

Likelihood-based heterogeneity inference reveals nonstationary effects in biohybrid cell-cargo transport

Jan Albrecht ^{1,*}, Lara S. Dautzenberg ¹, Manfred Oppen ^{2,3,4}, Carsten Beta ^{1,5,†} and Robert Großmann ^{1,‡}

¹*Institute of Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany*

²*Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, 10587 Berlin, Germany*

³*Institute of Mathematics, University of Potsdam, 14476 Potsdam, Germany*

⁴*Centre for Systems Modelling and Quantitative Biomedicine, University of Birmingham, Birmingham B15 2TT, United Kingdom*

⁵*Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan*



(Received 20 August 2025; accepted 7 December 2025; published 30 January 2026)

Variability of motility behavior in populations of microbiological agents is a ubiquitous phenomenon even in the case of genetically identical cells. Accordingly, passive objects introduced into such biological systems and driven by them will also exhibit heterogeneous motion patterns. Here, we study a biohybrid system of passive beads driven by active ameboid cells and use a likelihood approach to estimate the heterogeneity of the bead dynamics from their discretely sampled trajectories. We showcase how this approach can deal with information-scarce situations and provides natural uncertainty bounds for heterogeneity estimates. Using these advantages, we particularly uncover that the heterogeneity in the system is time-dependent.

DOI: [10.1103/pvsc-8sbp](https://doi.org/10.1103/pvsc-8sbp)

I. INTRODUCTION

During major concerts or other festivities with densely packed crowds, organizers occasionally distribute large balloons for the attendees to bounce around and interact with [1,2]. This is an example of a macroscopic active system interacting with and collectively driving passive objects; this paper studies a very similar system—albeit shrunk by 5 orders of magnitude.

Systems of self-propelled motile individuals [3] on their own already exhibit a wide range of interesting collective phenomena [4] like large-scale patterns [5] and nonequilibrium phase transitions [6–9]. Since real-life active systems are rarely clean and isolated, the interaction of such active particles with obstacles and movable passive objects is an active area of research [10]. In this regard, much work has focused on the diffusion of passive tracer particles in suspensions of microswimmers [11–13]. Another important research direction is studying composite systems, in which there is direct physical contact between the active and passive particles rather than hydrodynamic interaction [14–18]. While there are interesting effects in macroscopic realizations of such composite systems, too, the research focus has been on microscopic systems, where they have potential applications

in micromanipulation and -transport [19–21] as well as in targeted drug delivery [22,23]. To exploit these composite systems, a thorough understanding and modeling of the motion patterns of their constituents is required.

A ubiquitous aspect of biological active matter is its inherent interindividual variability that is present even for genetically identical microorganisms [24–26] and can manifest itself in their motility [27–29]. When passive objects are driven by a heterogeneous population of cells, the objects' dynamics inherit the heterogeneity, which may be further increased by fluctuations in the number of cells attached to each object [17]. Such population heterogeneity can pose challenges for model inference, as it complicates the interpretation of population-averaged quantities and can lead to unusual statistics, such as non-Gaussian displacement distributions [17,30,31]. Therefore, inferring the variability of a system constitutes an important part of its characterization.

Likelihood-based approaches can be used to obtain heterogeneity estimates directly from the measured trajectory data, rather than via an intermediate estimation of motility parameters [32–34]. Estimators derived from the likelihood, such as maximum-likelihood estimators (MLEs), have favorable statistical properties and provide natural uncertainty estimates for the inferred parameters [33,35]. Furthermore, these methods perform particularly well in situations with little available data, which is why they are also a popular tool in pharmacological research where they are used to infer interpatient variability [36].

In this article, we consider an experimental biohybrid cell-cargo system consisting of a carpet of *Dictyostelium discoideum* (*D. discoideum*) cells interacting with colloidal “cargo” beads, similar to concertgoers with large balloons. We have previously studied this system with a different bead size [17] and found Fickian, yet non-Gaussian diffusion due to

*Contact author: jan.albrecht@uni-potsdam.de

†Contact author: beta@uni-potsdam.de

‡Contact author: rgrossmann@uni-potsdam.de

heterogeneity in the system. Here, we use a likelihood-based approach to infer the heterogeneity in the dynamics of the cargo beads. Since this method can handle information-scarce situations, we use it to analyze short time snippets and through this uncover a drift of the heterogeneity over the course of the experiment.

The remainder of the paper is structured as follows. In Sec. II, we describe the system and deduce a model that captures the dynamics and heterogeneity of the beads. In Sec. III, we detail the likelihood method and use it for inference on the whole dataset before we investigate the time dependence in Sec. IV. We end with some more general remarks in Sec. V.

II. DATA AND MODEL

We study a composite system of motile ameboid cells and passive polystyrene beads: On top of a dense monolayer of *D. discoideum* cells, a small number of beads are added [17]. Interactions between the cells and the beads are mediated through unspecific adhesion [37] once a cell and a bead come into contact [38]. The beads experience fluctuating forces through locomotion of the cells as well as their membrane movements [39]. Each of the cell-cargo bonds can break either spontaneously [22] or through forces applied by other cells [39]. In this way, the layer of cells acts as an active bath for the beads that exhibit nonthermal stochastic trajectories.

The same experimental setup has been featured in Ref. [17]. There, a dataset with bead diameter of 46 μm was analyzed. Here, we focus on an experiment with beads 61 μm in diameter. We perform time-lapse recordings by imaging the system every 15 s during a total recording time of 4 h. The position of the beads was tracked, leading to a total of $N = 177$ trajectories. While we only present the 61 μm dataset in the main text, the results of our analysis on the dataset from Ref. [17] as well as two additional datasets with bead diameters of 29 and 100 μm are shown and discussed in the Supplemental Material [40]. The datasets for all bead sizes are published in Ref. [41].

First, we perform a statistical analysis analogous to Ref. [17], which will lead to the same conclusions about the form of the model to describe the dynamics of the population of beads. Once the model type is chosen, the following sections focus on the likelihood-based analysis of the data, which is specific to the paper at hand.

We particularly consider the mean-squared displacement (MSD) of the trajectories. For each trajectory, the squared displacement for a time delay t with respect to some reference time s is given by

$$\omega(t; s) = |\mathbf{r}(t + s) - \mathbf{r}(s)|^2. \quad (1)$$

The time-averaged mean-squared displacement (tamsd) for a trajectory of duration T starting at $t_0 = 0$ is then given by

$$\overline{\omega(t)} = \frac{1}{T - t} \int_0^{T-t} ds \omega(t; s). \quad (2)$$

We also consider the ensemble averages of the squared displacements as well as those of the tamsd over all trajectories, which leads to the ensemble-averaged MSD (eamsd) $\langle \omega(t) \rangle$ and the ensemble-averaged time-averaged MSD (eatamsd) $\langle \overline{\omega(t)} \rangle$, respectively. In case of the eamsd, the reference time

is the starting time of each trajectory. Figure 1(a) shows the different types of MSD calculated from the dataset. Starting from $t \approx 100$ s, the eatamsd has an exponent close to 1 (Fickian diffusion) over 2 orders of magnitude, indicating diffusive behavior at timescales longer than 2 min. The eamsd is more noisy but appears to be linear in time for even shorter timescales.

In addition to the MSD, we consider the displacement autocorrelation function (DACF). For a trajectory of duration T starting at $t_0 = 0$ and an interval length of τ , it is given by

$$C_\tau(\delta) = \frac{1}{T - (\tau + \delta)} \int_0^{T-(\tau+\delta)} dt [\Delta_\tau \mathbf{r}(t)] \cdot [\Delta_\tau \mathbf{r}(t + \delta)], \quad (3)$$

where $\Delta_\tau \mathbf{r}(t) := \mathbf{r}(t + \tau) - \mathbf{r}(t)$. We focus on the normalized correlation with rescaled time shift $\tilde{\delta} := \delta/\tau$:

$$\tilde{C}_\tau(\tilde{\delta}) := C_\tau(\tilde{\delta} \cdot \tau) / C_\tau(0). \quad (4)$$

The correlation $\tilde{C}_\tau(\tilde{\delta})$ calculated from the data is shown in Fig. 1(b) for a number of values for τ . For interval lengths τ larger than some critical value τ_c , the curves collapse onto one master curve (see also Ref. [17]). This master curve is given by the theoretical result of the correlation for a purely diffusive particle. For such a particle, the correlation is proportional to the overlap of the two intervals in Eq. (3) and therefore drops to zero for $\delta/\tau > 1$. The plot indicates that the value of τ_c for the experimental bead-cargo system lies between 1 and 2 min. This means that nonoverlapping displacements of the beads with interval lengths τ above 2 min are independent of each other. This is another indication that the dynamics of the beads is well described by Brownian motion above a timescale of 2 min. Many other models that could have been good candidates *a priori*, like Lévy walks or flights or processes driven by correlated noise, either contradict one of the above observations or become equivalent to pure diffusion for sufficiently long measurement times [40].

Finally, we look at the empirical displacement distributions of the dataset at different time lags τ . The results for steps in the x direction are shown in Fig. 1(c); the distributions for y steps (not shown) are qualitatively identical. In contrast to the results discussed in the paragraphs above, the displacement distributions do not show a signature of Brownian motion: While the displacement distributions of a purely diffusive particle are Gaussian, the experimental step-size distributions show clear exponential tails.

Going back to the plot of the MSDs in Fig. 1(a), we observe that the tamsd curves are not tightly packed as would be expected for a homogeneous ensemble but are rather spread out over almost 2 orders of magnitude. This spread hints at heterogeneity in the system. Variability in the dynamics of the individual beads can arise through random variations in the local cell density as well as through variability of the dynamics of the cells that drive the beads. Let us assume that the dynamics of each of the beads above a timescale of 2:00 min is well captured by standard diffusion

$$\dot{\mathbf{r}}^n(t) = \sigma_n \boldsymbol{\xi}^n(t), \quad (5)$$

as suggested by the MSDs and the DACF. Here, the index $n \in [1, \dots, N]$ indicates the individual beads and $\boldsymbol{\xi}^n(t)$ are

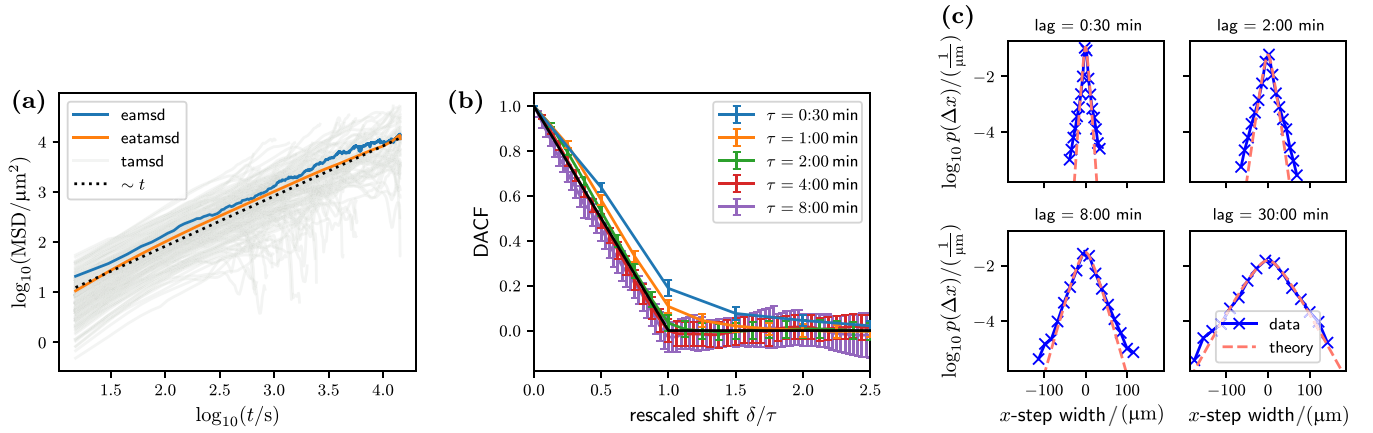


FIG. 1. Statistical characterization of bead trajectories driven by an active cell bath. (a) Mean-squared displacement. The blue and the orange lines show eamsd and eatamsd, respectively. The transparent gray lines show the tamsd of the individual trajectories [see Eq. (2)]. The dotted black line shows the theoretical tamsd of a long trajectory of a purely diffusive particle, which is proportional to t . (b) Rescaled and normalized displacement autocorrelation functions. The black line indicates the theoretical result for a purely diffusive particle. (c) Step-size distributions in the x direction. The blue curves are histograms calculated from the dataset. Each cross indicates the center of a bin interval. The red dashed lines are the predictions from the heterogeneity inference (see Sec. III).

independent white noise processes. We can then estimate the noise strength σ_n^2 for each of the trajectories using the maximum-likelihood estimator given in Eq. (A5). The empirical distribution of these noise strengths shown in Fig. 2(b) has a single peak and is well approximated by the probability density function (pdf) of a Γ distribution. We therefore model the system as diffusive particles driven by the stochastic differential equation (5) with noise strengths $\{\sigma_n^2\}_{n=1,\dots,N}$ identically and independently distributed according to a Γ distribution

$$\sigma_n^2 \sim \Gamma(\alpha, \beta), \quad (6)$$

with pdf

$$q_\Gamma(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (7)$$

and heterogeneity parameters $\theta := (\alpha, \beta)$. While other distributions might also fit the heterogeneity of the noise strengths, the use of the Γ function is motivated by its analytical properties, which ease the calculations in the following sections. Other choices do not qualitatively change the conclusions presented in the following [40].

In the next section, we will use the model comprised of Eqs. (5) and (6) to estimate the heterogeneity of the experimental system.

III. LIKELIHOOD-BASED INFERENCE

In the previous section, we have derived a hierarchical model: On the lower level, the noise strength σ_n^2 for each of the individual trajectories is unknown and needs to be inferred from the position data; on the higher level, the system heterogeneity, i.e., the statistical distribution of these unknown noise strengths over the dataset, is also unknown.

There are at least two distinct approaches to estimate the parameters of the heterogeneity distribution [32]. One is to treat the two levels of stochasticity individually. This means that, first, estimators $\hat{\sigma}_n^2$ are calculated independently for each trajectory. In a second step, these estimates can then be used

within some inference scheme to estimate the heterogeneity. We call this approach, which has been used in Ref. [17], a two-step estimate. Such a procedure is also known as a two-stage [42] or plug-in [43] approach. When doing so, all information included in a trajectory is reduced to a single number $\hat{\sigma}_n^2$ and any uncertainty about this estimate is ignored. One way to quantify the range of plausible values for the noise strength given a trajectory T^n is to look at the log-likelihoods of the noise strength:

$$\log p(T^n|\sigma_n^2) =: \ell_n(\sigma_n^2). \quad (8)$$

For Brownian motion as described by Eq. (5), it can be expressed as a product of Gaussians [44]; the full expression of $p(T^n|\sigma_n^2)$ is given in Eq. (A4). In Fig. 2(a), the log-likelihoods $\ell_n(\sigma_n^2)$ with respect to the trajectories in the experimental dataset are shown. Some of the log-likelihood functions are narrow with respect to the range of noise strengths in the dataset. For these trajectories, the reduction of the information in the trajectory to the one estimate value $\hat{\sigma}_n^2$ is justified. However, for other trajectories, the likelihood is very broad, indicating that there is a wide range of possible and relevant values for the noise strength.

In order to use the available information efficiently, an inference scheme for the heterogeneity should take this range of possible values into account. We therefore propose here to use a full-likelihood approach that allows inference of the heterogeneity parameters directly from the measured positions [32]. While Eq. (8) is the log-likelihood of the noise strength with respect to a trajectory, we now consider the log-likelihood of the heterogeneity parameters $\theta = (\alpha, \beta)$:

$$\log p(T^n|\theta) = \log \int d\sigma^2 p(T^n|\sigma^2) q_\Gamma(\sigma^2|\theta) =: \mathcal{L}_n(\theta). \quad (9)$$

The integration inside the logarithm combines all possible noise intensities σ_n^2 , each weighted by the likelihood. This way, an information bottleneck as for the two-step approach is avoided.

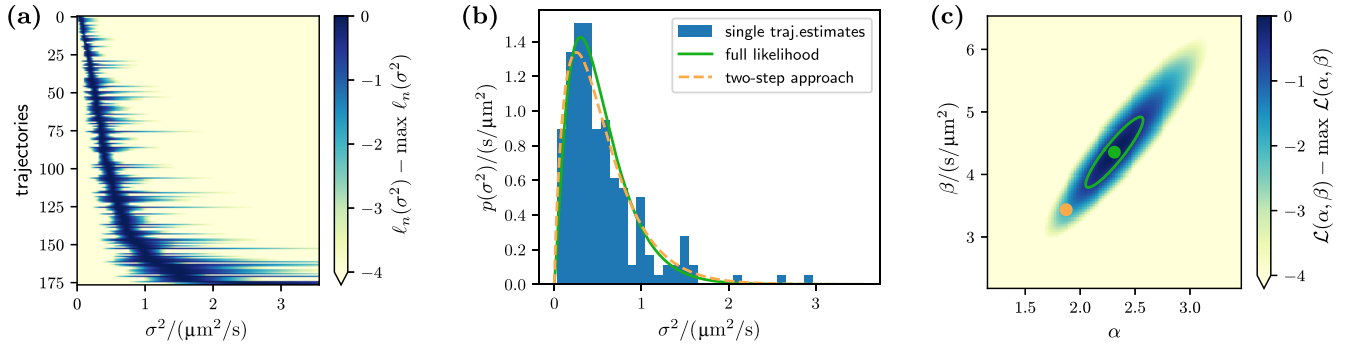


FIG. 2. Heterogeneity inference on bead trajectories driven by an active cell bath. A sparse version of each trajectory with $\Delta t = 2$ min is considered to ensure that the dynamics are in the diffusive regime. (a) Log-likelihood of the noise strength given the trajectories of the beads [see Eq. (8)]. The trajectories are sorted by their maximum-likelihood estimate for the noise strength. (b) Inferred probability density function of the noise strengths. The green curve represents the result of the full-likelihood approach, while the dashed orange curve is the result from a two-step approach. The histogram shows the MLEs of the noise strengths based on the functions shown in panel (a). (c) Log-likelihood $\mathcal{L}(\theta)$ of the heterogeneity parameters with respect to the sparse dataset. The green dot denotes the MLE $\hat{\theta}$ that maximizes $\mathcal{L}(\theta)$. The green ellipse is the uncertainty estimate calculated from the Hessian matrix at θ . The orange dot denotes the estimate obtained from a two-step inference approach.

The likelihoods for the individual trajectories can be combined into a log-likelihood of θ with respect to the full dataset $\mathcal{D} = \{\mathbf{T}^n\}_{n=1,\dots,N}$:

$$\mathcal{L}(\theta) := \log p(\mathcal{D}|\theta) = \sum_n \mathcal{L}_n(\theta). \quad (10)$$

The parameter value that maximizes a likelihood function $\hat{\theta}$ is known as the maximum-likelihood estimator. This estimator has favorable statistical properties like consistency and efficiency [35]. Furthermore, just like the shape of the noise strength log-likelihood was indicative of the range of relevant noise strengths, the Hessian matrix of $\mathcal{L}(\theta)$ at the MLE is connected to the uncertainty of $\hat{\theta}$: For a large number of trajectories, the negative inverse of the Hessian matrix approaches the variance matrix of the MLE [32,35]. For the model at hand, the integral in Eq. (9) and the elements of the Hessian matrix have analytical expressions in terms of modified Bessel functions and their derivatives (see Appendix A and Ref. [40]). There is no analytic expression for $\hat{\theta}$, but it can be easily obtained by numerical maximization of $\mathcal{L}(\theta)$.

We apply this full-likelihood approach to the dataset of bead trajectories in the experimental cell-cargo system. As described in Sec. II, the system has been imaged every 15 s, but the observed dynamics can only be assumed to be diffusive above a timescale of 2 min. We therefore work with sparse versions of the trajectories in which positions are removed such that the time difference between subsequent positions in the modified trajectory is 2 min. When creating these sparse trajectories, we start from the first position in every trajectory. In the following, all inference results will be based on this sparse dataset.

Figure 2(c) shows the log-likelihood of the heterogeneity parameters with respect to the sparse experimental data. Note that α is a dimensionless number, while β has inverse units of the noise strength. In addition, the MLE $\hat{\theta}$ is displayed as well as an ellipse that denotes the uncertainty estimate. The corresponding density of the heterogeneity distribution is shown in Fig. 2(b). The inference confirms that the effective noise strengths that drive the beads are strongly heterogeneous over

the dataset. Similarly to Ref. [17], we confirm that the inferred heterogeneity explains the behavior of the step-size distributions in Fig. 1, except for the outmost tails. The orange dotted line and the orange dot in Figs. 2(b) and 2(c), respectively, show an alternative heterogeneity estimate from the two-step approach. After calculating estimates of the noise strength for each trajectory by maximizing Eq. (8), these estimates $\{\hat{\sigma}_n^2\}$ are used in a maximum-likelihood scheme to optimize a Γ distribution. The estimator from the two-step approach is outside the uncertainty bounds of the full-likelihood approach.

The need for reliable uncertainty estimates and the likelihood framework increases when less data are available, since less data imply a larger uncertainty about the individual noise strength estimates. In the next section, we will apply the approach to a situation in which the available data are limited, namely, short trajectories: We analyze only snippets of trajectories to assess the stationarity of the bead dynamics.

IV. NONSTATIONARITY OF THE SYSTEM

The model in Eqs. (5) and (6) for the dynamics of the beads and their heterogeneity that we have used for inference in Sec. III is stationary—The noise strengths and therefore also the heterogeneity parameters are assumed not to change over time. In the following, we will relax this assumption.

While the results presented in Fig. 1 strongly suggest diffusive dynamics of the beads, it is still possible that the value of the noise strength changes on timescales longer than the measurement frequency [45]. Over the duration of the experiments, the beads are transported from one part of the cell carpet to the next, where other distinct cells are attached to the bead and also the cell density might be different. Furthermore, the behavior of the cells might change over time. Here, we are not primarily interested in the exact dynamics of the individual beads, but rather want to know if the overall heterogeneity distribution of the dynamics in the system changes over the course of the experiment, i.e., whether the system shows nonstationary behavior.

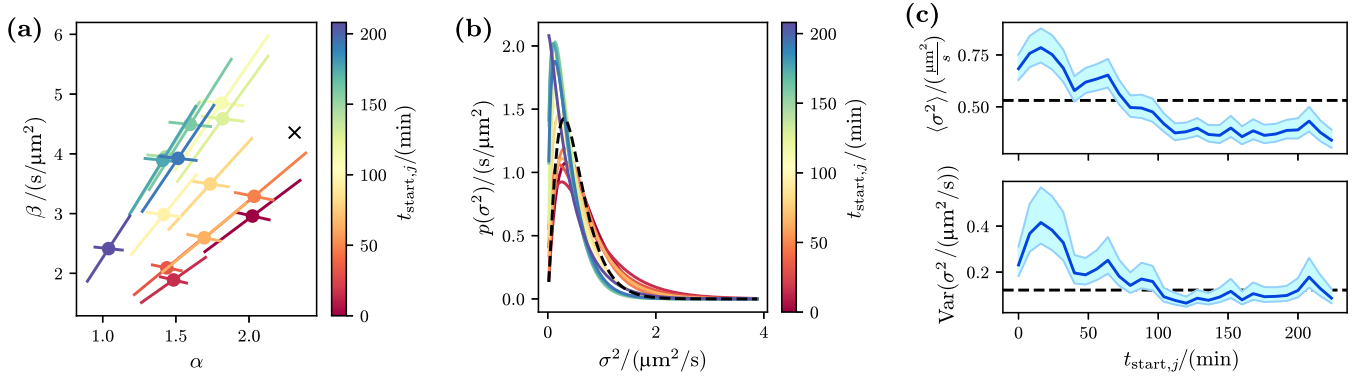


FIG. 3. Time dependence of heterogeneity inference. Inference was performed on reduced datasets containing only data points within a time window of 16 min. The start time $t_{\text{start},j}$ of this time window was shifted over the observation time of the experiment. (a) Inferred heterogeneity parameters θ_j . The black cross denotes the result of the inference from the complete dataset. Note that the error bars appear nonperpendicular due to the skewed aspect ratio. (b) Inferred heterogeneity distribution. The black dashed line denotes the result of the inference from the complete dataset, reflecting an average behavior. (c) Mean and variance of the inferred distribution over time. The dashed line denotes the result of the inference from the complete dataset. The plotted error bounds correspond to the highest and lowest values of mean and variance within the 1σ uncertainty bounds in θ space. In panel (c), the plotted $t_{\text{start},j}$ are spaced 8 min apart, which means that the windows partially overlap. Note that in panels (a) and (b), only every second point is plotted for better readability, leading to a spacing of 16 min.

In order to do so, we consider short time windows with onset time $t_{\text{start},j}$ and duration $w = 16$ min within which we assume the system to be stationary. The motion of individual *D. discoideum* cells is diffusive above a timescale of around 10 min with an average diffusion coefficient lower than $10 \mu\text{m}^2/\text{min}$ [17,46]. This means their expected displacement in 16 min is $(4 \times 10 \mu\text{m}^2/\text{min} \times 16 \text{ min})^{1/2} \approx 25 \mu\text{m}$ while being only 10–20 μm in diameter [47]. Therefore, only limited rearrangement of the active bath is possible during these time windows. The heterogeneity can then be estimated using the data in this time window only. Note that there will be a maximum of only nine positions per trajectory within each window width since we are using sparse trajectories with 2 min between data points. For each window j , the full-likelihood approach returns an estimate for the heterogeneity parameters θ_j together with an uncertainty estimate. By shifting the time window, we obtain a series of estimates $\{\theta_j\}$ and their uncertainties. Note that the full-likelihood approach is particularly well suited to deal with these short trajectories, since heterogeneity distributions can be inferred taking into account fluctuations at the single trajectory level under these information-scarce conditions, whereas a two-step approach (see Appendix B) is prone to inaccuracy as detailed in Sec. III.

The results in Fig. 3 show that the system is indeed not stationary but transient. The mean and the variance of the heterogeneity distribution decrease over the first two hours of the experiment before becoming much more stable during the second two hours. The dynamics in θ space with their uncertainties depicted in Fig. 3(a) show that the change in the first half of the experiment is indeed a significant, systematic shift rather than random fluctuations. During the second half, there is a large overlap of the uncertainties of the estimates. This indicates that the system has reached a quasistationary state. In Fig. 3(b), we can see that the heterogeneity distribution estimated from the full dataset constitutes a midpoint between the distributions inferred at the beginning and the end of the

experiment. The results above are robust against the changes of the window duration [40].

The drop in the average of the distribution means that the diffusion coefficient of the beads decreases over time. One important factor is probably the accumulation of cells around the beads, which is also visible in the raw footage from the microscope [40]. The accumulation arises due to the unspecific adhesion of *D. discoideum* cells [37] to the beads and, moreover, cells prefer structured environments with more contact area over flat surfaces [48]. Such an accumulation has several effects. The more cells are attached to a bead, the more the forces, which cells exert on the bead, cancel each other. Furthermore, it is less likely for a single cell to pull the bead off all other cells and move it by a larger distance. Both effects decrease the effective diffusion constant of the bead. Another aspect is that the maximum number of cells around a bead is limited by the steric repulsion of cells. Thus, the accumulation will naturally saturate at some density, which is similar for all beads. This decreases the contribution of cell density fluctuations to the variability of the dynamics of the beads, explaining the reduced variance.

Another possible contributing factor to the temporal change of the heterogeneity is a time dependence of the behavior of the cells themselves. *D. discoideum* cells are known to show quorum-sensing behavior that affects their motility [46]. The increase in concentration of the quorum-sensing factor could lead to a slowdown in their dynamics until a saturation of the sensing pathway sets in. A generally slower movement of each cell would lead to decreased mean as well as a decreased variance of the bead dynamics. Tracking and analyzing the motility of the cells would give a clearer picture about the influence of this factor; however, this is beyond the scope of the present study.

The nonstationarity of the data does not come as a surprise when we take a closer look at the MSD curves in Fig. 1(a). For a stationary system, the additional time averaging in the eatmsd leads to a less noisy curve compared to the eamsd

but does not shift the curve. The discrepancy of eamsd and eatamsd therefore indicates changes of the dynamics over time. The thorough likelihood-based analysis of the time windows reveals the nature of the nonstationarity in the system as well as the quasistationary behavior in the second half of the experiment.

V. CONCLUSION

We have presented a likelihood-based approach to study the time-dependent heterogeneous dynamics of microbeads in a bath of active ameboid cells. The likelihood approach allows to skip the intermediate estimation of trajectory-specific quantities—the individual effective diffusion coefficients in this case. This way, the available trajectory information is used efficiently and information bottlenecks are avoided. In addition, the approach provides natural uncertainty estimates that allow to judge the reliability of the heterogeneity estimates.

Using the likelihood-based approach, we show that the heterogeneity within the motion pattern of the population of beads changes significantly over the course of the experiment. During the first two hours, the dynamics “cool off,” becoming slower and less variable before reaching a quasistationary state for the remainder of the experiment. When conducting the experiment, the cells were given enough time to attach to the glass surface and start their basic motion pattern. Similarly, it was ensured that the beads had time to sediment before the start of the recording (see Ref. [40]). Despite the two individual subsystems being equilibrated prior to the recording, the composite system of beads and cells needs much longer to reach a steady state.

Not accounting for the dynamics of the heterogeneity can skew results when trying to quantify the variability in the system. But even when the heterogeneity quantification is not the goal, analyzing its time dependency can still be beneficial. It may be used to identify windows of stationary dynamics, which might be challenging to do by analyzing the imaging data or considering the extracted trajectories individually. The likelihood approach is especially well suited since it can deal with the short trajectories within the time windows.

The system studied here was described by a model that has an analytically tractable likelihood for the heterogeneity parameters. However, the same likelihood machinery can also be used on more complex systems with more involved models for which the exact likelihoods are untractable. In such cases, approximation techniques are required [32,33], but once appropriate likelihood expressions have been found, the method itself is also able to deal with additional terms such as particle-particle [49] or particle-wall [50] interactions. The more complex the model, the more difficult it is to estimate parameters and variability from limited data. Especially here, the likelihood approach can help to uncover heterogeneities as well as time dependence of the parameters.

ACKNOWLEDGMENTS

We thank Kirsten Sachse for supporting laboratory routines. This research has been partially funded by the Deutsche Forschungsgemeinschaft (DFG)—Project No. 318763901—SFB1294.

DATA AVAILABILITY

The data that support the findings of this article are openly available [41].

APPENDIX A: THEORETICAL MODELING

We consider an ensemble of N Brownian particles. The dynamics of each of them are described by

$$\dot{\mathbf{r}}^n(t) = \sigma_n \boldsymbol{\xi}^n(t), \quad (\text{A1})$$

where $\boldsymbol{\xi}^n(t)$ is d -dimensional white noise and σ_n^2 is the particle-specific noise intensity. These noise strengths are identically and independently distributed according to a Γ distribution $\Gamma(\alpha, \beta)$ with density

$$q_\Gamma(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\sigma^2 \beta} (\sigma^2)^{\alpha-1}. \quad (\text{A2})$$

The position of the particles is measured discretely at certain points in time, leading to observed trajectories of the form

$$\mathbf{T}^n = \{(\mathbf{t}_i^n, \mathbf{r}_i^n)\}_{i=1, \dots, M^n+1}. \quad (\text{A3})$$

Note that the upper index n indicates the particle, while the lower index i labels the specific measurement. This leads to observed displacements $\Delta \mathbf{r}_i^n := \mathbf{r}_{i+1}^n - \mathbf{r}_i^n$ and corresponding time lags $\Delta t_i^n := t_{i+1}^n - t_i^n$. A trajectory contains $M^n + 1$ positions and therefore M^n displacements.

Given the noise strength σ_n^2 , the probability of a measured trajectory follows from the Brownian dynamics and is

$$p(\mathbf{T}^n | \sigma_n^2) = \prod_{i=1}^{M^n} (2\pi \sigma_n^2 \Delta t_i^n)^{-\frac{d}{2}} \exp\left(-\frac{(\Delta \mathbf{r}_i^n)^2}{2 \sigma_n^2 \Delta t_i^n}\right), \quad (\text{A4})$$

where d is the spatial dimension of the process. For the experimental system considered in this paper, we have $d = 2$. Given a trajectory $\mathbf{T}^n$, Eq. (A4) is maximized by

$$\hat{\sigma}_n^2 := \frac{1}{d M^n} \sum_i \frac{(\Delta \mathbf{r}_i^n)^2}{\Delta t_i^n}, \quad (\text{A5})$$

which is the maximum-likelihood estimator for the noise strength.

The probability of a trajectory conditioned on the parameters of the Γ distribution is calculated by

$$p(\mathbf{T}^n | \alpha, \beta) = \int d\sigma^2 p(\mathbf{T}^n | \sigma^2) q_\Gamma(\sigma^2 | \alpha, \beta). \quad (\text{A6})$$

This is just the likelihood of the heterogeneity parameters (α, β) given trajectory $\mathbf{T}^n$. The above integral has an explicit analytical solution, which is given by

$$\begin{aligned} \log p(\mathbf{T}^n | \alpha, \beta) &= C + \frac{1}{2}(\alpha - b_n) \log\left(\frac{a_n}{\beta}\right) + \alpha \log \beta \\ &\quad - \log \Gamma(\alpha) + \log K_{(b_n - \alpha)}(2\sqrt{a_n \beta}) \\ &=: \mathcal{L}_n(\alpha, \beta), \end{aligned} \quad (\text{A7})$$

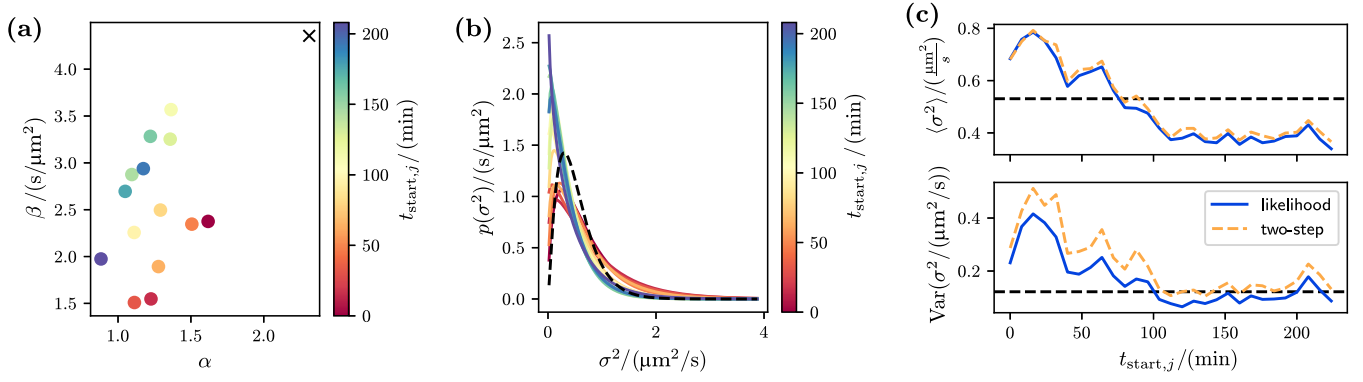


FIG. 4. Time dependence of heterogeneity inference using a two-step approach with $\hat{\sigma}_n^2$ obtained from single trajectory MLEs. Inference was performed on reduced datasets containing only data points within a time window of 16 min. The start time $t_{\text{start},j}$ of this time window was shifted over the observation time of the experiment. (a) Inferred heterogeneity parameters $\hat{\theta}_j$. (b) Inferred heterogeneity distribution. As a reference, the black cross and the black dashed line denote the result of the likelihood inference from the complete dataset. (c) Mean and variance of the inferred distribution over time. The dashed line denotes the result of the likelihood inference from the complete dataset.

with shorthands

$$C := \log 2 - b_n \log 2\pi - \frac{d}{2} \sum_{i=1}^{M^n} \log \Delta t_i^n, \quad (\text{A8a})$$

$$a_n := \frac{1}{2} \sum_{i=1}^{M^n} \frac{(\Delta \mathbf{r}_i^n)^2}{\Delta t_i^n}, \quad (\text{A8b})$$

$$b_n := \frac{M^n d}{2}, \quad (\text{A8c})$$

and where $K_\nu(x)$ is the modified Bessel function of the second kind. The likelihood of the heterogeneity parameters with respect to the whole dataset $\mathcal{D} = \{\mathbf{T}^n\}_{n=1,\dots,N}$ is then

$$\mathcal{L}(\alpha, \beta) = \log p(\mathcal{D}|\alpha, \beta) = \sum_{n=1}^N \mathcal{L}_n(\alpha, \beta). \quad (\text{A9})$$

The maximum of Eq. (A9) is the maximum-likelihood estimate $(\hat{\alpha}, \hat{\beta})$ for the heterogeneity parameters α and β .

In order to estimate the uncertainty of this estimate, we calculate the Hessian matrix of $\mathcal{L}(\alpha, \beta)$ at $(\hat{\alpha}, \hat{\beta})$. This can be understood as an empirical approximation of the negative Fisher information matrix $-I_{ij}(\hat{\theta}) = \mathbb{E}[\partial_{\theta_i} \partial_{\theta_j} \mathcal{L}(\theta)]|_{\theta=\hat{\theta}}$, where the expectation value is with respect to the true distribution of the data. In the limit of many trajectories per dataset, the maximum-likelihood estimates for different realizations of a dataset are expected to be normally distributed around the true value with the variance matrix being the inverse Fisher

information matrix $I^{-1}(\hat{\theta})$ [32,35]. In two dimensions, around 39.3% of the probability mass lies within the 1σ ellipse and 86.5% within the 2σ ellipse (see Ref. [40]).

The Hessian matrix of $\mathcal{L}(\alpha, \beta)$ is

$$\begin{aligned} H_{\mathcal{L}}(\hat{\alpha}, \hat{\beta}) &= \sum_n H_{\mathcal{L}_n}(\hat{\alpha}, \hat{\beta}) \\ &= \sum_n \begin{pmatrix} \partial_\alpha \partial_\alpha \mathcal{L}_n & \partial_\alpha \partial_\beta \mathcal{L}_n \\ \partial_\beta \partial_\alpha \mathcal{L}_n & \partial_\beta \partial_\beta \mathcal{L}_n \end{pmatrix} \Big|_{(\alpha,\beta)=(\hat{\alpha},\hat{\beta})}. \end{aligned} \quad (\text{A10})$$

This matrix can be calculated directly by differentiating Eq. (A7). In Ref. [40], we sketch a way that eases implementation.

APPENDIX B: TIME DEPENDENCE INFERRED USING A TWO-STEP APPROACH

In Fig. 4, the results of heterogeneity inference on 16:00 min time windows using the two-step approach are shown. The estimates for $\hat{\sigma}_n^2$ were obtained using the MLE given in Eq. (A5). The heterogeneity estimates were then calculated as MLEs of a Γ distribution with respect to the set of noise strength estimates $\{\hat{\sigma}_n^2\}_{n=1,\dots,N}$. In Fig. 4(c), we can see that the variance clearly deviates, while the estimated mean of the distributions is close to the results of the likelihood inference. Accordingly, we can also observe a shift of the estimates in heterogeneity parameters space displayed in panel (a). Importantly, the heterogeneity estimates have no uncertainty bars associated with them in contrast to the full-likelihood-based method.

- [1] F. Gu, B. Guiselin, N. Bain, I. Zuriguel, and D. Bartolo, Emergence of collective oscillations in massive human crowds, *Nature (London)* **638**, 112 (2025).
- [2] BBC Music, Orbital—Belfast (Glastonbury 2024), YouTube video (2024), <https://www.youtube.com/watch?v=9eDUY-UeDist=157s>, accessed 4 July, 2025.

- [3] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, Active Brownian particles: From individual to collective stochastic dynamics, *Eur. Phys. J. Spec. Top.* **202**, 1 (2012).
- [4] M. te Vrugt and R. Wittkowski, Metareview: A survey of active matter reviews, *Eur. Phys. J. E* **48**, 12 (2025).

- [5] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Hydrodynamics of soft active matter, *Rev. Mod. Phys.* **85**, 1143 (2013).
- [6] M. E. Cates and J. Tailleur, Motility-induced phase separation, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
- [7] H. Chaté, Dry aligning dilute active matter, *Annu. Rev. Condens. Matter Phys.* **11**, 189 (2020).
- [8] M. Bär, R. Großmann, S. Heidenreich, and F. Peruani, Self-propelled rods: Insights and perspectives for active matter, *Annu. Rev. Condens. Matter Phys.* **11**, 441 (2020).
- [9] R. Alert, J. Casademunt, and J.-F. Joanny, Active turbulence, *Annu. Rev. Condens. Matter Phys.* **13**, 143 (2022).
- [10] 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).
- [11] J. Shea, G. Jung, and F. Schmid, How boundary interactions dominate emergent driving of inertial passive probes in active matter, *J. Phys. A: Math. Theor.* **57**, 235006 (2024).
- [12] K. C. Leptos, J. S. Guasto, J. P. Gollub, A. I. Pesci, and R. E. Goldstein, Dynamics of enhanced tracer diffusion in suspensions of swimming eukaryotic microorganisms, *Phys. Rev. Lett.* **103**, 198103 (2009).
- [13] F. Guzmán-Lastra, H. Löwen, and A. J. T. M. Mathijssen, Active carpets drive non-equilibrium diffusion and enhanced molecular fluxes, *Nat. Commun.* **12**, 1906 (2021).
- [14] R. W. Carlsen and M. Sitti, Bio-hybrid cell-based actuators for microsystems, *Small* **10**, 3831 (2014).
- [15] V. A. Webster-Wood, M. Guix, N. W. Xu, B. Behkam, H. Sato, D. Sarkar, S. Sanchez, M. Shimizu, and K. K. Parker, Bio-hybrid robots: Recent progress, challenges, and perspectives, *Bioinspir. Biomim.* **18**, 015001 (2023).
- [16] D. B. Weibel, P. Garstecki, D. Ryan, W. R. DiLuzio, M. Mayer, J. E. Seto, and G. M. Whitesides, Microoxen: Microorganisms to move microscale loads, *Proc. Natl. Acad. Sci. USA* **102**, 11963 (2005).
- [17] R. Großmann, L. S. Bort, T. Moldenhawer, M. Stange, S. S. Panah, R. Metzler, and C. Beta, Non-Gaussian displacements in active transport on a carpet of motile cells, *Phys. Rev. Lett.* **132**, 088301 (2024).
- [18] A. Altschuler, O. L. Bonomo, N. Gorohovsky, S. Marchini, E. Rosen, O. Tal-Friedman, S. Reuveni, and Y. Roichman, Environmental memory facilitates search with home returns, *Phys. Rev. Res.* **6**, 023255 (2024).
- [19] A. Sokolov, M. M. Apodaca, B. A. Grzybowski, and I. S. Aranson, Swimming bacteria power microscopic gears, *Proc. Natl. Acad. Sci. USA* **107**, 969 (2010).
- [20] R. D. Leonardo, L. Angelani, D. Dell'Arciprete, G. Ruocco, V. Iebba, S. Schippa, M. P. Conte, F. Mecarini, F. D. Angelis, and E. D. Fabrizio, Bacterial ratchet motors, *Proc. Natl. Acad. Sci. USA* **107**, 9541 (2010).
- [21] S. Martel and M. Mohammadi, Using a swarm of self-propelled natural microrobots in the form of flagellated bacteria to perform complex micro-assembly tasks, in *2010 IEEE International Conference on Robotics and Automation* (IEEE, Anchorage, AK, USA, 2010), pp. 500–505.
- [22] O. Nagel, M. Frey, M. Gerhardt, and C. Beta, Harnessing motile amoeboid cells as trucks for microtransport and -assembly, *Adv. Sci.* **6**, 1801242 (2019).
- [23] S. J. Park, S.-H. Park, S. Cho, D.-M. Kim, Y. Lee, S. Y. Ko, Y. Hong, H. E. Choy, J.-J. Min, J.-O. Park, *et al.*, New paradigm for tumor theranostic methodology using bacteria-based micro-robot, *Sci. Rep.* **3**, 3394 (2013).
- [24] C. E. Meacham and S. J. Morrison, Tumour heterogeneity and cancer cell plasticity, *Nature (London)* **501**, 328 (2013).
- [25] S. Peled, S. D. Ryan, S. Heidenreich, M. Bär, G. Ariel, and A. Be'er, Heterogeneous bacterial swarms with mixed lengths, *Phys. Rev. E* **103**, 032413 (2021).
- [26] G. Ariel, A. Ayali, A. Be'er, and D. Knebel, Variability and heterogeneity in natural swarms: Experiments and modeling, in *Active Particles: Advances in Theory, Models, and Applications*, edited by N. Bellomo, J. A. Carrillo, and E. Tadmor (Springer International Publishing, Berlin, 2022), Vol. 3, pp. 1–33.
- [27] J. L. Spudis and D. E. Koshland, Non-genetic individuality: Chance in the single cell, *Nature (London)* **262**, 467 (1976).
- [28] A. J. Waite, N. W. Frankel, Y. S. Dufour, J. F. Johnston, J. Long, and T. Emonet, Non-genetic diversity modulates population performance, *Mol. Syst. Biol.* **12**, 895 (2016).
- [29] A. Klimek, J. C. J. Heyn, D. Mondal, S. Schwartz, J. O. Rädler, P. Sharma, S. Block, and R. R. Netz, Intrinsic cell-to-cell variance from experimental single-cell motility data, *PRX Life* **3**, 023015 (2025).
- [30] E. Lemaître, I. M. Sokolov, R. Metzler, and A. V. Chechkin, Non-Gaussian displacement distributions in models of heterogeneous active particle dynamics, *New J. Phys.* **25**, 013010 (2023).
- [31] A. G. Cherstvy, O. Nagel, C. Beta, and R. Metzler, Non-Gaussianity, population heterogeneity, and transient superdiffusion in the spreading dynamics of amoeboid cells, *Phys. Chem. Chem. Phys.* **20**, 23034 (2018).
- [32] J. Albrecht, M. Opper, and R. Großmann, Inferring parameter distributions in heterogeneous motile particle ensembles: A likelihood approach for second order Langevin models, *arXiv:2411.08692*.
- [33] M. Delattre and M. Lavielle, Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models, *Stat. Its Interface* **6**, 519 (2013).
- [34] U. Picchini, A. de Gaetano, and S. Ditlevsen, Stochastic differential mixed-effects models, *Scand. J. Stat.* **37**, 67 (2010).
- [35] M. J. Schervish, Theory of Statistics, *Springer Series in Statistics* (Springer, New York, 1995).
- [36] S. Donnet and A. Samson, A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models, *Adv. Drug Deliv. Rev.* **65**, 929 (2013).
- [37] W. F. Loomis, D. Fuller, E. Gutierrez, A. Groisman, and W.-J. Rappel, Innate non-specific cell substratum adhesion, *PLoS One* **7**, e42033 (2012).
- [38] V. Lepro, R. Großmann, S. Sharifi Panah, O. Nagel, S. Klumpp, R. Lipowsky, and C. Beta, Optimal cargo size for active diffusion of biohybrid microcarriers, *Phys. Rev. Appl.* **18**, 034014 (2022).
- [39] S. Sharifi Panah, R. Großmann, V. Lepro, and C. Beta, Cargo size limits and forces of cell-driven microtransport, *Small* **20**, 2304666 (2024).
- [40] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/pvsc-8sbp> for experimental methods and supplementary datasets, which includes Ref. [51].
- [41] L. S. Dautzenberg, J. Albrecht, R. Großmann, and C. Beta, Trajectories of polystyrene beads driven by a carpet of amoeboid

- cells [Data set], Zenodo (2025), doi: <https://doi.org/10.5281/zenodo.17804335>.
- [42] V. Genon-Catalot and C. Larédo, Estimation for stochastic differential equations with mixed effects, *Statistics* **50**, 1014 (2016).
- [43] M. Delattre, V. Genon-Catalot, and A. Samson, Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient, *ESAIM: PS* **19**, 671 (2015).
- [44] H. Risken, *The Fokker-Planck Equation. Methods of Solutions and Applications*, Springer Series in Synergetics, 2nd ed. (Springer, Berlin, 1996).
- [45] M. V. Chubynsky and G. W. Slater, Diffusing diffusivity: A model for anomalous, yet Brownian, diffusion, *Phys. Rev. Lett.* **113**, 098302 (2014).
- [46] L. Golé, C. Rivière, Y. Hayakawa, and J.-P. Rieu, A quorum-sensing factor in vegetative *Dictyostelium discoideum* cells revealed by quantitative migration analysis, *PLoS One* **6**, e26901 (2011).
- [47] S. Annesley and P. Fisher, *Dictyostelium discoideum*—A model for many reasons, *Mol. Cell. Biochem.* **329**, 73 (2009).
- [48] D. Arcizet, S. Capito, M. Gorelashvili, C. Leonhardt, M. Vollmer, S. Youssef, S. Rappl, and D. Heinrich, Contact-controlled amoeboid motility induces dynamic cell trapping in 3D-microstructured surfaces, *Soft Matter* **8**, 1473 (2012).
- [49] G. Ariel and A. Ayali, Locust collective motion and its modeling, *PLoS Comput. Biol.* **11**, e1004522 (2015).
- [50] S. Lambert, M. Duchêne, and S. Klumpp, Bayesian inference of wall torques for active Brownian particles, *Europhys. Lett.* **149**, 57001 (2025).
- [51] R. K. Sadhu, M. Luciano, W. Xi, C. Martinez-Torres, M. Schröder, C. Blum, M. Tarantola, S. Villa, S. Penič, A. Iglič, C. Beta, O. Steinbock, E. Bodenschatz, B. Ladoux, S. Gabriele, and N. S. Gov, A minimal physical model for curvotaxis driven by curved protein complexes at the cell's leading edge, *Proc. Natl. Acad. Sci. USA* **121**, e2306818121 (2024).