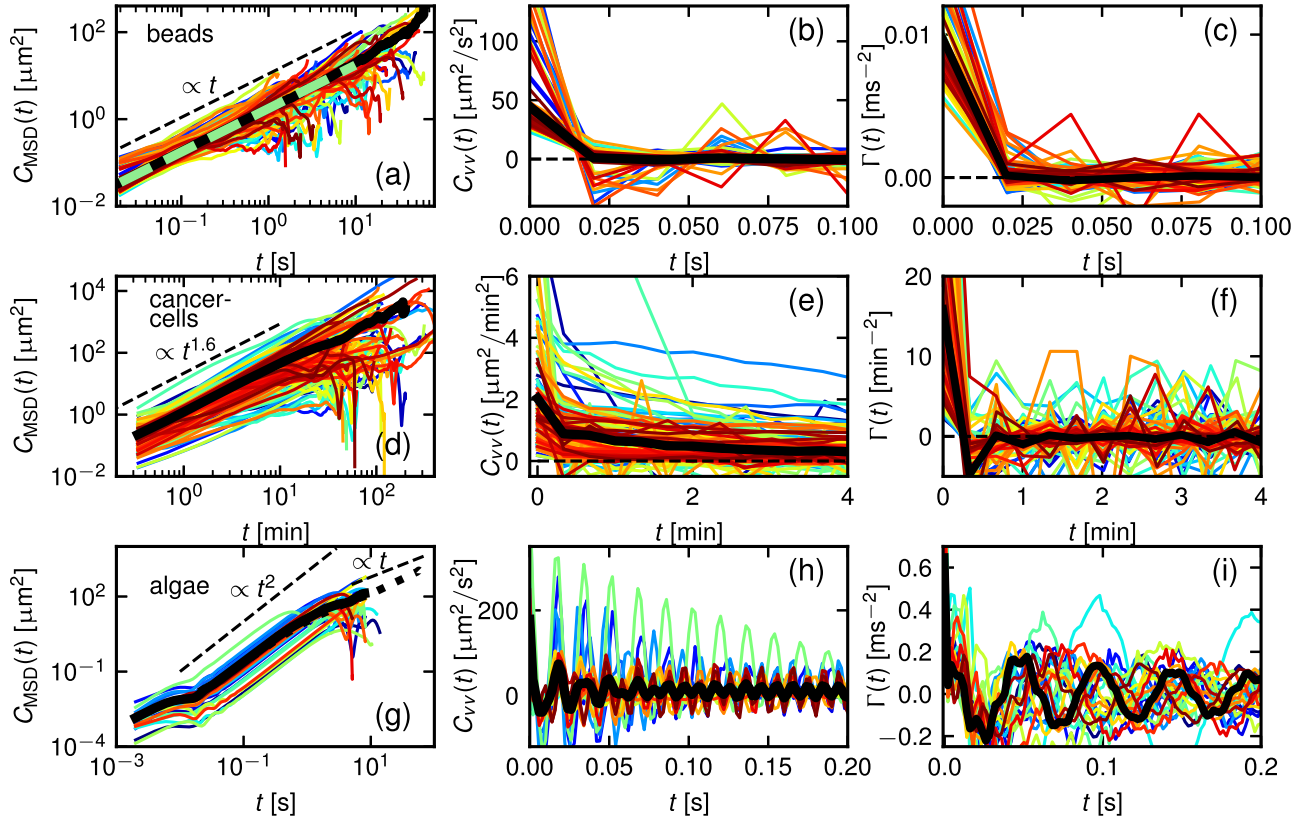


FIG. 3. Results for the time-averaged MSD, $C_{\text{MSD}}(t)$, defined in Eq. (4), the time-averaged VACF, $C_{vv}(t)$, defined in Eq. (3), and the friction kernel, $\Gamma(t)$, extracted from the trajectories for polystyrene beads in (a)–(c), for cancer cells in (d)–(f), and for algae cells in (g)–(i). Different colors represent results for individual moving objects. The black lines in the first two columns denote the ensemble average over all objects. In the third column, the black line denotes the friction kernel calculated from the ensemble-averaged VACF. The dashed green line in (a) is the theoretical prediction for the diffusion of a sphere with radius $r = 0.5 \mu\text{m}$ in water using the Einstein relation Eq. (5). The black dashed lines in the MSD plots indicate the scaling behavior. The thick dotted line in (g) shows the long-time diffusive regime from experimental data recorded at lower spatial resolution and for longer observation times.

Velocities follow from trajectories using finite differences [see Methods, Eq. (11)], which depend on the recording time step Δ and localization noise. The VACF for the polystyrene beads in Fig. 3(b) exhibits a single peak at time zero, which reflects that consecutive displacements are completely uncorrelated, and it shows that the persistence time τ_m is much smaller than Δ . This implies that the friction kernel $\Gamma(t)$ is a δ function, which perfectly matches the extracted kernel shown in Fig. 3(c). We explain the extraction scheme of the friction kernel in detail in the SM [37]. For the actively moving cancer cells, the ensemble-averaged MSD over all cells exhibits superdiffusive behavior $\propto t^{1.6}$ with a slightly decreasing slope for longer times shown in Fig. 3(d), where the long-time diffusive regime is not well resolved due to larger noise (caused by less averaging) for long times. The VACFs of the actively moving cancer cells are decaying on the scale of a few minutes, as shown in Fig. 3(e). Similarly to the polystyrene beads, the friction kernel shown in Fig. 3(f) exhibits a δ peak at time zero, followed by a dip at the first time step $t = \Delta = 20 \text{ s}$ and is then essentially zero. This dip, which is also present for the algae cell kernels at the first time step $t = \Delta = 0.002 \text{ s}$ in Fig. 3(i), originates from localization noise (see the SM for more information [37]), as this noise reduces the correlation of consecutive velocities $v(t)$ and $v(t + \Delta)$, which influences

the extracted kernel [12]. For algae cells, the MSD exhibits an extended ballistic regime up to a few seconds followed by the long-time diffusive regime, as shown in Fig. 3(g), whereas the VACF in Fig. 3(h) shows strong oscillations that correspond to the flagella beat cycle with a frequency of $\sim 50 \text{ Hz}$. The extracted friction kernel $\Gamma(t)$ in Fig. 3(i) shows oscillations as well, but at a slightly different frequency, reflecting the complex dependence of the friction kernel on the VACF [12].

The extracted friction kernels imply different modes of motion. For the algae cells in Fig. 3(i), the kernels exhibit a δ peak at time zero followed by a decaying oscillation, which was previously modeled by [12]

$$\Gamma(t) = 2a\delta(t) + be^{-t/\tau} \left(\cos(\Omega t) + \frac{1}{\tau\Omega} \sin(\Omega t) \right). \quad (6)$$

The extracted kernels of the cancer cells in Fig. 3(f) suggest a simple persistent random walk [42] [note that an earlier study modeled a different cancer cell line by a slightly more complex model including a negative exponential decay in the friction kernel $\Gamma(t)$ [16]]. The persistent random walk is described by the GLE Eq. (1) with

$$\Gamma(t) = 2\delta(t)/\tau_m, \quad (7)$$

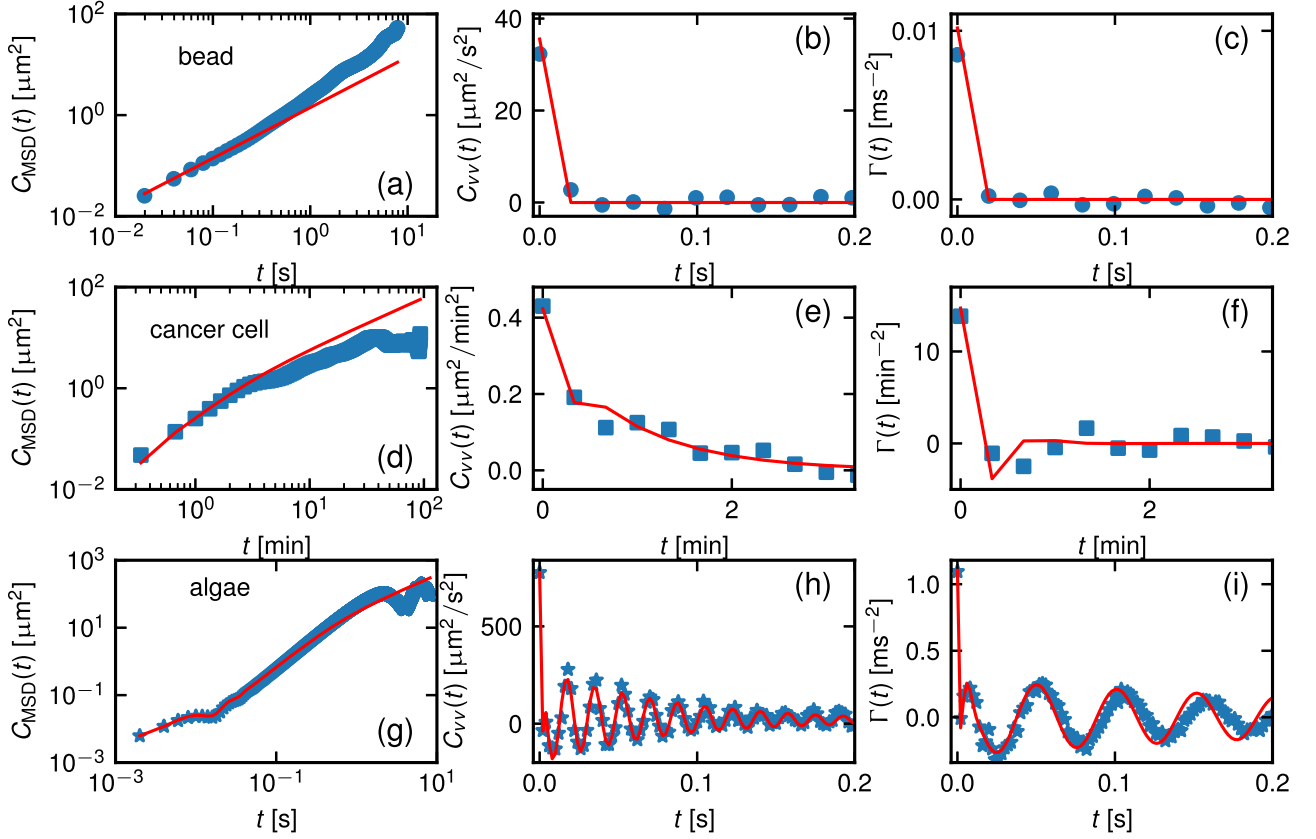


FIG. 4. Experimental results for the MSD, $C_{\text{MSD}}(t)$, VACF, $C_{vv}(t)$, and friction kernel, $\Gamma(t)$, of a single bead (a)–(c), a single cancer cell (d)–(f), and a single algae cell (g)–(i) (blue symbols), all picked at random from the full ensemble of trajectories. The red lines denote predictions of the model defined in Eq. (8) fitted to the bead MSD, of the model defined in Eqs. (1) and (7) fitted to the cancer cell VACF, and of the model defined in Eqs. (1) and (6) fitted to the algae cell VACF. We show corresponding results for individual trajectories that yield the best fits to the models and the distribution of R^2 , which measures the goodness of the fit, for all trajectories in the SM [37].

where the transition from the ballistic to the long-time diffusive regime, defining the persistence time τ_m , occurs for cancer cells within the experimentally observed timescales; the mean value $\tau_m \approx 2$ min is of the order of minutes and thus is much larger than the time step of $\Delta = 20$ s. Therefore, the observed superdiffusive scaling behavior of the ensemble-averaged MSD in Fig. 3(d) is caused by the slow transition from ballistic $\propto t^2$ to the long-time diffusive regime $\propto t$, which is confirmed by fits to the model to single-cell data explained below. In fits to the cancer cells, the persistence time of the polystyrene beads, estimated from the equipartition theorem as $\tau_m = m/\gamma \approx 0.1 \mu\text{s}$, is much smaller than the experimental time resolution $\Delta = 0.02$ s, so that only the long-time diffusive behavior is observed. Hence, we model the polystyrene beads by pure diffusion, in which case the GLE Eq. (1) simplifies to

$$\dot{x}(t) = \sqrt{D}\xi(t), \quad (8)$$

where $\xi(t)$ is uncorrelated white noise with $\langle \xi(t) \rangle = 0$ and $\langle \xi(0)\xi(t) \rangle = 2\delta(t)$.

In Fig. 4 we show fits of the analytical models to the MSD, VACF, and friction kernel data of a single randomly selected moving object, using for the polystyrene bead Eq. (8), for the breast-cancer cell Eqs. (1) and (7), and for the algae cell Eqs. (1) and (6). The GLE describes the data perfectly, which

confirms the applicability of the GLE to describe the active or passive motion of single moving objects. Moreover, our fitting procedure (described in the Methods section) captures localization noise effects, which manifest, for instance, in a local minimum of the friction kernel $\Gamma(t)$ at $t = \Delta$ in Fig. 4(i).

C. Reproducing experimental parameter distributions by simulation

Even though all polystyrene beads are nearly identical, with the standard deviation of the radius being $\Delta r \sim 25$ nm according to the manufacturer and as confirmed by our measurement of the bead radii by atomic force microscopy (explained in the SM), the extracted MSDs, VACFs, and friction kernels of single beads scatter significantly around the mean, as seen in Figs. 3(a)–3(c). Hence, the fitted single-bead parameters show a large spread, shown in Fig. 5(a), which does not reflect the true parameter spread that the beads actually exhibit, as we demonstrate further below.

For living organisms it is often not known how similar individuals of one population are in terms of their genotype or phenotype. Naturally, the question arises: How much of the spread of the parameters extracted from experimental data originates from differences among individuals, and how much comes from the motion stochasticity and experimental errors?

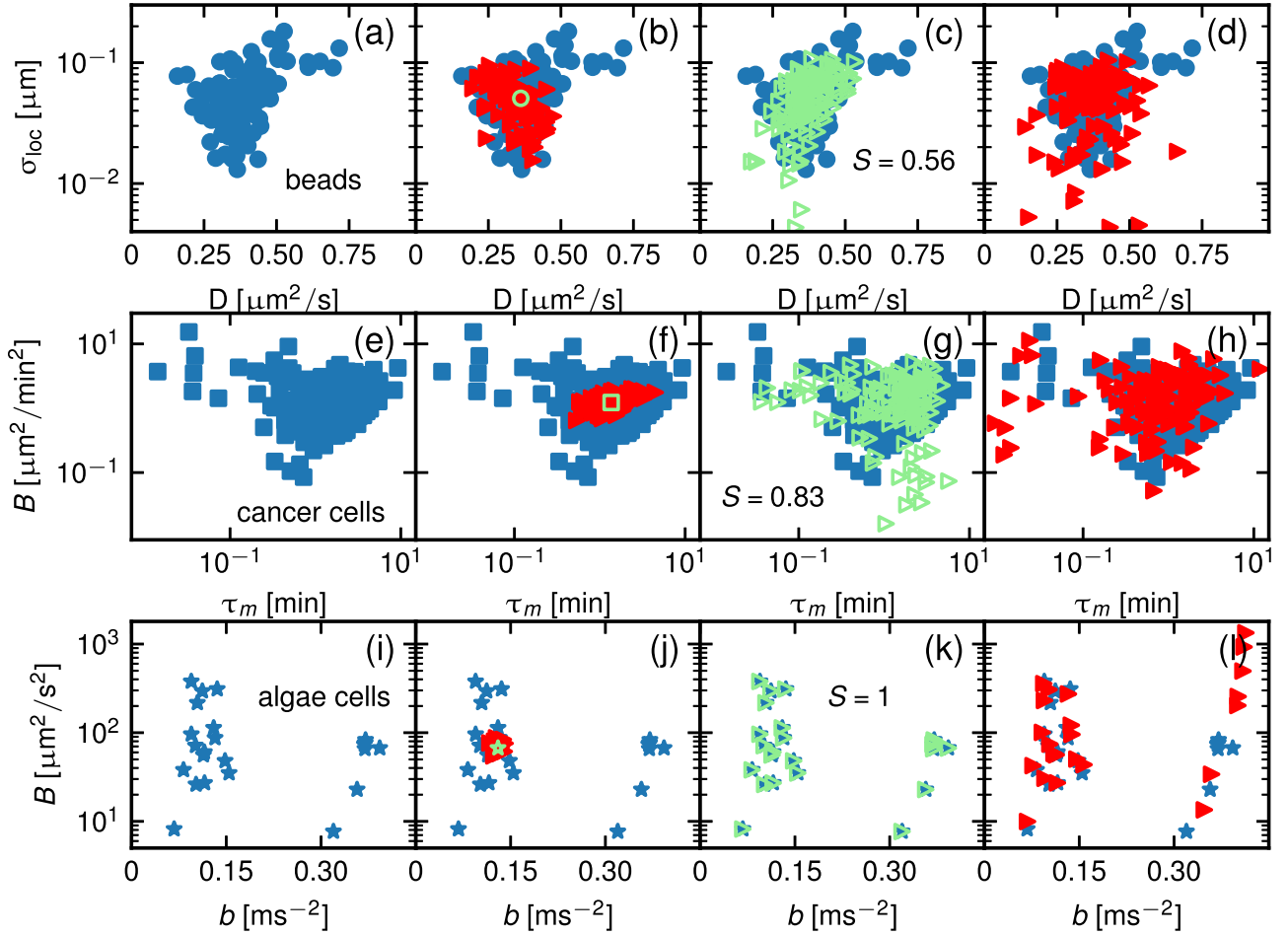


FIG. 5. Distributions of GLE model parameters extracted from the experiments are shown as blue symbols. For polystyrene beads we show the localization noise width σ_{loc} and diffusivity D in (a)–(d), for cancer cells the mean squared velocity B and the persistence time τ_m in (e)–(h), and for algae cells the mean squared velocity B and the kernel oscillation amplitude b [see Eq. (6)] in (i)–(l). The green symbols in (b), (f), and (j) represent the respective median of the distribution of blue symbols, and the red filled triangles represent the parameters extracted from simulations using the median parameter set as input, the experimental time step for analysis, and the experimental trajectory length distribution, shown in Figs. 2(a), 2(d), and 2(g), for the simulation lengths. Empty triangles in (c), (g) denote Gaussian distributions of simulation input parameters with covariance ratios $S = 0.56$ and 0.83 , respectively, which correspond to the estimated covariance ratio S^* shown in Fig. 6(a). In (k) simulation input parameters (green triangles) equal the parameters extracted from the experiment, which corresponds to $S = 1$. Simulations using the empty green triangle parameters in (c), (g), and (k) and the respective model suggested by the data Eqs. (6), (7), and (8) and the experimental trajectory lengths lead to the filled red triangles in (d), (h), and (l), which by definition of S^* have the same covariance as the blue symbols.

To proceed, we imagine the scenario that all individuals are completely identical and are represented by the median of the extracted parameter distribution. Here, we use the median instead of the mean, because a least-squares fit of a noisy exponential function leads to a log-normal distribution of the fitting parameters with the median being the best representation of the true parameter value, as explained in the SM [37]. We simulate the same number of experimentally recorded trajectories, N , with the same trajectory length distribution as in the experiments, shown in Figs. 2(a), 2(d), and 2(g), using the GLE model with the median parameters estimated from the experimental trajectories. GLE simulations with the oscillating kernel Eq. (6) are performed using Markovian embedding techniques [14]. To reach the continuum limit in simulations, we employ a time discretization constant h , which is 20–200 times smaller than the experimental resolution Δ . We then use

every (Δ/h) th point of the simulation to obtain a trajectory with the experimental time resolution Δ and add the localization noise to the trajectories, which is estimated directly by our fit described in the Methods section. From such simulated trajectories with identical parameters, we extract the VACF and perform the same fit that we used to extract the parameters from the experiment. If the spread of the distribution extracted from the N simulations with identical GLE parameters is comparable to the spread of the distribution extracted from the experiment, then it is likely that all moving objects are actually completely identical in their motion parameters.

In fact, the spread of the experimentally determined single breast-cancer cell parameters shown in Fig. 5(e) and of the single algae cell parameters shown in Fig. 5(i) are much larger than the spread of the respective parameters extracted from simulations with identical input parameters, shown in

Figs. 5(f) and 5(j) by the red triangles. This strongly suggests that the ensemble of cancer cells as well as the ensemble of algae cells are not characterized by identical parameters of motion. In contrast, in Fig. 5(b) the spread of the bead parameter distribution extracted from the simulations with identical input parameters is almost as large as the spread of the parameter distribution extracted from the experiment, indicating that most of the observed spread originates from the finite length and the experimental noise, in line with the small variance of the bead radii determined by our atomic force microscopy (AFM) measurement.

The simulation results in Figs. 5(f) and 5(j) suggest that the individual cancer and algae cells are characterized by non-identical GLE parameters within their respective populations. To quantitatively estimate the actual parameter distribution of individual cells, we test whether simulations using input parameters from a Gaussian distribution with a certain covariance Cov^{inp} result in an extracted parameter distribution with a covariance Cov^{sim} that agrees with the covariance of the parameter distribution extracted from the experiment Cov^{exp} . For this, we use a permutation test that compares the covariances Cov^{sim} and Cov^{exp} , as explained in detail in the Methods section.

In short, we denote the summed ratio of the d -dimensional input parameter covariance for the simulation, Cov^{inp} , and the covariance of the distribution extracted from the experiment, Cov^{exp} , as

$$S = d^{-2} \sum_{i,j} \text{Cov}_{ij}^{\text{inp}} / \text{Cov}_{ij}^{\text{exp}}, \quad (9)$$

where the covariance of a distribution of parameter vectors $\vec{z}$ is defined in the standard way as $\text{Cov}_{ij} = \langle (z_i - \langle z_i \rangle)(z_j - \langle z_j \rangle) \rangle$, with z_i the i th component of the d -dimensional vector $\vec{z}$ and $\langle z_i \rangle$ the mean of z_i ; d denotes the number of parameters. We consider input covariances in the range from $S = 0$, which corresponds to all moving objects described by identical input parameters, to $S = 1$, where we use the parameter distribution extracted from the experiment as the simulation input parameters. For intermediate values $0 < S < 1$, simulation input distributions are drawn from Gaussian distributions with covariances that are uniformly rescaled by S . Exemplary simulation input parameter distributions are shown as green symbols in Figs. 5(c), 5(g), and 5(k).

The main result of our method is the value S^* , for which the covariance of the distribution extracted from the experiment Cov^{exp} agrees, according to a statistical test described below, with the covariance of the distributions extracted from simulations Cov^{sim} with the given input covariance $\text{Cov}^{\text{inp}*}$. Hence, we call S^* the estimated covariance ratio and the corresponding input covariance $\text{Cov}^{\text{inp}*}$ the intrinsic parameter covariance of individual objects. Examples of distributions that are extracted from simulations with the green input distributions $\text{Cov}^{\text{inp}*}$ in Figs. 5(c), 5(g), and 5(k) are shown in Figs. 5(d), 5(h), and 5(l) (red triangles), which agree with the respective experimental distribution (blue symbols) according to our statistical test. In brief, the employed statistical test compares the covariance of the experimentally determined parameter covariance Cov^{exp} to the covariance of a parameter

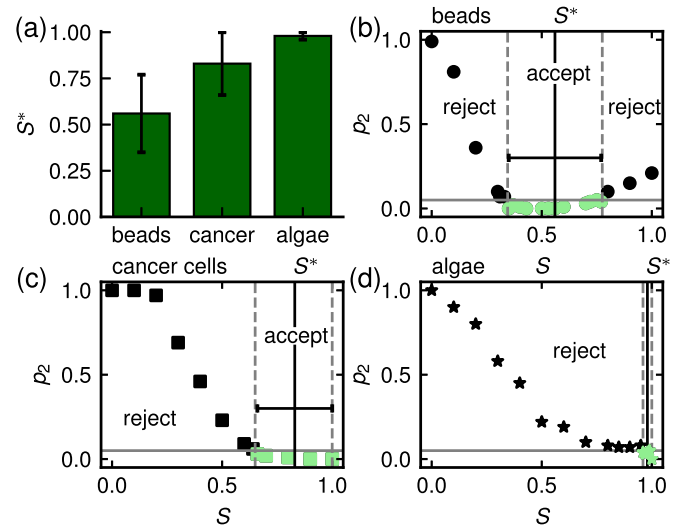


FIG. 6. (a) Estimated covariance ratio S^* of the intrinsic to the experimentally observed parameters as defined in Eq. (9). Likelihood p_2 that the covariance extracted from simulation Cov^{sim} disagrees with the experimental covariance Cov^{exp} according to a permutation test as a function of S (different S originate from different simulation input covariances Cov^{inp}) for (b) beads, (c) cancer cells, (d) algae. Data below the acceptance threshold of $p_2^* = 0.05$ (gray horizontal line) correspond to accepted S values and are shown in green. The acceptance region is indicated by dashed vertical lines, the center defines S^* and the width defines the error bars shown in (a).

distribution extracted from a simulation Cov^{sim} and estimates how likely the two distributions originate from the same input covariance. If several values of S are accepted by our statistical test, meaning that they lead to agreeing covariances Cov^{sim} and Cov^{exp} according to the test, we choose the midpoint of the accepted range as S^* and use the upper and lower acceptance bounds to estimate the error of S^* , as demonstrated in Fig. 6 (details are described in the Methods section).

As an example, the diffusion of polystyrene beads in water is described by the two-dimensional parameter vector comprising the diffusivity D and the localization noise width σ_{loc} , $\vec{z} = (D, \sigma_{\text{loc}})$; the latter defines the uncertainty of the position data (see Methods). If the covariance of a parameter distribution extracted from the experiment was given by $\text{Cov}^{\text{exp}} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ with a median of $\vec{z}^{\text{med}} = (0, 1)$, simulation input parameters for $S = 0.1$ would be drawn from a Gaussian distribution with covariance $\text{Cov}^{\text{inp}} = \begin{pmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{pmatrix}$ and median $\vec{z}^{\text{med}} = (0, 1)$ (here we omit units for clarity). As mentioned earlier, we use the median for generating simulation input parameters because it is a better representation of typical parameters than the mean when using least-squares fits (see the SM [37]). If the parameter distribution covariance Cov^{sim} extracted from the simulation with $S = 0.1$ is the same as the parameter covariance extracted from the experiment Cov^{exp} according to our statistical test, we do the same for $S = 0.2$. Then we repeat the test, increasing S in steps of 0.1, up to a final value of $S = 1$. In the regime of accepted S values, we use smaller steps to determine S^* . For accepted values in the range $S = 0.05$ – 0.15 , we would estimate $S^* = 0.1 \pm 0.05$

for the covariance ratio. The evaluation of S^* is graphically depicted in Figs. 6(b)–6(d).

For the polystyrene beads, we find $S^* = 0.56 \pm 0.21$ as shown in Fig. 6(a), which translates into an estimate for the standard deviation of the bead radius $\Delta r = 86 \pm 63$ nm via the Einstein law Eq. (5) and using the viscosity of water at room temperature $\eta \approx 1$ mPs. The bead radius standard deviation of $\Delta r \sim 25$ nm given by the manufacturer and confirmed by our AFM measurements is close to the lower bound of our result. Nevertheless, it should be noted that other experimental factors, such as the bead-surface properties, contribute to the diffusivity variance between individual beads, thereby increasing the variance of the hydrodynamic radius Δr , which explains our relatively large estimate of Δr .

For the cancer cells, the estimated covariance ratio is given by $S^* = 0.83 \pm 0.17$ shown in Fig. 6(a), which indicates a spread over nearly two orders of magnitude of the intrinsic squared cell speed B and the intrinsic persistence time τ_m , as demonstrated by the green triangles in Fig. 5(g), even though all cells are from the same cell line and supposedly share the same genetic code. This demonstrates a surprisingly large phenotype variance among cancer cells of the same cell line. The distribution extracted from the algae trajectories represents the true cell-to-cell variance accurately, as seen by comparing Figs. 5(k) and 5(l), and as reflected by the large estimated covariance ratio $S^* = 0.98 \pm 0.02$ shown in Fig. 6(a).

In addition to extracting the intrinsic spread of parameters characterizing the motion of individual objects, our approach allows us to estimate how long experimental recordings have to be in order to obtain a certain level of S^* . The variance contribution due to the stochasticity of the motion decreases inversely proportional to the trajectory length, as explained in the SM [37,43]. Therefore, if stochasticity of the motion is the main contribution to the variance, doubling the length of all trajectories of a given population of moving objects leads to an increased S_{new}^* determined by $1 - S_{\text{new}}^* = (1 - S^*)/2$, and thus $S_{\text{new}}^* = S^* + (1 - S^*)/2$. In the case of polystyrene beads, for instance, this would mean that for experimental trajectories twice as long as in the actual experiment, the estimated covariance ratio would be expected to increase from $S^* = 0.53$ to roughly $S_{\text{new}}^* = 0.76$ if the spread $1 - S^*$ stems mainly from stochasticity, i.e., if the influence of other experimental factors on the parameter spread is negligible.

III. DISCUSSION AND CONCLUSIONS

We use the GLE Eq. (1) to describe active cell motion as well as passive motion of polystyrene beads and extract all GLE parameters on the single trajectory level. The spread of the parameter distributions extracted from the experiments is due to a combination of intrinsic differences among the individual moving objects, the finite trajectory length, and noise in the experiment. Our approach determines how much of the extracted total parameter spread actually originates from intrinsic differences between individuals of the observed population and how much stems from the experimental finite observation time. On top of that, it allows us to estimate how long experimental trajectories have to be in order to reach a certain accuracy level of the experimentally extracted

parameter distributions. Our method readily includes experimental factors such as localization noise, discrete time steps, and different trajectory lengths. We confirm the robustness of our model by applying it to data of passively moving polystyrene beads in water and to synthetic data sets with known origin in the SM [37].

Compared to the polystyrene beads, which are relatively similar to each other, the cancer cells and algae cells show significant differences within the studied populations. For the motion of the breast-cancer cells, the value of $S^* = 0.83$ means that most of the observed spread in the mean squared velocity B and in the persistence time τ_m originates from sizable intrinsic differences on the cell-to-cell level; this is noteworthy, as these cells carry the same genetic material. Such phenotypic differences were previously observed [44] and are connected to different levels of expressed proteins in the cells [45]. For the algae cells, we conclude from the value $S^* = 0.98$ that the observed GLE parameter distribution almost entirely reflects intrinsic cell-to-cell differences, which is part of the reason why the observed GLE parameters could previously be directly used to distinguish different swimming styles of these confined algae cells on the single cell level [12].

In principle, one would expect the localization noise width σ_{loc} of rigid objects of identical size to have a rather small variance. The estimated spread of the localization noise width of the polystyrene beads in Fig. 5(a), however, is of the same order of magnitude as the mean. This comes partly from size differences between beads and presumably also from so-called motion blur, which is an additional localization error due to the finite illumination time in every microscopy image, during which the observed object is moving. Motion blur increases the variance of localization noise widths extracted from MSDs [46], which in turn increases the estimated variance of the localization noise width. To separate the effects of motion blur and other localization noise, one could add the effect of motion blur to the synthetic trajectories and in the fit of the correlation function, which is left for future work.

The usage of the GLE Eq. (1) to describe active systems relies on the Gaussianity of the underlying motion process [12,20]. In the hypothetical case of non-Gaussian single-cell velocity distributions, one could use filter approaches to remove non-Gaussian components in the data [47] and then apply the same methodology we describe in this paper. Alternatively, for non-Gaussian systems one could use the GLE in a heuristic manner or use previously introduced models to describe the organism motion [23,38,48–51]. While we show the GLE model with time-independent parameters to reproduce the experimental trajectories of single algae cells perfectly, we cannot exclude that over much longer, experimentally inaccessible times, the cells would exhibit time-dependent motion parameters. Such questions might be addressable in the future with extended experimental trajectories. Since the GLE Eq. (1) is valid for general observables, our approach can be applied to time-series data characterizing any organism or object of choice—be it active or passive.

In summary, our approach allows us to disentangle the intrinsic individual-to-individual parameter variance of moving objects from noise and stochastic effects. As every distribution of parameters extracted from experiments inevitably is noisy due to finite recording time, we anticipate various applications

of our method in infection biology, ecology, medicine, and all other fields that rely on the knowledge of the heterogeneity of individual moving objects. We mention in the end that our method is also applicable to time-series data of general observables and thus is not restricted to positional trajectories.

IV. METHODS, EXPERIMENTAL DETAILS, AND CELL PREPARATION

A. Statistical test

To compare two d -dimensional distributions $\rho^{\text{exp}}(\vec{z})$ and $\rho^{\text{sim}}(\vec{z})$, which are determined by a finite number of observations $\vec{z}_i^{\text{exp}}$ for ρ^{exp} and $\vec{z}_i^{\text{sim}}$ for ρ^{sim} , we use a so-called permutation test [52]. We define the distance measure

$$T = \sum_{i \leq j}^d (\text{Cov}_{ij}^1 - \text{Cov}_{ij}^2)^2, \quad (10)$$

which measures the difference between two distributions in terms of their covariances Cov^1 and Cov^2 . We denote the value of T for the distributions $\rho^{\text{exp}}(\vec{z})$ and $\rho^{\text{sim}}(\vec{z})$ by T_0 .

The statistical test starts by randomly drawing samples from the joint pool of observations $\vec{z}_i^{\text{exp}}$ and $\vec{z}_i^{\text{sim}}$, which leads to the exchanged distributions $\rho_{\text{exch}}^{\text{exp}}(\vec{z})$ and $\rho_{\text{exch}}^{\text{sim}}(\vec{z})$. The exchanged distributions exhibit a new value of T called T_{exch} . This process of sampling exchanged distributions from the joint pool of observations is repeated $n_{\text{exch}} = 10^4$ times, leading to 10^4 values of T_{exch} . Comparing T_0 to T_{exch} leads to the probability of rejecting the hypothesis that the two distributions have the same covariance, given that the hypothesis is actually true, as $p_1 = n_{\text{rej}}/n_{\text{exch}}$, with n_{rej} being the number of exchanges for which $T_{\text{exch}} > T_0$ [52]. We use the acceptance level $p_1^* = 0.05$, meaning we accept the hypothesis that the distributions $\rho^{\text{exp}}(\vec{z})$ and $\rho^{\text{sim}}(\vec{z})$ have identical covariances when $p_1 \leq p_1^*$.

In practice, we extract $M = 100$ distributions $\rho^{\text{sim}}(\vec{z})$ from simulations and compare each to the experimental distribution using the statistical test described above. This means we simulate a total of $M \times N$ trajectories. We define $p_2 = M_{\text{rej}}/M$ as the ratio of rejected distributions M_{rej} , where $p_1 > p_1^*$, to the total number of simulated distributions M . The hypothesis that the covariance from simulation Cov^{sim} originates from the same intrinsic covariance $\text{Cov}^{\text{inp}*}$ as the experimental covariance Cov^{exp} is accepted as true, if $p_2 \leq p_2^*$, where p_2^* is another acceptance level. One expects the hypothesis to be rejected on average $p_1^* \times M$ out of M times, if it is true. Therefore, we set the acceptance level of p_2 as $p_2^* = p_1^* = 0.05$. Results for p_2 depending on S are shown in Figs. 6(b)–6(d) for different moving objects.

B. Fit to discretized data including localization noise

Every experimental recording of trajectories inevitably includes noise. In the case of cell-center trajectories, the finite camera pixel size as well as thermal and electronic fluctuations in the camera sensors are just a few of many possible origins of localization noise present in the trajectories. Therefore, our fit to the experimental VACF of individual trajectories needs to account for finite time discretization and

noise [16]. Here, we explain our fitting procedure. We denote discrete values of a function $f(t)$ as $f(i\Delta) = f_i = f^i$ and the discretization time step as Δ . Velocities are computed at half time steps as

$$v_{i+\frac{1}{2}} = \frac{x_{i+1} - x_i}{\Delta}. \quad (11)$$

From the discrete velocities, the VACF defined by Eq. (3) is computed according to

$$C_{vv}^i = \frac{1}{n-i} \sum_{j=0}^{n-i-1} v_{j+\frac{1}{2}} v_{j+i+\frac{1}{2}}, \quad (12)$$

with n being the the number of trajectory steps. To account for localization noise, we add Gaussian uncorrelated noise of width σ_{loc} at every time step to the trajectory, which leads to the noisy MSD as [16]

$$C_{\text{MSD}}^{\text{noise}}(t) = C_{\text{MSD}}^{\text{theo}}(t) + 2(1 - \delta_{t0})\sigma_{\text{loc}}^2, \quad (13)$$

where $C_{\text{MSD}}^{\text{theo}}(t)$ is the theoretical expression for the model MSD given by Eqs. (17), (18), and (20) for the different models Eqs. (6), (7), and (8), and δ_{t0} is the Kronecker delta reflecting the uncorrelated nature of the localization noise. Since the observed trajectories are sampled with a finite time step Δ , we discretize the relation

$$C_{vv}(t) = \frac{1}{2} \frac{d^2}{dt^2} C_{\text{MSD}}(t), \quad (14)$$

which leads to

$$\begin{aligned} C_{vv}^{\text{fit}}(i\Delta) &= \frac{C_{\text{MSD}}^{\text{noise}}[(i+1)\Delta] - 2C_{\text{MSD}}^{\text{noise}}(i\Delta) + C_{\text{MSD}}^{\text{noise}}[(i-1)\Delta]}{2\Delta^2}. \end{aligned} \quad (15)$$

Finally, fits are performed by minimizing the cost function

$$E_{\text{cost}} = \sum_{i=0}^{n-1} [C_{vv}^{\text{exp}}(i\Delta) - C_{vv}^{\text{fit}}(i\Delta)]^2 \quad (16)$$

with `scipy`'s least-squares function in `python` and using Eq. (15) to determine the parameters of $C_{vv}^{\text{fit}}(t)$. As the MSD and VACF follow from the GLE Eq. (1) and the friction kernel in Eqs. (6) and (7) or the LE in Eq. (8), the parameters to be optimized are the kernel parameters, together with the mean squared velocity B and the localization noise width σ_{loc} . Resulting fits are shown in Figs. 4(e) and 4(h), where the data are fitted up to 3.3 min for the cancer cells and up to 0.2 s for the algae cells in order to disregard the noisy part of the VACF at longer times. For the polystyrene beads, which perform purely diffusive motion, we use Eq. (13) to fit the experimental MSD as the VACF is only nonzero for the first two time steps. A resulting fit is shown in Fig. 4(a), where the MSD is fitted up to 0.2 s. The time range over which we perform the fit corresponds to roughly 3% of the mean trajectory length for the different moving objects, for which the time-averaged correlation functions exhibit low errors [43]. For consistency, we use the same fitting range for the synthetic data.

Before we fit the data on the individual trajectory level, we estimate the order of magnitude of the mean squared velocity B by $B_{\text{est}} = C_{vv}^0$ for the cancer cells and algae cells. For the

freely diffusing beads, we estimate the diffusion coefficient D by the integral over the VACF, which leads to $D_{\text{est}} = C_{vv}^0 \frac{\Delta}{2}$. Furthermore, for the cancer cells we estimate the order of magnitude of the persistence time τ_m by $\tau_m^{\text{est}} = \frac{2}{\Delta \Gamma_0}$. We apply the following boundaries for the optimization of cancer cell parameters: B is constrained between $B_{\text{est}}/4$ and $4B_{\text{est}}$, τ_m between $\tau_m^{\text{est}}/4$ and $4\tau_m^{\text{est}}$, and the localization noise width σ_{loc} between 0.01 and 1 μm . For the algae cells, B is assumed to lie between $B_{\text{est}}/3$ and $3B_{\text{est}}$, the amplitudes a and b are assumed to lie between $0.01\Gamma_0$ and Γ_0 , the memory time τ is bound between 0.05 and 5 s, the kernel frequency Ω is between 15 and 250 Hz, and the localization noise width σ_{loc} is between 0 and 0.05 μm . Further, for polystyrene beads we use bounds of $4D_{\text{est}}$ and $D_{\text{est}}/4$ for the diffusivity D , and the localization noise width σ_{loc} is constrained between 0.001 and 0.2 μm . Finally, we discard fit results for which the localization noise width σ_{loc} is close to the boundaries and likely to be an unphysical fit result, when σ_{loc} is smaller than 1.5 times the lower boundary or larger than 0.99 times the upper boundary. This leads to four discarded trajectories out of 135 total trajectories for the breast-cancer cells, no discarded trajectories for the 23 algae cells, and 19 discarded trajectories out of 144 for the polystyrene beads. For the diffusion of the polystyrene beads, we additionally discard diffusion constants above $0.8 \mu\text{m}^2/\text{s}$, which is roughly twice the theoretical value of a sphere with radius $r = 0.5 \mu\text{m}$ according to the Stokes-Einstein law Eq. (5), which leads to four additionally discarded trajectories.

C. MSD expressions

The diffusion coefficient defined in Eq. (5) determines the behavior of the MSD for pure diffusion as

$$C_{\text{MSD}}(t) = 2Dt. \quad (17)$$

In the case of the persistent random walk, the MSD takes the form

$$C_{\text{MSD}}(t) = 2B\tau_m(t - \tau_m(1 - e^{-t/\tau_m})), \quad (18)$$

where the diffusion constant is given by $D = B\tau_m$ and the VACF follows a single exponential decay,

$$C_{vv}(t) = Be^{-t/\tau_m}. \quad (19)$$

For the oscillating memory kernel Eq. (6), the MSD takes the form [12]

$$C_{\text{MSD}}(t) = \frac{B}{\tau^4} \left(\frac{k_1 t}{\omega_1^2 \omega_2^2 \omega_3^2} + \sum_{i=1}^3 \frac{e^{-\sqrt{-\omega_i^2} t} - 1}{\sqrt{-\omega_i^2} \prod_{j \neq i} (\omega_i^2 - \omega_j^2)} \right) \times \left[\frac{k_1}{\omega_i^2} + k_2 + k_3 \omega_i^2 \right], \quad (20)$$

where the constants ω_i and k_i have been previously calculated [12].

D. Polystyrene beads

Materials. Fluorescently labeled polystyrene beads (diameter: 1000 nm, excitation wavelength: 505 nm, emission wavelength: 515 nm) were purchased from Thermo Fisher

Scientific (order number: F8776). For single-particle tracking experiments, the beads are diluted 1:1000 in deionized water and sonicated for five minutes at room temperature. Afterwards, 8 μL of bead-containing solution is injected into polydimethylsiloxane (PDMS) microwells, which are formed on glass cover slips as described previously [53].

Imaging. Single-particle tracking of the fluorescently labeled polystyrene beads is performed using a Nikon Ti-E Eclipse fluorescence microscope (Nikon, Düsseldorf, Germany), which is equipped with focus stabilization, a white-light source (Prior Lumen 200; Prior Scientific, Cambridge, UK), and an Andor Zyla 4.2 sCMOS camera (Oxford Instruments, Oxford, UK). Fluorescence excitation and emission are controlled using a GFP filter set (Nikon GFP-1828A; EX 482/18, DM 495, BA 520/28). The beads are imaged using a 100 $\times$ Plan-Apo oil immersion objective (numerical aperture: 1.45) and the following setting of the sCMOS camera: 2 $\times$ 2 binning, 10 ms exposure time, acquisition rate of 50.41 frames per second at a field of view of 1024 $\times$ 1024 pixel², corresponding to 133.12 $\times$ 133.12 μm^2 .

Image analysis. Single-particle tracking analysis is performed as previously described [54]. In brief, the beads are detected by a local maximum of the fluorescence intensity and only considered for further analysis if the corresponding intensity value exceeds a user-defined threshold, which is chosen slightly above the level of readout noise of the sCMOS camera. The center position of each detected bead is determined by fitting a two-dimensional Gaussian distribution to its intensity distribution. Trajectories are generated by a nearest-neighbor linking scheme involving a distance threshold as described previously [55]. Finally, as convection of the aqueous solution (contained in the PDMS wells) cannot be ruled out even for these relatively small wells, the bead trajectories are corrected for the potential occurrence of convection-based distortions of bead motion. Convection affects the displacements of all beads in the same way, and there is no further source of motional correlation for these randomly moving beads, except for hydrodynamic interactions, which are weak for the high dilution of beads employed here [56]. The convection is determined (i) by calculating the displacements of all beads between consecutive frames, followed (ii) by calculating the average value of all obtained two-dimensional displacement vectors, and finally (iii) by integrating all average displacement vector values from the first to the last frame (whereas the average displacement vector between the first two frames is set to zero). This running integral yields an estimate for the convection-based motion of the aqueous solution, which we subtract from the position of the beads to correct the trajectories for convection.

E. Breast-cancer cells

Micropatterning. The human breast-cancer cells of the cell line MDA-MBA-231 in our experiments move on one-dimensional lanes that are produced by coating with fibronectin (FN). We transfer fibronectin (FN) (YO Proteins, Ronning, Sweden) to an imaging dish featuring a polymer coverslip bottom (ibidi, Gräfelfing, Germany) using polydimethylsiloxane (PDMS) stamps with a 15- μm -wide lane pattern. The microcontact printing protocol, including

the fabrication of the PDMS stamps, has been detailed previously [57].

Cell culture. We culture MDA-MB-231 cells that had been stably transduced with histone-2B mCherry (gifted from Timo Betz, University of Göttingen, Germany) in Leibovitz's CO₂-buffered L-15 medium with 2 mM Glutamax (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C. The growth medium is supplemented by 10% fetal bovine serum (Thermo Fisher), and cells are passaged every 2–3 days using Accutase (Thermo Fisher). For experiments, we seed about 5000 cells per dish. After incubation for 2–3 h, cells adhere to the micropatterns and we exchange the medium.

Microscopy. The microscopy images [see, for instance, Fig. 1(b)] originate from time-lapse imaging on an inverted fluorescence microscope (Nikon Eclipse Ti, Nikon, Tokyo, Japan) outfitted with an XY-motorized stage, Perfect Focus System (Nikon), and a heating chamber (Okolab, Pozzuoli, Italy) maintained at 37 °C. We set up an acquisition protocol to sequentially scan and image fields of view employing the motorized stage, the Perfect Focus System, a 10 CFI Plan Fluor DL objective (Nikon), a CMOS camera (PCO edge 4.2, Excelitas PCO, Kelheim, Germany), and the acquisition software NIS Elements (Nikon). Prior to the time-lapse measurement, we obtain epifluorescence images of the FN patterns. Subsequently, we capture phase-contrast images of the cells and epifluorescence images of their nuclei at 20 s intervals for a total of 48 h.

Image analysis. To obtain trajectories from the microscopy data of the fluorescently labeled cell nuclei, we employ MATLAB [58] scripts built upon previous work [30] for image analysis. Here, the geometric center of the fluorescently labeled nucleus is used as the cell position. Detection of FN lanes is accomplished through a Hough transformation of the fluorescence signal from labeled FN. Tracking of nuclei positions involves setting a threshold after applying a back-

ground correction and a band-pass filter to fluorescent images of the nuclei. The position of the nuclei is adjusted to ensure that the position aligns with the FN lanes such that nuclei center positions cannot be extracted offside of the FN lane.

F. Algae cells

Wild-type *Chlamydomonas reinhardtii* cells are recorded in two-dimensional confinement between two anti-adhesively coated glass plates with a separation of roughly 10 μm by high-speed video microscopy (Olympus IX83/IX73) at 500 frames per second with a 40× phase-contrast objective (Olympus, 0.65 NA, Plan N, PH2) connected to a metal oxide semiconductor (CMOS) camera (Phantom Miro C110, Vision Research, pixel size = 5.6 μm). Trajectories of the algae cells are extracted by binarizing the image sequences with appropriate threshold parameters and tracking their centers using standard MATLAB routines [58]. More experimental details are given in [12,32]. The *Chlamydomonas reinhardtii* cells were previously shown to exhibit two distinct swimming styles [12]; only data of cells with synchronous flagella motion are considered in this work. The studied algae cells originate from multiple isolated colonies and hence are presumably genetically heterogeneous.

ACKNOWLEDGMENTS

We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) through Grant No. CRC 1449 “Dynamic Hydrogels at Biointerfaces,” Project ID 431232613, Project A03. D.M. and P.S. acknowledge support from the DBT/Wellcome Trust India Alliance Fellowship (grant number IA/I/16/1/502356) awarded to P.S.

DATA AVAILABILITY

The data that support the findings of this article are openly available [59].

- [1] E. Levien, J. Min, J. Kondev, and A. Amir, Non-genetic variability in microbial populations: survival strategy or nuisance? *Rep. Prog. Phys.* **84**, 116601 (2021).
- [2] M. M. Conner and G. C. White, Effects of individual heterogeneity in estimating the persistence of small populations, *Nat. Res. Model.* **12**, 109 (1999).
- [3] G. A. Fox, B. E. Kendall, J. W. Fitzpatrick, and G. E. Woolfenden, Consequences of heterogeneity in survival probability in a population of Florida scrub-jays, *J. Anim. Ecol.* **75**, 921 (2006).
- [4] R. Avraham, N. Haseley, D. Brown, C. Penaranda, H. B. Jijon, J. J. Trombetta, R. Satija, A. K. Shalek, R. J. Xavier, A. Regev *et al.*, Pathogen cell-to-cell variability drives heterogeneity in host immune responses, *Cell* **162**, 1309 (2015).
- [5] A. E. Fleming-Davies, V. Dukic, V. Andreassen, and G. Dwyer, Effects of host heterogeneity on pathogen diversity and evolution, *Ecol. Lett.* **18**, 1252 (2015).
- [6] Asadullah, S. Kumar, N. Saxena, M. Sarkar, A. Barai, and S. Sen, Combined heterogeneity in cell size and deformability promotes cancer invasiveness, *J. Cell Sci.* **134**, jcs250225 (2021).
- [7] W. C. Aird, Endothelial cell heterogeneity, *Crit. Care Med.* **31**, S221 (2003).
- [8] R. Rigler, A. Pramanik, P. Jonasson, G. Kratz, O. Jansson, P.-Å. Nygren, S. Ståhl, K. Ekberg, B.-L. Johansson, S. Uhlen *et al.*, Specific binding of proinsulin c-peptide to human cell membranes, *Proc. Natl. Acad. Sci. USA* **96**, 13318 (1999).
- [9] E. R. Jerison and S. R. Quake, Heterogeneous T cell motility behaviors emerge from a coupling between speed and turning in vivo, *Elife* **9**, e53933 (2020).
- [10] S. Studenski, S. Perera, K. Patel, C. Rosano, K. Faulkner, M. Inzitari, J. Brach, J. Chandler, P. Cawthon, E. B. Connor *et al.*, Gait speed and survival in older adults, *Jama* **305**, 50 (2011).
- [11] P. Maiuri, J.-F. Rupprecht, S. Wieser, V. Rupprecht, O. Bénichou, N. Carpi, M. Coppey, S. De Beco, N. Gov, C.-P. Heisenberg *et al.*, Actin flows mediate a universal coupling between cell speed and cell persistence, *Cell* **161**, 374 (2015).
- [12] A. Klimek, D. Mondal, S. Block, P. Sharma, and R. R. Netz, Data-driven classification of individual cells by their non-Markovian motion, *Biophys. J.* **123**, 1173 (2024).

- [13] F. N. Brünig, J. O. Daldrop, and R. R. Netz, Pair-reaction dynamics in water: competition of memory, potential shape, and inertial effects, *J. Phys. Chem. B* **126**, 10295 (2022).
- [14] F. N. Brünig, O. Geburtig, A. v. Canal, J. Kappler, and R. R. Netz, Time-dependent friction effects on vibrational infrared frequencies and line shapes of liquid water, *J. Phys. Chem. B* **126**, 1579 (2022).
- [15] B. A. Dalton, H. Kiefer, and R. R. Netz, The role of memory-dependent friction and solvent viscosity in isomerization kinetics in viscogenic media, *Nat. Commun.* **15**, 3761 (2024).
- [16] B. G. Mitterwallner, C. Schreiber, J. O. Daldrop, J. O. Rädler, and R. R. Netz, Non-Markovian data-driven modeling of single-cell motility, *Phys. Rev. E* **101**, 032408 (2020).
- [17] C. Ayaz, L. Tepper, F. N. Brünig, J. Kappler, J. O. Daldrop, and R. R. Netz, Non-Markovian modeling of protein folding, *Proc. Natl. Acad. Sci. USA* **118**, e2023856118 (2021).
- [18] B. A. Dalton, C. Ayaz, H. Kiefer, A. Klimek, L. Tepper, and R. R. Netz, Fast protein folding is governed by memory-dependent friction, *Proc. Natl. Acad. Sci. USA* **120**, e2220068120 (2023).
- [19] B. J. A. Héry and R. R. Netz, Derivation of a generalized Langevin equation from a generic time-dependent Hamiltonian, *J. Phys. A* **57**, 505003 (2024).
- [20] R. R. Netz, Derivation of the nonequilibrium generalized Langevin equation from a time-dependent many-body Hamiltonian, *Phys. Rev. E* **110**, 014123 (2024).
- [21] D. Martin, J. O'Byrne, M. E. Cates, É. Fodor, C. Nardini, J. Tailleur, and F. van Wijland, Statistical mechanics of active Ornstein-Uhlenbeck particles, *Phys. Rev. E* **103**, 032607 (2021).
- [22] B. G. Mitterwallner, L. Lavacchi, and R. R. Netz, Negative friction memory induces persistent motion, *Eur. Phys. J. E* **43**, 67 (2020).
- [23] G. M. Viswanathan, M. G. Da Luz, E. P. Raposo, and H. E. Stanley, *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters* (Cambridge University Press, Cambridge, 2011).
- [24] T. K. Moon, The expectation-maximization algorithm, *IEEE Sign. Proc. Mag.* **13**, 47 (1996).
- [25] A. M. Law, W. D. Kelton, and W. D. Kelton, *Simulation Modeling and Analysis* (McGraw-Hill, New York, 2007), Vol. 3.
- [26] R. R. Barton and L. W. Schruben, Resampling methods for input modeling, in *Proceeding of the 2001 Winter Simulation Conference (Cat. No. 01CH37304)* (IEEE, Piscataway, NJ, 2001), Vol. 1, pp. 372–378.
- [27] A. B. Massada and Y. Carmel, Incorporating output variance in local sensitivity analysis for stochastic models, *Ecol. Modell.* **213**, 463 (2008).
- [28] J. E. Straub, M. Borkovec, and B. J. Berne, Dynamic friction on intramolecular degrees of freedom, *J. Phys. Chem.* **91**, 4995 (1987).
- [29] B. J. Berne, M. E. Tuckerman, J. E. Straub, and A. L. R. Bug, Friction on rigid and flexible bonds, *J. Chem. Phys.* **93**, 5084 (1990).
- [30] C. Schreiber, B. Amiri, J. C. Heyn, J. O. Rädler, and M. Falcke, On the adhesion–velocity relation and length adaptation of motile cells on stepped fibronectin lanes, *Proc. Natl. Acad. Sci. USA* **118**, e2009959118 (2021).
- [31] B. Amiri, J. C. Heyn, C. Schreiber, J. O. Rädler, and M. Falcke, On multistability and constitutive relations of cell motion on fibronectin lanes, *Biophys. J.* **122**, 753 (2023).
- [32] D. Mondal, A. G. Prabhune, S. Ramaswamy, and P. Sharma, Strong confinement of active microalgae leads to inversion of vortex flow and enhanced mixing, *eLife* **10**, e67663 (2021).
- [33] H. Mori, Transport, collective motion, and Brownian motion, *Prog. Theor. Phys.* **33**, 423 (1965).
- [34] R. Zwanzig, Memory effects in irreversible thermodynamics, *Phys. Rev.* **124**, 983 (1961).
- [35] C. Ayaz, L. Scalfi, B. A. Dalton, and R. R. Netz, Generalized Langevin equation with a nonlinear potential of mean force and nonlinear memory friction from a hybrid projection scheme, *Phys. Rev. E* **105**, 054138 (2022).
- [36] T. Indei, J. D. Schieber, A. C. Cordóba, and E. Pilyugina, Treating inertia in passive microbead rheology, *Phys. Rev. E* **85**, 021504 (2012).
- [37] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/77t1-2qvv> for additional analysis of experimental data, derivations of formulas, and test of the methods used for determining the covariance ratio of the intrinsic to the experimentally observed parameters.
- [38] A. V. Chechkin, F. Seno, R. Metzler, and I. M. Sokolov, Brownian yet non-Gaussian diffusion: from superstatistics to subordination of diffusing diffusivities, *Phys. Rev. X* **7**, 021002 (2017).
- [39] D. Mizuno, C. Tardin, C. F. Schmidt, and F. C. MacKintosh, Nonequilibrium mechanics of active cytoskeletal networks, *Science* **315**, 370 (2007).
- [40] R. R. Netz, Fluctuation-dissipation relation and stationary distribution of an exactly solvable many-particle model for active biomatter far from equilibrium, *J. Chem. Phys.* **148**, 185101 (2018).
- [41] R. R. Netz, Multi-point distribution for Gaussian non-equilibrium non-Markovian observables, [arXiv:2310.08886](https://arxiv.org/abs/2310.08886).
- [42] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg, Cell motility as persistent random motion: theories from experiments, *Biophys. J.* **89**, 912 (2005).
- [43] H. Flyvbjerg and H. G. Petersen, Error estimates on averages of correlated data, *J. Chem. Phys.* **91**, 461 (1989).
- [44] B. Snijder and L. Pelkmans, Origins of regulated cell-to-cell variability, *Nat. Rev. Mol. Cell Biol.* **12**, 119 (2011).
- [45] S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger, Non-genetic origins of cell-to-cell variability in trail-induced apoptosis, *Nature (London)* **459**, 428 (2009).
- [46] A. J. Berglund, Statistics of camera-based single-particle tracking, *Phys. Rev. E* **82**, 011917 (2010).
- [47] R. R. Netz, Temporal coarse-graining and elimination of slow dynamics with the generalized Langevin equation for time-filtered observables, *Phys. Rev. E* **111**, 054132 (2025).
- [48] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [49] P. Dieterich, R. Klages, R. Preuss, and A. Schwab, Anomalous dynamics of cell migration, *Proc. Natl. Acad. Sci. USA* **105**, 459 (2008).

- [50] R. K. Sadhu, A. Iglič, and N. S. Gov, A minimal cell model for lamellipodia-based cellular dynamics and migration, *J. Cell Sci.* **136**, jcs260744 (2023).
- [51] J. Heyn, M. Atienza Juanatey, M. Falcke, and J. Rädler, Cell-mechanical parameter estimation from 1D cell trajectories using simulation-based inference, *bioRxiv* 2024.09.06.611766 (2024).
- [52] F. Pesarin and L. Salmaso, The permutation testing approach: a review, *Statistica* **70**, 481 (2010).
- [53] M. Wallert, C. Nie, P. Anilkumar, S. Abbina, S. Bhatia, K. Ludwig, J. N. Kizhakkedathu, R. Haag, and S. Block, Mucin-inspired, high molecular weight virus binding inhibitors show biphasic binding behavior to influenza a viruses, *Small* **16**, 2004635 (2020).
- [54] M. Müller, D. Lauster, H. H. Wildenauer, A. Herrmann, and S. Block, Mobility-based quantification of multivalent virus-receptor interactions: new insights into influenza a virus binding mode, *Nano Lett.* **19**, 1875 (2019).
- [55] Y. Kerkhoff and S. Block, Analysis and refinement of 2D single-particle tracking experiments, *Biointerphases* **15**, 021201 (2020).
- [56] Y. von Hansen, A. Mehlich, B. Pelz, M. Rief, and R. R. Netz, Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference, *Rev. Sci. Instrum.* **83**, 095116 (2012).
- [57] C. Schreiber, F. J. Segerer, E. Wagner, A. Roidl, and J. O. Rädler, Ring-shaped microlanes and chemical barriers as a platform for probing single-cell migration, *Sci. Rep.* **6**, 26858 (2016).
- [58] D. Blair and E. Dufresne, The Matlab particle tracking code repository, Particle-tracking code available at <http://physics.georgetown.edu/matlab> (2008).
- [59] A. Klimek, Tools and data for analyzing cell statistics in simulations, GitHub repository, https://github.com/kanton42/cell_statistics (2025).