From collective motion to single particle trajectory segmentation

Insights into particle dynamics at the microscopic scale

El Korde, Ismail

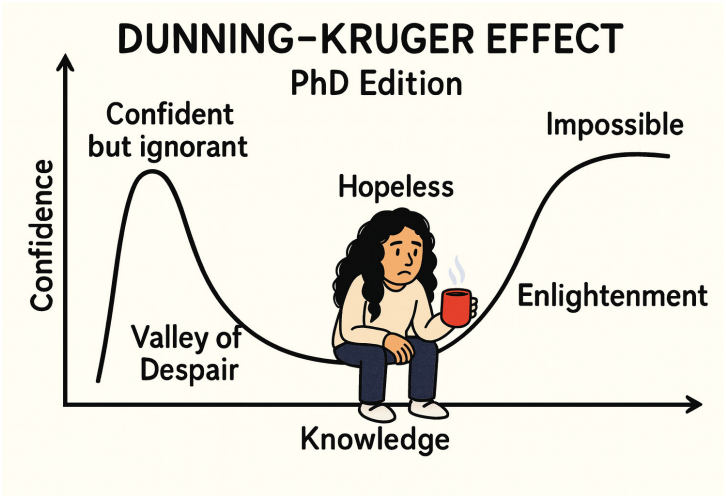
2025

Document Version:
Publisher's PDF, also known as Version of record

From collective motion to single particle trajectory segmentation

Insights into particle dynamics at the microscopic scale

ISMAIL EL KORDE
DIVISION OF PHYSICAL CHEMISTRY | FACULTY OF SCIENCE | LUND UNIVERSITY



Printed by Media-Tryck, Lund 2025
NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Faculty of Science
Department of Chemistry
Division of Physical Chemistry

ISBN 978-91-8096-138-7

From collective motion to single particle trajectory segmentation

From collective motion to single particle trajectory segmentation

Insights into particle dynamics at the microscopic scale

by Ismail El Korde

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Science, Lund University, Sweden.
To be defended on Thursday, the 11th of December 2025 at 13:00 in lecture hall A, Kemicentrum, Lund University.

Faculty opponent
Prof. Ignacio Pagonabarraga
University of Barcelona

Organization
LUND UNIVERSITY

Department of Chemistry
Box 124
SE–221 00 LUND
Sweden

Author(s)
Ismail El Korde

Document name
DOCTORAL DISSERTATION

Date of disputation
2025-12-11

Sponsoring organization

Title and subtitle
From collective motion to single particle trajectory segmentation: Insights into particle dynamics at the microscopic scale

Abstract
The study of physical systems at the microscopic scale often reveals emergent behaviors driven by thermal fluctuations and interactions. The characterization of such systems permits a deeper understanding of how these driving forces give rise to collective phenomena and functional responses. This thesis investigates two classes of microscopic systems: momentum-conserving active matter, or so-called "wet" active matter, where the constituents have the ability to transform energy into motion and interact through hydrodynamic interactions; and passive Brownian particles, where diffusion is transient due to thermal fluctuations and molecular interactions.

In wet active matter, interactions between active stresses and fluid flows lead to complex spatiotemporal dynamics that differ from those in dry active systems. While giant number fluctuations (GNFs), characterized by superlinear scaling of density fluctuations with system size, are a feature of dry active matter with long-range orientational order, theoretical analyses suggest that long-range nematic order is unstable in wet active matter due to hydrodynamic instabilities at finite wavelengths. Consequently, the mechanism responsible for GNFs in dry systems may not operate at large scales in wet active systems. Using large-scale lattice Boltzmann simulations of dilute suspensions of pusher-type microswimmers in three-dimensional unbounded fluids, this work demonstrates that enhanced, super-Gaussian number fluctuations emerge above the transition to collective motion. However, these fluctuations are confined to length scales comparable to the swimmers' persistence length ℓ_p, which also sets the typical size of transient nematic domains. At larger scales, number fluctuations return to Gaussian statistics. We also address the role of dimensionality by investigating two-dimensional microswimmer suspensions, our results show that the transition is almost completely smoothened out for small to intermediate values of ℓ_p. Furthermore, for large values of ℓ_p , we observe a discontinuous transition to a characteristic, stationary state with extensile flow structures spanning the whole system.

The second project addresses passive Brownian particles switching between diffusive states, typical in molecular binding or conformational changes. A computationally efficient segmentation method combines Gaussian filtering of displacement time series with Gaussian mixture models to classify states. Tests on synthetic two-state trajectories show high accuracy across diffusion constants and lifetimes, with robustness to motion blur and localization noise. This approach offers a transparent, lightweight alternative for real-time single-particle tracking analysis.

Key words
Active matter, lattice Boltzmann, microswimmers, collective motion, single particle tracking, segmentation

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-8096-138-7 (print)
978-91-8096-139-4 (pdf)

Recipient's notes

Number of pages
123

Price

Security classification

Signature _____ Date 2025-11-03

From collective motion to single particle trajectory segmentation

Insights into particle dynamics at the microscopic scale

by Ismail El Korde

LUND
UNIVERSITY

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the doctoral student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration front: Collective motion across scale (DALL-E)

Cover illustration back: Dunning-Kruger effect according to a PhD student (DALL-E)

Funding information: The thesis work was financially supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Faculty of Science, Department of Chemistry, Division of Physical Chemistry

isbn: 978-91-8096-138-7 (print)
isbn: 978-91-8096-139-4 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2025

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Dedicated to my parents
Siham and Ahmed

Contents

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

i **Pseudogiant Number Fluctuations and Nematic Order in Microswimmer Suspensions**
Ismail El Korde, Dóra Bárdfalvy, Jason M. Lewis, Alexander Morozov, Cesare Nardini, and Joakim Stenhammar
Phys. Rev. Lett. **2025**, 135, 098302

ii **Collective behaviour in unbounded microswimmer suspensions: The role of dimensionality**
Ismail El Korde, Alexander Morozov, Cesare Nardini, and Joakim Stenhammar
Manuscript

iii **A light-weight segmentation method for multi-state Brownian trajectories**
Ismail El Korde, Jason M. Lewis, Erik Clarkson, Tommy Dam, Peter Jönsson, Tobias Ambjörnsson, and Joakim Stenhammar
Submitted to npj Biological Physics and Mechanics

Author Contributions

i **Pseudogiant Number Fluctuations and Nematic Order in Microswimmer Suspensions**
I performed the simulations and the analysis. I contributed to the writing of the manuscript.

ii **Collective behaviour in unbounded microswimmer suspensions: The role of dimensionality**
I performed the simulations, the analysis, and contributed to the writing of the manuscript.

iii **A light-weight segmentation method for multi-state Brownian trajectories**

I contributed to the conceptualisation of the research and the design of the model. I conducted the software implementation and data analysis. I wrote the manuscript with input from my collaborators.

All papers are reproduced with permission of their respective publishers.

Acknowledgements

First and foremost, I would like to express that I have been an exceptionally fortunate and deeply content person to have been surrounded by individuals from whom I have learned immensely, both intellectually and culturally. I am profoundly grateful to all of you, and it has truly been an honour to meet and share this journey with you.

I would like to extend my heartfelt gratitude to my supervisor, **Joakim**, for the invaluable opportunity you granted me. Working under your guidance has been an extraordinary privilege. I have learned a great deal by your side and have developed many essential skills throughout this journey. One of the aspects I appreciated most was the level of autonomy you entrusted me with, an autonomy that came hand in hand with a genuine sense of responsibility, which helped me to grow both personally and professionally. Thank you for the wisdom and insight you so generously shared with me. You were far more than a supervisor; you were a constant source of support and encouragement, particularly during the most challenging times. I am truly grateful for all I have learned from you and will never forget the opportunity you provided me with. Without this opportunity, I would not have met so many wonderful people who have become true friends, individuals I sincerely hope to remain in contact with throughout my life.

I also want to convey my sincerest thanks to **Damien**, without whom I would not have reached this point. Thank you for the two exceptional years we spent working side by side. I learned a great deal from you, and your support has never wavered, from our very first meeting until today. I am sincerely thankful for your guidance, kindness, and patience.

My deepest thanks go to my late friend **Matthias**. I wish you could have been here to share this moment with me. Thank you for all the support you gave me, for always being there to help, often several times a day. You were someone without whom I could not have come this far. I learned so much from you, and I will be forever grateful for having known you.

My closest friend during this journey **Jason**. Thank you for all the long nights we spent working together, for the challenges we faced, and for the many lessons I learned from you. I truly consider you one of my closest friends, and I am deeply grateful for the part you played in shaping the person I have become.

I am also profoundly grateful to **Madeleine**. Thank you for your unwavering support during the most difficult times and for always standing by me. I wish you all the success and happiness you deserve in the future.

Arif, **Antara**, **Iria**, **Pappu**, **Najla**, and **Nikol**. You were all an integral part of my PhD

journey, the people I could turn to in moments of joy as well as difficulty. You were with me through every stage of my thesis, and I cannot imagine what I would have done without you. Thank you.

I would also like to thank both **David**s. Without you, the atmosphere at work would not have been the same. You were always there to listen, to share a laugh, and to celebrate the small victories. Thank you for all the wonderful moments we shared.

I would like to express my gratitude to **everyone in both Physical Chemistry and Computational Chemistry**. Each of you has contributed, in one way or another, to my academic and personal development, and to the achievement I am proud to celebrate today.

Finally, I would like to express my deepest gratitude to my parents, **Ahmed** and **Siham**, my aunt and second mother, **Karima**, and to my siblings, **Mohammed** and **Meryem**, for always standing by my side. Without your love, patience, and unwavering support, I would not be the person I am today. I thank you all from the bottom of my heart.

Popular scientific summary

At very small scales, matter is restless. Heat makes molecules jiggle, tiny swimmers push on the surrounding liquid, and together they create patterns and motion that are not obvious from the parts alone. This thesis examines two settings: (1) "wet" active matter, fluids filled with microscopic self propelled swimmers that consume energy to move and interact through the fluid, and (2) ordinary particles that drift randomly yet switch between slow and fast diffusion, as happens when molecules bind or change shape.

In active fluids, the swimmers do not just move; they stir the liquid, and the liquid stirs them back. That feedback produces collective motion: bursts, swirls, and fleeting patches where many swimmers line up. In dry active systems, these alignments can cause "giant number fluctuations", meaning that if you count particles in a box, the counts vary far more than simple randomness would allow. Theory indicates that in wet active matter such long-range alignments cannot persist, because fluid instabilities break them up. Using large-scale lattice Boltzmann simulations of dilute suspensions of pusher type swimmers in three-dimensional, unbounded liquids, this thesis shows that unusually large clumps and gaps appear once collective motion sets in, but only up to a characteristic length given by the persistence length ℓ_p, which also fixes the typical size of transient nematic domains. At larger scales, number fluctuations return to ordinary Gaussian statistics. Dimensionality matters as well: in two dimensions, the onset of collective motion is largely smoothed out when ℓ_p is small to intermediate, whereas for large ℓ_p there is a discontinuous jump to a stationary state filled with extensile flow structures that span the whole system.

The second part of the thesis tackles a different but common behaviour: particles that wander randomly yet switch between slow and fast modes, as biomolecules do when they bind, unbind, or change conformation. I develop a lightweight analysis method that filters noisy position data and then uses a simple statistical model to label which mode a particle is in at each moment. On carefully designed test data, the approach identifies state changes accurately across a wide range of diffusion speeds and lifetimes, and it remains reliable even when measurements are blurred or noisy. Because it is transparent and computationally cheap, it is well suited for analysis in real-time during single-particle tracking experiments.

From collective motion to single particle trajectory segmentation: Insights into particle dynamics at the microscopic scale

Chapter 1

Introduction

Robert Brown's observation (1827) of the irregular motion of microscopic particles provided an early empirical basis for what is now called Brownian motion [1]. Subsequent theory connected these observations to molecular agitation: Einstein (1905) and Smoluchowski (1906) [2, 3] showed that the mean-squared displacement grows linearly with time,

$$\langle \Delta r^2(t) \rangle = 2\, d\, D\, t,$$

and related the diffusivity D to temperature T and viscous drag (e.g., the Stokes-Einstein relation $D = k_{\mathrm{B}}T/(6\pi\mu a)$ for a sphere of radius a in a fluid of viscosity μ) [2, 4]. Langevin (1908) formulated the corresponding stochastic equation of motion [5], and later developments linked such dynamics to Fokker-Planck descriptions and, more broadly, fluctuation-dissipation relations [6–8]. At microscopic scales, thermal fluctuations therefore render particle trajectories effectively stochastic; although random motion has macroscopic analogues (e.g., granular diffusion), thermal Brownian motion is most naturally observed for microscopic particles in fluids.

The microscopic realm, however, encompasses systems with very different driving. Some are well approximated by equilibrium thermal fluctuations, while others are maintained far from equilibrium by sustained energy input at the particle scale. *Active matter* [9–11] refers to assemblies of units that convert environmental or stored energy into motion. Examples include vibrated rods, swimming bacteria, cytoskeletal filaments driven by molecular motors, and catalytically propelled Janus colloids [11–18]. Such systems can exhibit sustained self-driven transport, emergent order (polar or nematic alignment), motility-induced phase separation, and, in fluid-immersed ('wet') settings, hydrodynamically mediated collective flows such as mesoscale turbu-

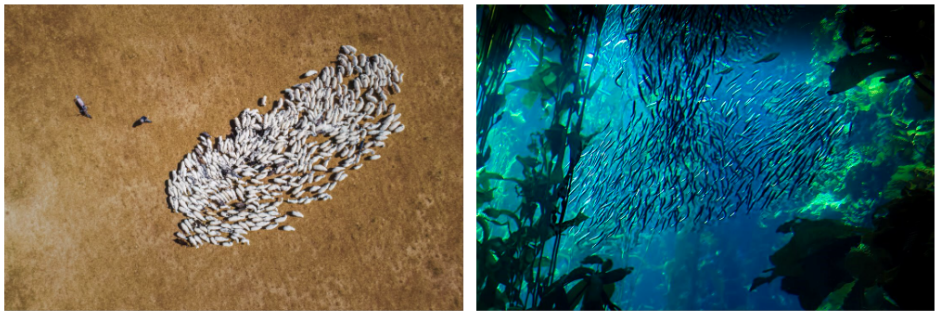


Figure 1.1: Natural instances of collective motion: (left) sheep forming a cohesive herd [35], (right) fish forming a school [36].

lence with jets and vortices, so-called *bacterial turbulence* [14, 15, 19–24]. By contrast, *passive tracers* do not generate motion; their motion is induced by thermal fluctuations of the surrounding fluid, arising from incessant microscopic collisions with solvent molecules. In homogeneous Newtonian fluids they undergo Gaussian diffusion with a single diffusivity, but in heterogeneous or complex media, crowded cytoplasm, or spatially varying viscosity, they often display non-Gaussian displacement distributions, transient sub- or super-diffusion, and time- or space-dependent effective mobilities [25–29].

In light of these considerations, the present thesis addresses both classes: Part I focuses on wet active matter in bacterial suspensions [14, 15], quantifying the emergence of anomalous density fluctuations near the onset of *bacterial turbulence*, how the associated scaling depends on control parameters, and where departures from Gaussian statistics are confined to mesoscopic scales set by the persistence length of the individual swimmer. Part II turns to single-particle tracking, introducing a fast, transparent segmentation pipeline [30–34] that denoises displacements, clusters trajectories into regions of similar diffusivity, and estimates the corresponding transport parameters. Together, these studies examine how minimal, coarse-grained descriptions can be distilled from complex dynamics across both active and passive settings.

1.1 Active matter and giant number fluctuations

Here we begin with representative *dry active matter* systems [37], where particles move on or through a momentum-dissipating substrate so that interactions, rather than long-ranged hydrodynamics, mediate collective behaviour. The *Vicsek model* is the canonical polar flock [19–21]: point-like agents advance at fixed speed while rotating their headings towards the local mean direction with noise, producing true long-

range orientational order in two dimensions, travelling high-density bands, and a strong coupling between orientation and density. A closely matched experimental realisation is provided by *Quincke rollers* [38–40]: dielectric colloids set into steady rolling by an applied electric field, which self-propel along a surface, exhibit effective short-range alignment (with electro-hydrodynamic couplings), form polar flocks with travelling bands, and permit direct tests of fluctuation scaling in the ordered phase. Another important class comprises *substrate-based active nematics* [22, 41] built from apolar, elongated constituents with head-tail symmetry; theory anticipates, and both simulations and experiments confirm, large fluctuation effects in these systems.

A hallmark of these dry realisations of active matter is *anomalous density fluctuations* [12, 21, 42, 43], far larger than those expected from systems at equilibrium. The strong coupling between local order and density, amplified by propagating orientational distortions and banding instabilities, generically boosts number fluctuations beyond equilibrium scaling.

To make this precise, recall the equilibrium baseline set by the central-limit theorem. For the number of particles N inside an observation window with mean $\langle N \rangle$, the standard deviation scales as

$$\Delta N \sim \langle N \rangle^{1/2}.$$

Giant Number Fluctuations (GNFs) are defined as a super-Poissonian growth of fluctuations with system size,

$$\Delta N \sim \langle N \rangle^{\alpha} \qquad \text{with } \alpha > 1/2,$$

a behaviour predicted for dry polar and nematic orders (e.g., within Toner-Tu theory for flocks) [20, 42] and observed in simulations and several experiments [43, 44].

By contrast, *wet active matter* involves constituents suspended in a viscous fluid that conserves momentum [45, 46]. Canonical examples are bacterial suspensions of *E. coli* or *B. subtilis*, which develop *bacterial turbulence* with continually forming jets and swirls [13–15, 47]. A closely related wet platform is the *microtubule-kinesin active nematic* at an oil-water interface: motor-driven extensile stresses generate spontaneous flows, create and transport $\pm 1/2$ defects, and sustain persistent streaming [16–18, 22].

Returning to GNFs specifically in wet systems: number fluctuations in bacterial suspensions [48] are often enhanced over *mesoscopic* ranges because advection correlates distant regions; however, whether such enhancement persists to the largest accessible scales, and thus qualifies as *asymptotic* GNFs in the same sense as in dry ordered flocks, *remains an open question* [22, 37, 41].

While the first part interrogates nonequilibrium fluctuations in fluid-coupled active suspensions, the second part tackles a statistical problem in equilibrium systems: de-

tecting time-varying mobility states from finite, noisy trajectories. Both share a common aim, extracting reliable coarse-grained descriptions from fluctuating dynamics. This parallel motivates the thesis structure.

1.2 Segmentation of single-particle trajectories in passive Brownian systems

Over the past three decades, advances in optical microscopy have greatly expanded our ability to probe biological systems at the molecular scale. Among these, single-molecule fluorescence microscopy [49–51], used together with single-particle tracking (SPT) algorithms [52–58], has transformed the study of single-molecule behaviour in living cells by enabling high spatiotemporal resolution tracking of individual molecules. This combination has yielded important insights into intracellular transport, protein-protein interactions, and the mechanisms that govern cellular function [49, 53, 59–64].

Classical mean-square displacement (MSD) analysis works well for single-state Brownian motion, yet it becomes unreliable when diffusion varies across time or space [26, 32]. Such variability is common in SPT as trajectories are shaped by transient and complex interactions with the cellular environment, producing mixtures of diffusive states [64–68]. Disentangling these states and their transitions is essential for quantitative studies of protein-ligand binding and cellular signaling [69–71].

A rich ecosystem of alternatives exists. Hidden Markov models (HMMs) jointly infer state labels, diffusion coefficients, and transition probabilities and can even infer the number of states, but the computation scales poorly with model complexity and trajectory length [31, 33]. Deep-learning methods achieve strong accuracy for segmentation and even anomalous-diffusion classification, but they typically require large training sets and lack physical interpretability [72, 73]. Together, these trade-offs motivate algorithms that are accurate yet lightweight and transparent.

The work summarised in this thesis adopts a deliberately simple and interpretable strategy: lightly denoise step displacements and then group recurring movement patterns into a small number of mobility states using a data-driven, model-based separation. This produces time-resolved state labels for each trajectory; quantitative parameters such as diffusion coefficients and lifetimes are then estimated afterwards with standard tools. The emphasis is on clarity and speed rather than heavy modelling or large-scale training, making the approach practical for routine SPT workflows.

Part I

Collective motion in suspensions of microswimmers

Chapter 1

Fluid dynamics at low Reynolds number

The motion of Newtonian fluids is governed by the conservation of mass, momentum, and energy, expressed mathematically through the Navier-Stokes and continuity equations. Fundamentally, they are the local (differential) expressions of the conservation laws of mass and momentum, allowing the description of fluid dynamics [74, 75]:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{1.1}$$

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} \right) = -\nabla p + \mu \nabla^2 \mathbf{u} + \left(\tfrac{1}{3}\mu + \lambda \right) \nabla(\nabla \cdot \mathbf{u}) + \mathbf{f}. \tag{1.2}$$

$$\rho \left(\frac{\partial e}{\partial t} + \mathbf{u} \cdot \nabla e \right) = -p(\nabla \cdot \mathbf{u}) + \Phi + \nabla \cdot (k \nabla T) + \rho q_{\text{ext}} \tag{1.3}$$

where ρ is the density, $\mathbf{u}$ is the fluid velocity, t is the time, p is the pressure, μ is the dynamic viscosity, λ is the bulk viscosity, $\mathbf{f}$ is the force density, e is the specific internal energy, k is the thermal conductivity, T is the temperature, q_{ext} is the external heat source per unit mass, and Φ is the viscous dissipation due to deformation. For many biological systems immersed in a fluid, *e.g.* bacterial suspensions, a convenient approximation is to consider a fluid with constant density and temperature. In this incompressible and isothermal limit, the governing equations simplify dramatically to [4, 74, 76–78]:

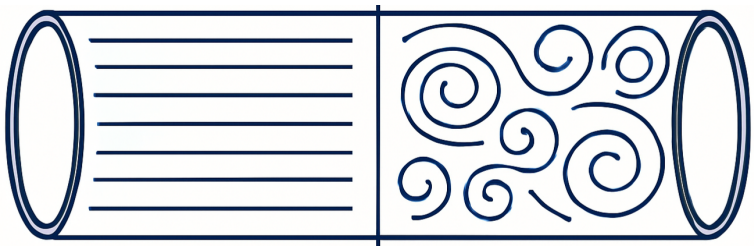


Figure 1.1: Comparison of laminar and turbulent flows inside cylindrical pipes. Laminar flow (left) exhibits smooth, parallel streamlines, whereas turbulent flow (right) shows chaotic, swirling structures.

$$\nabla \cdot \mathbf{u} = 0 \tag{1.4}$$

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{f} \tag{1.5}$$

Eq. (1.4) describes the incompressibility of the fluid, meaning that the fluid does not expand or contract, whereas in Eq. (1.5), the left-hand term corresponds to inertial effects, while the right-hand terms describe the effects of pressure gradients, external forces, and viscosity.

To characterise the nature of the fluid flow, a dimensionless quantity, the *Reynolds number* Re, is used to measure the balance between inertial and viscous contributions:

$$\mathrm{Re} = \frac{\rho u \ell}{\mu} \tag{1.6}$$

where u is a characteristic flow speed, ℓ is a characteristic length scale (such as the diameter of a pipe or the size of a swimmer), and μ is the dynamic viscosity of the fluid.

Reynolds number allows a qualitative classification of flows into three regimes: the laminar regime, where viscous contributions dominate the inertial ones; the transitional regime, where inertial effects begin to emerge and the flow becomes sensitive to perturbations; and the turbulent regime, where inertial forces are dominant and the flow exhibits chaotic, vortical structures. This contrast between laminar and turbulent pipe flow is illustrated in Fig. 1.1. A further classification of laminar flows arises in the limit of very low Reynolds number, specifically when $\mathrm{Re} \ll 1$, commonly referred to as the *Stokesian regime*. In this regime, viscous forces overwhelmingly dominate inertial effects, resulting in a simplification of the governing equations. Neglecting inertial terms in the Navier-Stokes equations yields the Stokes equations [4, 76, 78]:

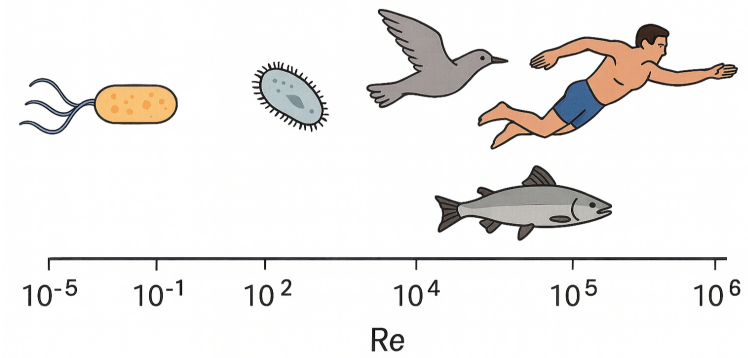


Figure 1.2: Representative Reynolds numbers across biological length scales, from bacteria ($Re \approx 10^{-5}$) to human swimming ($Re \approx 10^{6}$).

$$\begin{aligned} -\nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{f} &= 0, \\ \nabla \cdot \mathbf{u} &= 0. \end{aligned} \tag{1.7}$$

These equations describe incompressible and linear flow fields, where the fluid adjusts instantaneously to applied forces due to the absence of inertia. Typical Reynolds numbers across biological length scales are summarised in Fig. 1.2.

The Stokes regime exhibits *time-reversal symmetry*, implying that reversing the sequence of motions results in an exact reversal of the flow, a feature absent at higher Reynolds numbers. This regime is particularly relevant for the study of microscale systems suspended in a fluid. A canonical example, central to this thesis, is the locomotion of bacteria, which typically operate at Reynolds numbers on the order of Re $\approx 10^{-5}$. At these scales, traditional swimming strategies based on inertia (such as those used by fish or humans) become ineffective. The absence of inertia imposes strict physical constraints on motion at the microscale, where any reciprocal (*i.e.*, time-reversible) motion fails to produce net displacement. This limitation is formalised by the *Scallop Theorem*, which states that in a Newtonian fluid governed by the Stokes equations, net locomotion is impossible through time-symmetric actuation alone [79, 80]. Consequently, microorganisms must adopt *non-reciprocal* swimming strategies that break time-reversal symmetry in order to achieve propulsion. These strategies involve generating flows that differ between the forward and backward strokes of their gait, thereby circumventing the constraints of kinematic reversibility. A prominent example is the use of rotating helical flagella, employed by many motile bacteria such as *Escherichia coli*, which continuously rotate to generate net propulsion [76, 81].

Because the Stokes equations are linear, their solutions can be expressed as superpo-

sitions of flow singularities [76, 78]. These singularities serve as building blocks that represent the fluid disturbance caused by microswimmers or particles. Studying these singular solutions provides insights into the fluid dynamics of microorganisms and the hydrodynamic interactions governing their motion. Similarly to the traditional treatment of equations in electrodynamics [82], the Stokes equations can be addressed using a common set of analytical methods, notably Green's functions and multipole expansions, to obtain solutions for the flow and pressure fields generated by singularities and to analyse far-field behaviour.

A simple practical case is the directed motion of a colloidal particle suspended in a Newtonian fluid, where a point force $\mathbf{F}$ is applied to the particle. This results in a force density $\mathbf{f}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}')\,\mathbf{F}$, where $\mathbf{r}'$ is the position of the particle and δ is the Dirac delta function. Setting boundary conditions such that both the fluid velocity field and the pressure field vanish infinitely far from the source, we obtain the Green's function for the fluid velocity field in $3D$ [76, 83, 84]:

$$\mathbf{u}(\mathbf{r}) = \mathbf{G}(\mathbf{r} - \mathbf{r}') \cdot \mathbf{F}, \quad \text{with} \quad \mathbf{G}(\mathbf{r}) = \frac{1}{8\pi\mu}\left(\frac{1}{r}\mathbb{1} + \frac{\mathbf{rr}}{r^3}\right), \tag{1.8}$$

where $r = |\mathbf{r}|$, $\mathbf{G}(\mathbf{r})$ commonly known as the *Oseen tensor*, and $\mathbb{1}$ the identity tensor. The velocity field due to a point force in Eq. (1.8) is known as the *Stokeslet*, corresponding to a force monopole acting on the fluid.

Higher-order moments of the flow field can be obtained through a multipole expansion, in which the *Stokeslet* represents the leading-order (monopole) term. Of particular interest is the next-order term corresponding to a force dipole, whose symmetric part is known as the *stresslet* [76, 83–85]:

$$\mathbf{u}(\mathbf{r}) = \frac{\kappa}{8\pi}\left(\frac{3(\mathbf{p}\cdot\mathbf{r})^2}{r^5} - \frac{1}{r^3}\right)\mathbf{r}, \tag{1.9}$$

where $\mathbf{p}$ denotes the unit orientation vector of the force dipole, and $\kappa = \pm\frac{Fl}{\mu}$ is the dipolar strength associated with an extended dipole composed of two opposite forces of magnitude F separated by a distance l. For a 2D fluid, the corresponding Oseen tensor is [86]

$$\mathbf{G}(\mathbf{r}) = \frac{1}{4\pi\mu}\left(\frac{\mathbf{rr}}{r^2} - \ln(r)\,\mathbb{1}\right), \tag{1.10}$$

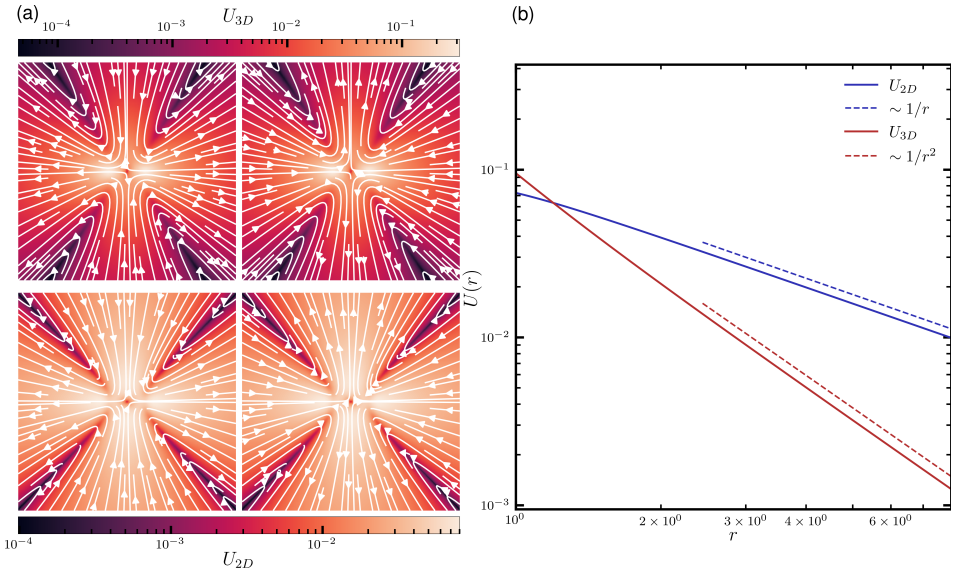


Figure 1.3: Velocity fields and asymptotic decay of the stresslet in 2D and 3D. (a) (Top) Velocity field generated by (left) a regularised 3D pusher and (right) a regularised puller in the $z = 0$ plane. (Bottom) Velocity field generated by (left) a regularised 2D pusher and (right) a regularised 2D puller (b) Far-field decay of stresslet magnitudes, showing the r^{-1} scaling in 2D and the r^{-2} scaling in 3D.

and the corresponding stresslet is [76, 86]

$$\mathbf{u}(\mathbf{r}) = \frac{\kappa}{4\pi}\left(\frac{2\,(\mathbf{p}\cdot\mathbf{r})^2}{r^4} - \frac{1}{r^2}\right)\mathbf{r}. \tag{1.11}$$

The corresponding velocity fields and their far-field decays are shown in Fig. 1.3. The sign of κ is positive when the forces are directed away from each other (extensile or "pusher" dipole) and negative when they point towards each other (contractile or "puller" dipole). Unlike the $2D$ Stokeslet, which produces $\log r$ growth and makes the unbounded problem ill-posed, a force-dipole (stresslet) decays as $1/r$ in $2D$ and $1/r^2$ in $3D$, thereby satisfying $u \to 0$ as $r \to \infty$.

For microorganisms such as *Escherichia coli* or *Chlamydomonas reinhardtii*, which swim in a torque-free and force-free manner [76], the stresslet provides a quantitative description of the flow field generated by their swimming motion. These long-range flow fields induce complex hydrodynamic interactions among microorganisms, which can lead to emergent collective behaviours observed in bacterial suspensions, as will be discussed in Chapter 2.

Chapter 2

Interactions and the emergence of collective motion

As discussed in Chapter 1, at low Reynolds numbers, the locomotion of microorganisms is dominated by viscous forces and is well described by the Stokes equations. Crucially, breaking time-reversal symmetry through their swimming gait enables net propulsion. Individual microswimmers therefore generate flows in the surrounding fluid that can be represented, to a first-order approximation, by force dipoles [79]. This force dipole perturbs the fluid and produces a long-ranged dipolar flow field [76] as shown in Eqs. (1.9)-(1.11).

As illustrated in Fig. 2.1 for an *E. coli* bacterium, this description is an accurate approximation for dilute suspensions in an unbounded medium, where the separation between swimmers is large enough that higher-order contributions are negligible relative to the stresslet, short-ranged non-hydrodynamic interactions can be ignored, and near-boundary effects are absent. This regime is commonly referred to as the far-field approximation. The flows generated by these stresslets influence the surrounding fluid and mediate interactions, *i.e.* hydrodynamic interactions, between microorganisms.

While the analysis of an isolated microswimmer provides valuable insight into individual locomotion, real biological systems often consist of suspensions containing large numbers of microorganisms. In these systems, the long-range flow fields produced by each swimmer superpose, giving rise to complex, many-body emergent phenomena. Hydrodynamic interactions in such suspensions influence both the translational and rotational dynamics of the microswimmers: advection by the fluid drives positional changes, and velocity gradients in the fluid induce rotations that alter swimmer orientation thus shaping the emergent collective behaviour. As illustrated in 2.2, for pusher

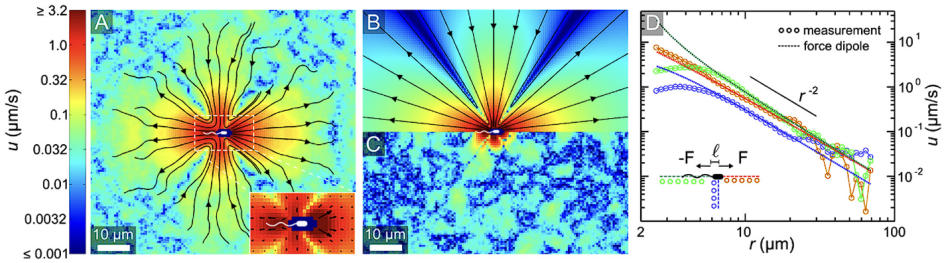


Figure 2.1: Flow field generated by a swimming *E. coli* bacterium. (A) Experimentally measured velocity field in the plane of the swimmer, shown as a colour map of the speed magnitude u with streamlines overlaid in black. The inset highlights the near-field velocity distribution. (B) Best fit to the analytical flow field of a pusher-type force dipole. (C) Residual flow field obtained by subtracting the fitted field from the experimental data, which shows that higher-order contributions dominate in the near field. (D) Radial decay of the speed magnitude u extracted from experimental data in the parallel direction to the swimming orientation (red and green circles) and perpendicular to it (blue circles). In the far field, the data exhibit the characteristic r^{-2} scaling expected for a force dipole. Reproduced from [87].

microswimmers, interactions tend to align neighbouring swimmers, promoting nematic alignment along their swimming direction, whereas for puller microswimmers, they induce effective repulsion, causing the swimmers to move apart. Understanding these interactions provides a foundation for analysing the emergent collective dynamics observed in bacterial suspensions [14, 88].

2.1 Emergence of bacterial turbulence in microswimmer suspensions

2.1.1 Microswimmer model

We consider spatial dimensions $d \in \{2, 3\}$, with swimmer positions $\mathbf{r} \in \mathbb{R}^d$ and unit orientation vectors $\mathbf{p} \in \mathbb{S}^{d-1}$. The suspending fluid is Newtonian with dynamic viscosity μ and mass density ρ. Each swimmer self-propels at constant speed v_s along its body axis $\mathbf{p}$. The far-field hydrodynamic signature of a force- and torque-free swimmer is represented as an extended force dipole of length l: two equal and opposite point forces $\pm F\,\mathbf{p}$ applied at positions $\mathbf{r}$ and $\mathbf{r} - l\mathbf{p}$. We parameterise the dipole by the stresslet strength κ:

$$\kappa \equiv \frac{F\,l}{\mu}, \tag{2.1}$$

with $\kappa > 0$ for pushers and $\kappa < 0$ for pullers.

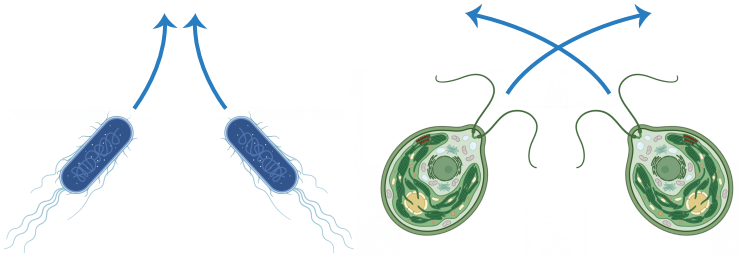


Figure 2.2: Mutual reorientation due to hydrodynamic interactions. (Left) The hydrodynamic interactions between pushers promote side-by-side alignment. (Right) The hydrodynamic interactions between pullers promote anti-alignment.

The dynamics of individual microswimmers in the presence of fluid-mediated interactions can be described by the following equations of motion:

$$\dot{\mathbf{r}} = v_s \mathbf{p} + \mathbf{u}(\mathbf{r}), \tag{2.2}$$

$$\dot{\mathbf{p}} = (\mathbb{1} - \mathbf{pp}) \cdot [\mathbf{W} + B\mathbf{E}] \cdot \mathbf{p}. \tag{2.3}$$

Here, B is the Bretherton parameter where $B = 0$ corresponds to spherical swimmers and $B \to 1$ to needle-like ones [83], $\mathbf{W}$ is the vorticity tensor and $\mathbf{E}$ is the rate-of-strain tensor of the surrounding fluid, defined respectively as

$$\mathbf{W} = \frac{1}{2}\left[\boldsymbol{\nabla}\mathbf{u} - (\boldsymbol{\nabla}\mathbf{u})^{\top}\right], \tag{2.4}$$

$$\mathbf{E} = \frac{1}{2}\left[\boldsymbol{\nabla}\mathbf{u} + (\boldsymbol{\nabla}\mathbf{u})^{\top}\right]. \tag{2.5}$$

Eq. (2.2) describes the displacement of microswimmers due to self-propulsion and fluid advection, whereas Eq. (2.3) corresponds to the Jeffery equation [89], governing the orientation of a rigid ellipsoid immersed in a viscous fluid at low Reynolds number in the presence of fluid flows. In addition to hydrodynamic reorientation, swimmers undergo run-and-tumble dynamics: tumbles occur as a Poisson process at rate λ, followed by an isotropic reorientation of $\mathbf{p}$. In quiescent fluid this produces a persistent random walk with persistence length $\ell_p = v_s/\lambda$. The rationale for modelling microswimmers as rigid ellipsoids is twofold. (i) Many natural microswimmers, such as bacteria and sperm cells, are elongated and approximately ellipsoidal [76]. (ii) Treating the body as rigid simplifies the problem and improves computational efficiency: propulsion typically arises from appendages (e.g., flagella) rather than body

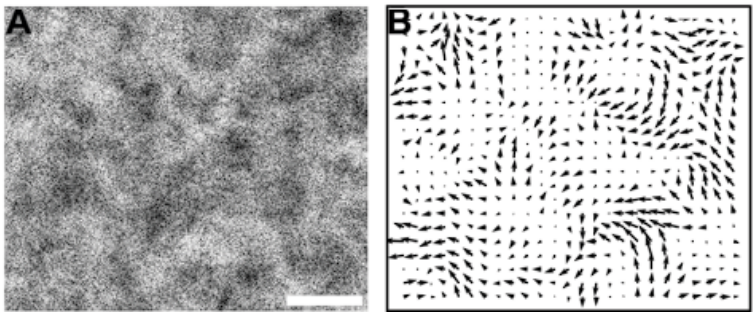


Figure 2.3: Active turbulence in a suspension of *E. coli*. **(A)** Bright-field snapshot of a dense bacterial suspension in the turbulent state. **(B)** In-plane velocity field $\mathbf{u}(x, y)$ extracted from (A) by particle-image velocimetry (PIV), revealing vortices and jets characteristic of active turbulence. Reproduced from [48].

deformation, which enables the use of established analytical treatment within Stokes flow [89], notably via Eq. (2.3), and eliminates the need to compute complex fluid-structure interactions.

2.1.2 Kinetic theory and the mean-field approximation

To describe the emergence of collective motion in dilute suspensions of microswimmers, where far-field hydrodynamic interactions dominate, kinetic theory offers a framework that describes the macroscopic behaviour of the system starting from a microscopic description of its constituents [90–94].

We consider a suspension of N identical microswimmers, each labelled by an index $i \in \{1, 2, \ldots, N\}$ having a position $\mathbf{r}_i \in \mathbb{R}^d$ and orientation $\mathbf{p}_i \in \mathbb{S}^{d-1}$, where d is system dimension. Let the full N-particle distribution function be

$$\psi^{(N)}(\mathbf{r}_1, \mathbf{p}_1, \ldots, \mathbf{r}_N, \mathbf{p}_N, t). \tag{2.6}$$

Each swimmer evolves according to Eq. (2.2) and Eq. (2.3) in addition to isotropic tumbling with rate λ. The N-particle distribution function evolves according to the Fokker-Planck equation [93, 95]:

$$\begin{aligned} \partial_t \psi^{(N)} + \sum_{i=1}^{N} \boldsymbol{\nabla}_{\mathbf{r}_i} \cdot \left(\dot{\mathbf{r}}_i \, \psi^{(N)} \right) + \sum_{i=1}^{N} \boldsymbol{\nabla}_{\mathbf{p}_i} \cdot \left(\dot{\mathbf{p}}_i \, \psi^{(N)} \right) \\ = \lambda \sum_{i=1}^{N} \left(\left\langle \psi^{(N)} \right\rangle_{\mathbf{p}_i'} - \psi^{(N)} \right), \end{aligned} \tag{2.7}$$

where $\boldsymbol{\nabla}_{\mathbf{r}_i} \equiv \partial/\partial \mathbf{r}_i$ is the spatial gradient with respect to the position of particle i, $\boldsymbol{\nabla}_{\mathbf{p}_i}$ is the surface gradient on the unit sphere in orientation space

$$\boldsymbol{\nabla}_{\mathbf{p}_i} = \big(\mathbb{1} - \mathbf{p}_i \mathbf{p}_i\big) \cdot \frac{\partial}{\partial \mathbf{p}_i},$$

and the orientational average over the i-th swimmer defined as

$$\big\langle \psi^{(N)} \big\rangle_{\mathbf{p}_i'} := \int_{\mathbb{S}^{d-1}} \frac{1}{\Omega_d}\, \psi^{(N)}\big(\mathbf{r}_1, \mathbf{p}_1, \ldots, \mathbf{r}_i, \mathbf{p}_i', \ldots, \mathbf{r}_N, \mathbf{p}_N, t\big)\, d\mathbf{p}_i'.$$

The terms on the left hand side describe the temporal evolution of the probability density, and the terms on the right represent source and sink contributions ensuring conservation of probability in the orientational space for run-and-tumble motion. We introduce $s \in \{1, \ldots, N\}$ as the number of particles retained in a reduced description. The associated s-particle (marginal) distribution is obtained by integrating out the degrees of freedom of the remaining $N - s$ particles:

$$\begin{aligned}\psi^{(s)}(\mathbf{r}_1, \mathbf{p}_1, \ldots, \mathbf{r}_s, \mathbf{p}_s, t) = \frac{N!}{(N-s)!N^s} \int \prod_{j=s+1}^{N} \left(d\mathbf{r}_j \int_{\mathbb{S}^{d-1}} \frac{d\mathbf{p}_j}{\Omega_d} \right) \\ \times\, \psi^{(N)}(\mathbf{r}_1, \mathbf{p}_1, \ldots, \mathbf{r}_s, \mathbf{p}_s, \mathbf{r}_{s+1}, \mathbf{p}_{s+1}, \ldots, \mathbf{r}_N, \mathbf{p}_N, t).\end{aligned} \tag{2.8}$$

We obtain a set of equations which connect the evolution of the s-particle probability density function with the $(s+1)$-particle probability density function forming a hierarchical set of equations known as the BBGKY hierarchy [96–100]. At the mean-field level, we consider the one-particle distribution $\psi^{(1)}(\mathbf{r}, \mathbf{p}, t)$ $(s = 1)$. Integrating Eq. (2.7) over the degrees of freedom of the remaining $N-1$ swimmers then yields the Smoluchowski equation governing the evolution of $\psi^{(1)}$ in the absence of interparticle correlations [93, 94]:

$$\begin{aligned}\partial_t \psi^{(1)}(\mathbf{r}, \mathbf{p}, t) + \boldsymbol{\nabla}_{\mathbf{r}} \cdot \Big(\dot{\mathbf{r}} \psi^{(1)}(\mathbf{r}, \mathbf{p}, t) \Big) + \boldsymbol{\nabla}_{\mathbf{p}} \cdot \Big(\dot{\mathbf{p}} \psi^{(1)}(\mathbf{r}, \mathbf{p}, t) \Big) = \\ -\lambda \psi^{(1)}(\mathbf{r}, \mathbf{p}, t) + \lambda \int_{\mathbb{S}^{d-1}} \frac{1}{\Omega_d} \psi^{(1)}(\mathbf{r}, \mathbf{p}', t)\, d\mathbf{p}',\end{aligned} \tag{2.9}$$

Eq (2.9) provides a *mean-field* description of a dilute suspension of microswimmers obeying the set of equations of motion in Eq. (2.2) and Eq. (2.3). A trivial solution

to Eq. (2.9) is a solution that does not exhibit any spatial, orientational, or temporal structure corresponding to a uniform, stationary state where all particles are independent and identically distributed both in positions and orientations, the *homogeneous isotropic* state ψ_{HI}:

$$\psi_{\mathrm{HI}} = \frac{1}{V_d \Omega_d}, \tag{2.10}$$

where V_d is the volume of the system and Ω_d is the surface area of a d-sphere of unit radius.

2.1.3 Linear stability of the homogeneous isotropic state in the mean-field approximation

After establishing the homogeneous isotropic state as a solution for the system, we study its stability under a linear perturbation [91, 93, 94]. The perturbed probability density is defined as:

$$\psi(\mathbf{r}, \mathbf{p}, t) = \psi_{\mathrm{HI}} + \delta\psi(\mathbf{r}, \mathbf{p}, t), \tag{2.11}$$

where $\delta\psi(\mathbf{r}, \mathbf{p}, t)$ satisfies:

$$\int d\mathbf{r}\, d\mathbf{p}\, \delta\psi(\mathbf{r}, \mathbf{p}, t) = 0. \tag{2.12}$$

By inserting Eq. (2.11) in Eq. (2.9), and after standard manipulations, the perturbations to the density field $\delta\rho$ and the fluid velocity field $\delta\mathbf{u}$ due to $\delta\psi(\mathbf{r}, \mathbf{p}, t)$ can be obtained. For suspensions of puller microswimmers, both the density and the fluid velocity field are stable. In contrast, pushers exhibit an orientational instability arising from mutual swimmer reorientation, which manifests at the largest length scale of the suspension for densities above a critical density n_c. For an infinite system of dimension $d \in \{2, 3\}$ [90, 92–94]:

$$n_c^{\infty} = \frac{(d+2)\lambda}{B\kappa}. \tag{2.13}$$

In a confined system, finite-size effects must be considered, leading to a corrected critical density for $d = 3$ [101]:

$$\frac{n_c(H)}{n_c^\infty} = 1 + \frac{3}{10}\frac{2\pi\ell_p}{H} + \frac{1}{5}\left(\frac{2\pi\ell_p}{H}\right)^2, \tag{2.14}$$

where $\ell_p = v_s/\lambda$ is the persistence length and H is the smallest confinement dimension.

For densities $n > n_c$, regions of nematic alignment form as pusher swimmers mutually align. In a momentum-conserving fluid, however, these nematic patches are *generically* unstable due to the Simha-Ramaswamy instability (SR) [9, 102]. In pusher suspensions, a small bend of the director generates shear that advects and rotates neighbours, further amplifying the bend. This long-wavelength flow-orientation feedback destabilises the aligned state and drives spontaneous, chaotic flows commonly referred to as bacterial turbulence (Fig. 2.3) [13, 45].

Bacterial turbulence thus emerges via a cascade of hydrodynamic instabilities:

$$\text{HI state} \xrightarrow{n > n_c} \text{nematic alignment} \xrightarrow{\text{SR}} \text{bacterial turbulence}$$

2.1.4 Active nematics

For dense active suspensions, the kinetic theory framework becomes harder to use as other contributions need to be taken into account, *e.g.* short-range interactions, as well as higher order correlations, hindering theoretical progress. Alternative methods are thus needed to describe these systems.

For such dense suspensions, *e.g.* suspensions of elongated swimmers or mixtures of filamentous particles and motor proteins, the particles tend to align one with the other and a head-tail symmetry emerges, similar in nature to the one observed for *liquid crystals* [103]. A natural choice is then to build on theories developed for passive liquid crystals to describe the isotropic-nematic (IN) phase transition [103]. An example of such theory is the Landau-de Gennes theory where the system is described via a Landau-Ginzburg-like free energy density. The free energy density f is obtained by expanding in rotationally invariant combinations of the so-called nematic order parameter $\mathbf{Q}$ defined for a uniaxial nematic system as

$$\mathbf{Q} = \frac{d}{d-1} S\Big(\mathbf{pp} - \frac{\mathbb{1}}{d}\Big), \tag{2.15}$$

where d is the dimension of the system and S the largest eigenvalue of $\mathbf{Q}$ describing the magnitude of nematic alignment ($S = 1$ for a nematically aligned system and

$S = 0$ an isotropic system). The free energy density f that describes this system around the IN phase transition is

$$f = \frac{A}{2}\,\mathrm{Tr}\,\mathbf{Q}^2 + \frac{B}{3}\,\mathrm{Tr}\,\mathbf{Q}^3 + \frac{C}{4}\left(\mathrm{Tr}\,\mathbf{Q}^2\right)^2 + \frac{K}{2}\,|\boldsymbol{\nabla}\mathbf{Q}|^2, \tag{2.16}$$

where A, B, and C are material parameters controlling, respectively, the location, the order, and the stability of the IN phase transition. The parameter K denotes the elastic constant that penalises spatial distortions of the nematic order parameter within the one-constant approximation, *i.e.*, assuming equal bend, twist, and splay moduli [103].

In the presence of fluid flows, the hydrodynamics of liquid crystals are described by the nematodynamic equation [104]:

$$\partial_t\mathbf{Q} + \mathbf{u}\cdot\boldsymbol{\nabla}\mathbf{Q} - \mathbf{S} = \Gamma\mathbf{H} \tag{2.17}$$

where $\mathbf{S}$ is a co-rotation term defined as

$$\mathbf{S} = (\chi\,\mathbf{E} + \mathbf{W})\cdot\left(\mathbf{Q} + \frac{\mathbb{1}}{d}\right) + \left(\mathbf{Q} + \frac{\mathbb{1}}{d}\right)\cdot(\chi\,\mathbf{E} - \mathbf{W}) - 2\chi\left(\mathbf{Q} + \frac{\mathbb{1}}{d}\right)(\mathbf{Q} : \boldsymbol{\nabla}\mathbf{u}), \tag{2.18}$$

where χ denotes the flow-alignment parameter (tumbling parameter); $\mathbf{E}$ and $\mathbf{W}$ are, respectively, the rate-of-strain and the vorticity tensors, as defined in Eqs. (2.5)–(2.4); and $\Gamma\mathbf{H}$ represents a relaxation term driving the system towards equilibrium with Γ the rotational diffusivity and $\mathbf{H}$ defined as

$$\mathbf{H} = -\frac{\delta f}{\delta\mathbf{Q}} + \frac{\mathbb{1}}{d}\mathrm{Tr}\left(\frac{\delta f}{\delta\mathbf{Q}}\right). \tag{2.19}$$

The evolution of the fluid velocity field is governed by Navier-Stokes for an incompressible fluid:

$$\rho\left(\partial_t\mathbf{u} + \mathbf{u}\cdot\boldsymbol{\nabla}\mathbf{u}\right) = \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}, \qquad \boldsymbol{\nabla}\cdot\mathbf{u} = 0, \tag{2.20}$$

where the stress tensor $\boldsymbol{\sigma} = -p\mathbb{1} + \boldsymbol{\sigma}^{\mathrm{viscous}} + \boldsymbol{\sigma}^{\mathrm{elastic}}$ contains the contributions of the pressure and the viscous and elastic stresses, defined as

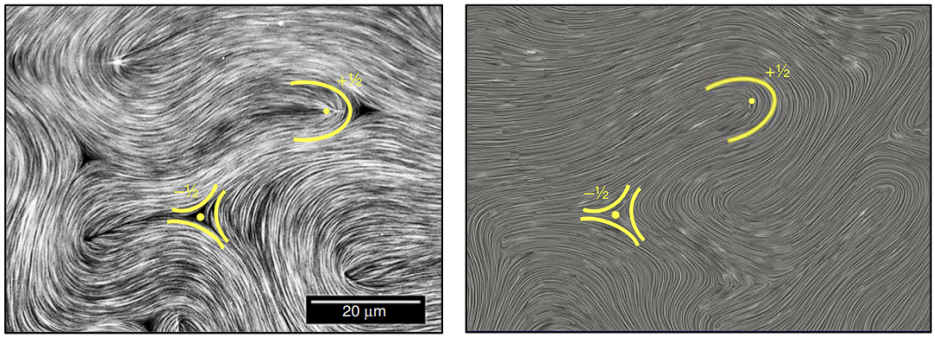


Figure 2.4: Active nematic turbulence. (Left) Fluorescence confocal micrograph of a microtubule–kinesin active nematic in the turbulent state. Yellow overlays mark comet-like $+1/2$ and trefoil-like $-1/2$ defects. (Right) Continuum nematohydrodynamic simulations reproduce similar swirling textures and the same $\pm 1/2$ topological defects. Reproduced from [22].

$$\boldsymbol{\sigma}^{\text{viscous}} = 2\mu\mathbf{E}, \tag{2.21}$$

$$\begin{aligned}\boldsymbol{\sigma}^{\text{elastic}} = {} & 2\chi\left(\mathbf{Q}+\frac{\mathbb{1}}{d}\right)(\mathbf{Q}:\mathbf{H}) - \chi\mathbf{H}\cdot\left(\mathbf{Q}+\frac{\mathbb{1}}{d}\right) \\ & - \chi\left(\mathbf{Q}+\frac{\mathbb{1}}{d}\right)\cdot\mathbf{H} + \mathbf{Q}\cdot\mathbf{H} - \mathbf{H}\cdot\mathbf{Q} \\ & - \boldsymbol{\nabla}\mathbf{Q} : \frac{\delta f}{\delta(\boldsymbol{\nabla}\mathbf{Q})}\end{aligned} \tag{2.22}$$

where μ is the viscosity, and : denotes the full index contraction.

Thus, the evolution of the nematic order parameter is strongly coupled to the flow velocity field through the advection term $\mathbf{u}\cdot\boldsymbol{\nabla}\mathbf{Q}$ and the co-rotation term $\mathbf{S}$ and similarly for the evolution of the flow field through the elastic stress $\boldsymbol{\sigma}^{\text{elastic}}$. Eq. (2.17) and Eq. (2.20) are the nematohydrodynamic equations of motion for liquid crystals.

Active nematics are obtained simply by adding an additional term to the stress tensor $\boldsymbol{\sigma}$, the active stress

$$\boldsymbol{\sigma}^{\text{active}} = -\zeta\mathbf{Q}, \tag{2.23}$$

where ζ is the activity coefficient: $\zeta > 0$ for pusher (extensile) and $\zeta < 0$ for pullers (contractile) [22]. The dependence of the active stress $\boldsymbol{\sigma}^{\text{active}}$ on the nematic order parameter $\mathbf{Q}$ arises from the coarse-graining of the dipolar fluid velocity fields induced

by the swimmers [22, 102, 105]. It should be noticed that, in contrast with kinetic theory where the microscopic description is at the centre of the theory, active nematics is a phenomenological theory that does not depend on the microscopic details of the system, but rather on a set of parameters that describe the system at the macroscopic scale. The emergence of active turbulence is also observed in active nematics, both in simulations [41] and experiments [16], due to the Simha-Ramaswamy instability. Thus, the uniform nematic order breaks down into a dynamic steady state characterised by creation, self-propulsion, and annihilation of topological defects ($\pm 1/2$ for $d = 2$, and disclination lines for $d = 3$). An example of active-nematic turbulence with clearly visible ±1/2 defects is shown in Fig. 2.4. The flow field exhibits vortices and jets of a characteristic size set by the active length scale ℓ_a [22]

$$\ell_a \sim \left(\frac{K}{|\zeta|} \right)^{\frac{1}{2}}. \tag{2.24}$$

Interestingly, as will be showcased later in this thesis, we also observe the emergence of a characteristic length scale at the onset of bacterial turbulence, the persistence length $\ell_p = v_s/\lambda$. Although distinct in microscopic origin from the active length scale $\ell_a \sim \sqrt{K/|\zeta|}$ of active nematics, it plays an analogous role in that it governs all emergent length scales of dilute microswimmer suspensions [106].

Chapter 3

Model and Method

Understanding the collective dynamics of microswimmer suspensions requires a modelling framework that captures both the behaviour of individual swimmers and the fluid-mediated interactions that couple them over long distances. Purely agent-based approaches are well suited to describing self-propulsion and Vicsek-type reorientation dynamics; however, they are tractable only numerically and are computationally demanding. By contrast, continuum models are analytically tractable but provide only an approximate description. Moreover, continuum descriptions are often mean-field; incorporating stochasticity and correlations typically requires significant additional effort.

We employ a hybrid model in which an off-lattice, agent-based representation of individual swimmers is coupled to a lattice Boltzmann (LB) fluid solver. Each swimmer is modelled as an extended force dipole that both generates and responds to local fluid flows. The LB method provides an efficient, weakly compressible approximation to isothermal hydrodynamics in two and three spatial dimensions, accurate at low Mach number ($\mathrm{Ma} \ll 1$). This enables *ab initio* studies of emergent, large-scale phenomena while retaining computational efficiency for systems as large as $N \sim 10^6$ swimmers. We first describe the swimmer model (Sec. 3.1), then formulate the LB solver and forcing scheme (Sec. 3.2), and finally describe the off-lattice point-force implementation (Sec. 3.3).

3.1 Microswimmer model

As discussed in Section 2.1.1, we model microswimmers as extended force dipoles of length l with dipolar strength κ ($\kappa > 0$ for pushers and $\kappa < 0$ for pullers). We work

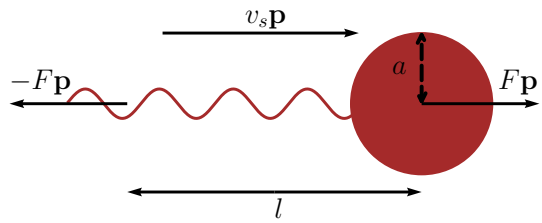


Figure 3.1: Schematic representation of the pusher model, where F denotes the applied force, l the swimmer length, v_s the swimming speed, a the effective body radius, and $\mathbf{p}$ the swimmer's orientation.

in the needle-like (infinitely slender) limit of the Bretherton parameter, $B \to 1$, where the swimmer position $\mathbf{r}$ and orientation $\mathbf{p}$ evolve according to Eqs. (2.2)–(3.2):

$$\dot{\mathbf{r}} = v_s\,\mathbf{p} + \mathbf{u}(\mathbf{r})\,. \tag{3.1}$$

$$\dot{\mathbf{p}} = (\mathbb{1} - \mathbf{pp}) \cdot \nabla\mathbf{u} \cdot \mathbf{p}. \tag{3.2}$$

Although hydrodynamics treats swimmers as needle-like dipoles, for reporting volume fraction it is convenient to define an *effective hydrodynamic radius* $\tilde{a}$ via a force-drag balance at speed v_s. Treating the body centre as a sphere of radius a, we write

$$F\,\mathbf{p} \;=\; 6\pi\mu\,a\,(v_s\,\mathbf{p} - \mathbf{u}_{\text{flag}})\,, \tag{3.3}$$

where $\mathbf{u}_{\text{flag}}$ is the Stokeslet induced at the body centre by the flagellar force $-F\,\mathbf{p}$ [93]. Using the $3D$ Oseen tensor,

$$\mathbf{u}_{\text{flag}} \;=\; \frac{1}{8\pi\mu l}\big(\mathbb{1} + \mathbf{pp}\big) \cdot (-F\,\mathbf{p}) \;=\; -\,\frac{F}{4\pi\mu l}\,\mathbf{p}. \tag{3.4}$$

Substituting Eq. (3.4) into Eq. (3.3) yields

$$F \;=\; 6\pi\mu\,\tilde{a}\,v_s, \qquad \tilde{a} \;=\; \frac{a}{1 - \frac{3}{2}\frac{a}{l}} \quad \text{for } \frac{a}{l} \ll 1 \text{ and } l > \tfrac{3}{2}a, \tag{3.5}$$

and using $F = |\kappa|\,\mu/l$ we find

$$a \;=\; \frac{2\,|\kappa|\,l}{12\pi v_s l^2 + 3\,|\kappa|}. \tag{3.6}$$

In $3D$, Eqs. (3.5)–(3.6) provide a practical mapping from stresslet parameters to an effective size for defining a volume fraction that can be compared to experiments. However, for $2D$, we only report number densities.

3.2 The lattice Boltzmann method

3.2.1 From Boltzmann equation to Navier-Stokes equation

We first establish the link between the Boltzmann equation and the Navier-Stokes equation in the simple case of an isothermal three-dimensional gas before delving into the implementation of the LB method.

Let $f(\mathbf{r}, \mathbf{v}, t)$ be the single-particle distribution function of a single fluid particle at time t. The Boltzmann equation [107] with an external body-force density $\mathbf{F}/\rho$ reads

$$\partial_t f + \mathbf{v}\cdot\nabla f + \frac{\mathbf{F}}{\rho}\cdot\nabla_{\mathbf{v}} f \;=\; \Omega[f]. \tag{3.7}$$

Here ∇ is the spatial gradient operator, $\nabla_{\mathbf{v}}$ is the velocity-space gradient, ρ is the mass density, and $\Omega[f]$ is the collision operator. The macroscopic fields are velocity moments of f:

$$\rho = \int f\, d\mathbf{v}, \qquad \rho\mathbf{u} = \int \mathbf{v}\, f\, d\mathbf{v}, \qquad \mathbf{P} = \int \mathbf{v}\mathbf{v}\, f\, d\mathbf{v}, \tag{3.8}$$

where $\mathbf{u}$ is fluid velocity, and $\mathbf{P}$ is the momentum-flux tensor. For convenience, we introduce the *peculiar velocity* $\mathbf{c} = \mathbf{v} - \mathbf{u}$ and the decomposition

$$\mathbf{P} = \rho\,\mathbf{u}\mathbf{u} + p\,\mathbb{1} + \mathbf{\Pi}, \tag{3.9}$$

with p the pressure and $\mathbf{\Pi}$ the non-equilibrium (viscous) stress.

We use the single-relaxation-time BGK (Bhatnagar-Gross-Krook) collision operator [108], which drives f toward a local equilibrium f^{eq} on a timescale τ:

$$\Omega[f] = -\frac{1}{\tau}\big(f - f^{\mathrm{eq}}\big), \tag{3.10}$$

where f^{eq} is the isothermal Maxwell-Boltzmann distribution given by [109]

$$f^{\mathrm{eq}}(\mathbf{r}, \mathbf{v}, t) = \frac{\rho}{(2\pi R T_0)^{3/2}} \exp\left[-\frac{(\mathbf{v} - \mathbf{u})^2}{2RT_0}\right], \tag{3.11}$$

with R the specific gas constant. The collision operator conserves mass and momentum:

$$\int \Omega[f] d\mathbf{v} = 0, \quad \int \Omega[f] \mathbf{v} d\mathbf{v} = 0. \tag{3.12}$$

The first three moments of the equilibrium distribution f^{eq} are given by:

$$\int f^{\mathrm{eq}} d\mathbf{v} = \rho, \quad \int \mathbf{v}\, f^{\mathrm{eq}} d\mathbf{v} = \rho \mathbf{u}, \quad \int \mathbf{c}\mathbf{c}\, f^{\mathrm{eq}} d\mathbf{v} = p\, \mathbb{1}, \tag{3.13}$$

where $p = \rho R T_0 = \rho c_s^2$ with c_s the speed of sound.

Using the BGK collision operator, the Boltzmann equation takes the form

$$\partial_t f + \mathbf{v} \cdot \nabla f + \frac{\mathbf{F}}{\rho} \cdot \nabla_{\mathbf{v}} f = -\frac{1}{\tau}\big(f - f^{\mathrm{eq}}\big). \tag{3.14}$$

Taking the zeroth and first velocity moments of Eq. (3.14) and using Eqs. (3.9)–(3.12), we obtain the mass and momentum conservation equations:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0. \tag{3.15}$$

$$\rho\big(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}\big) = -\nabla p - \nabla \cdot \mathbf{\Pi} + \mathbf{F}. \tag{3.16}$$

At this point $\mathbf{\Pi}$ is unknown and finding it yields the Navier-Stokes equation. By performing a so-called Chapman-Enskog expansion about local equilibrium in the weakly compressible hydrodynamic limit, we find:

$$\mathbf{\Pi} = -\boldsymbol{\sigma} = -\mu\Big(\nabla \mathbf{u} + \nabla \mathbf{u}^{\top} - \tfrac{2}{3}(\nabla \cdot \mathbf{u})\, \mathbb{1}\Big), \tag{3.17}$$

where $\boldsymbol{\sigma}$ is the viscous (deviatoric) stress, and $\mu = p\tau = \rho c_s^2 \tau$ the viscosity for the BGK operator. Finally, by substituting Eq. (3.17) into Eq. (3.16) and assuming a constant viscosity μ and vanishing bulk viscosity, we obtain the Navier-Stokes equation for a compressible fluid:

$$\rho\left(\partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u}\right) = -\nabla p + \mu\,\nabla^2\mathbf{u} + \tfrac{\mu}{3}\,\nabla(\nabla\cdot\mathbf{u}) + \mathbf{F}. \tag{3.18}$$

After establishing the link between the Boltzmann equation and both the continuity equation and the Navier-Stokes equation, we will show how the LB method is developed starting from the Boltzmann equation.

3.2.2 Lattice Boltzmann method

As indicated by its name, the LB method evolves the phase-space distribution $f(\mathbf{r},\mathbf{v},t)$ on a Cartesian lattice in physical space using a finite set of discrete velocities. The Discretisation must preserve the low-order moments so that the Navier-Stokes equations are recovered. A classical route is to expand the phase-space distribution f in Hermite polynomials [109]:

$$f(\mathbf{r},\mathbf{v},t) = \omega(\mathbf{v}) \sum_{n=0}^{\infty} \frac{1}{n!}\, \mathbf{a}^{(n)}(\mathbf{r},t) : \mathbf{H}^{(n)}\Big(\frac{\mathbf{v}}{c_s}\Big), \tag{3.19}$$

$$\mathbf{a}^{(n)}(\mathbf{r},t) \equiv \int f(\mathbf{r},\mathbf{v},t)\; \mathbf{H}^{(n)}\Big(\frac{\mathbf{v}}{c_s}\Big)\, d\mathbf{v}, \tag{3.20}$$

where $\mathbf{H}^{(n)}$ is the d-dimensional Hermite polynomial of order n, $\mathbf{a}^{(n)}$ the corresponding coefficient, and $\omega(\mathbf{v})$ the weight function defined as:

$$\omega(\mathbf{v}) = \frac{1}{(2\pi c_s^2)^{d/2}}\; \exp\big[-|\mathbf{v}|^2/(2c_s^2)\big]. \tag{3.21}$$

Two features are crucial here. First, the tensor Hermite polynomials are orthogonal with respect to the Gaussian weight ω:

$$\int_{\mathbb{R}^d} \omega(\mathbf{v})\, \mathbf{H}^{(n)}\Big(\frac{\mathbf{v}}{c_s}\Big) : \mathbf{H}^{(m)}\Big(\frac{\mathbf{v}}{c_s}\Big)\, d\mathbf{v} = n!\,\delta_{nm}.$$

Second, the lowest coefficients coincide with the conserved hydrodynamic moments:

$$a^{(0)} = \rho, \tag{3.22}$$

$$\mathbf{a}^{(1)} = \rho\mathbf{u}. \tag{3.23}$$

Thus, the orthogonality of the Hermite polynomial ensures that density ρ and fluid velocity $\mathbf{u}$ are not affected by the truncation of higher-order terms. Retaining terms up to second order in the flow velocity $\mathbf{u}/c_s$, the equilibrium distribution reads

$$f^{\mathrm{eq}}(\mathbf{r}, \mathbf{v}, t) \approx \rho\, \omega(\mathbf{v}) \left[1 + \frac{\mathbf{v}\cdot\mathbf{u}}{c_s^2} + \frac{(\mathbf{v}\cdot\mathbf{u})^2}{2c_s^4} - \frac{\mathbf{u}\cdot\mathbf{u}}{2c_s^2}\right]. \tag{3.24}$$

Conveniently, Eq. (3.24) is the product of a Gaussian $\omega(\mathbf{v})$ and a polynomial in $\mathbf{v}$. This allows to reduce the continuous velocity dependence to a *finite* set of representative velocities $\{\mathbf{c}_i\}$ and weights $\{w_i\}$ by means of a Gauss-Hermite quadrature such that the set $\{(\mathbf{c}_i, w_i)\}$ obey the following identities which ensure rotational isotropy

$$\sum_i w_i \;=\; 1, \qquad \sum_i w_i\, \mathbf{c}_i \;=\; \mathbf{0}, \tag{3.25}$$

$$\sum_i w_i\, c_{i\alpha} c_{i\beta} \;=\; c_s^2 \delta_{\alpha\beta}, \qquad \sum_i w_i\, c_{i\alpha} c_{i\beta} c_{i\gamma} \;=\; 0, \tag{3.26}$$

$$\sum_i w_i\, c_{i\alpha} c_{i\beta} c_{i\gamma} c_{i\delta} \;=\; c_s^4 \big(\delta_{\alpha\beta}\delta_{\gamma\delta} + \delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}\big), \tag{3.27}$$

$$\sum_i w_i\, c_{i\alpha} c_{i\beta} c_{i\gamma} c_{i\delta} c_{i\eta} \;=\; 0, \tag{3.28}$$

where Greek indices denote Cartesian coordinates. Satisfying these moment constraints up to fourth order ensures recovery of the isothermal Navier-Stokes equations. The sets $\{(\mathbf{c}_i, w_i)\}$ are of crucial importance since they define the nature of the lattice used for the LB method. After discretising the velocity space, the discrete equilibrium distribution reads [109]

$$f_i^{\mathrm{eq}} = \rho\, w_i \left[1 + \frac{\mathbf{c}_i\cdot\mathbf{u}}{c_s^2} + \frac{(\mathbf{c}_i\cdot\mathbf{u})^2}{2c_s^4} - \frac{\mathbf{u}\cdot\mathbf{u}}{2c_s^2}\right], \tag{3.29}$$

and the viscosity is $\mu = \rho c_s^2\big(\tau - \frac{\Delta t}{2}\big)$, which requires $\tau > \frac{\Delta t}{2}$. A similar treatment applies to the out-of-equilibrium distribution f.

3.2.3 Discretisation of the forcing term

Accurately incorporating body forces in the lattice Boltzmann (LB) method is essential: if forcing is treated inconsistently, the recovered hydrodynamics can violate exact momentum conservation. Several forcing formulations exist [110, 111], among these, the scheme of Guo *et al.* [112] is the most commonly used because it achieves second-order accuracy in time, preserves the correct low-order moments, and recovers the isothermal Navier-Stokes equations for finite τ and Δt. In what follows we adopt the Guo formulation, which yields a simple discrete forcing term S_i satisfying

$$S_i(\mathbf{r},t) = \left(1 - \frac{\Delta t}{2\tau}\right) w_i \left(\frac{\mathbf{c}_i - \mathbf{u}}{c_s^2} + \frac{(\mathbf{c}_i \cdot \mathbf{u})\mathbf{c}_i}{c_s^4} \right) \cdot \mathbf{F}(\mathbf{r},t), \tag{3.30}$$

where Δt is the discretised time [112]. The moments can be obtained using the discretised distribution f_i:

$$\rho = \sum_i f_i \tag{3.31}$$

$$\mathbf{u} = \frac{1}{\rho} \sum_i \mathbf{c}_i f_i + \frac{\Delta t}{2\rho} \mathbf{F}. \tag{3.32}$$

After discretising the velocity space and spacetime, and, consequently, the phase-space distribution f, the collision operator $\Omega(f)$, and the forcing term $(\mathbf{F}/\rho) \cdot \nabla_{\mathbf{v}} f$, we obtain the discrete-velocity Boltzmann equation:

$$\partial_t f_i + \mathbf{c}_i \cdot \nabla f_i \;=\; \Omega_i(f) + S_i. \tag{3.33}$$

By inserting the BGK collision operator and the discrete forcing term S_i in Eq. (3.33), we obtain the lattice Boltzmann equation [109, 113]:

$$\tilde{f}_i(\mathbf{r} + \mathbf{c}_i \Delta t,\; t + \Delta t) = \tilde{f}_i(\mathbf{r},t) - \frac{\Delta t}{\tau} \Big[\tilde{f}_i(\mathbf{r},t) - f_i^{\mathrm{eq}}(\rho,\mathbf{u}) \Big] + \Delta t\, S_i(\mathbf{r},t), \tag{3.34}$$

where $\tilde{f}$ is defined as

$$\tilde{f}_i(\mathbf{r},t) \equiv f_i(\mathbf{r},t) - \frac{\Delta t}{2} S_i(\mathbf{r},t). \tag{3.35}$$

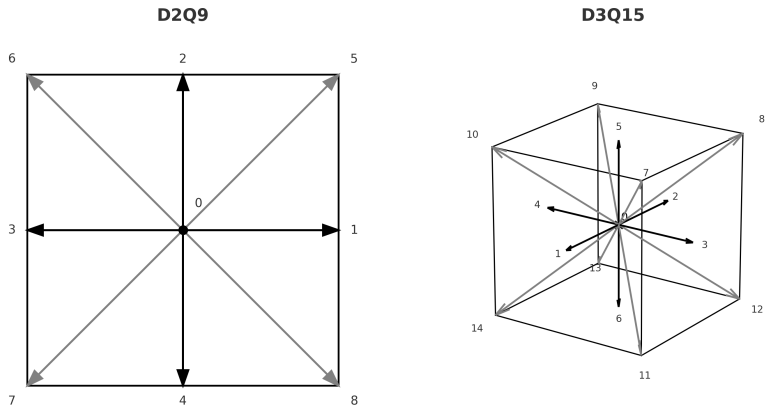


Figure 3.2: Schematic representation of the discrete velocity sets used in LB. Left: $D2Q9$ lattice with nine discrete velocities in two dimensions, including the rest particle (0), four axial directions (1-4), and four diagonal directions (5-8). Right: $D3Q15$ lattice with fifteen discrete velocities in three dimensions, including the rest particle (0), six axial directions (1-6), and eight body-diagonal directions (7-14). Arrows indicate the discrete velocity vectors $\mathbf{c}_i$ in lattice units.

Eq. (3.34) can be seen as a two-step process where the terms on the right-hand side correspond to a collision step where the distribution f relaxes towards f^{eq}, while the left-hand side term represent a streaming step where f propagates to the neighbouring lattice points defined by the set $\{(\mathbf{c}_i, w_i)\}$.

3.2.4 Examples of LB lattices

In our work, we have used both $2D$ ($D2Q9$) and $3D$ ($D3Q15$) implementations of the LB method:

$D3Q15 \quad (d = 3)$. In the $D3Q15$ lattice Boltzmann model, we fix the time step $\Delta t = 1$ and the lattice spacing to $\Delta H = 1$ such that $c_s^2 = 1/3$. The velocity space is discretised into fifteen vectors $\{\mathbf{c}_i\}_{i=0}^{14}$: a rest particle $\mathbf{c}_0 = (0, 0, 0)$, six axis-aligned directions $(\pm 1, 0, 0)$, $(0, \pm 1, 0)$, $(0, 0, \pm 1)$, and eight body diagonals $(\pm 1, \pm 1, \pm 1)$. Each vector is assigned a weight w_i to ensure isotropy and correct recovery of hydrodynamic moments: $w_0 = 2/9$ for the rest particle, $w_{1-6} = 1/9$ for the axis directions, and $w_{7-14} = 1/72$ for the diagonals.

$D2Q9 \quad (d = 2)$. In the $D2Q9$ lattice Boltzmann model, we fix the time step $\Delta t = 1$ and the lattice spacing to $\Delta H = 1$ such that $c_s^2 = 1/3$. The velocity space is discretised into nine vectors $\{\mathbf{c}_i\}_{i=0}^{8}$: a rest particle $\mathbf{c}_0 = (0, 0)$, four axis-aligned

directions $(\pm 1, 0)$ and $(0, \pm 1)$, and four diagonals $(\pm 1, \pm 1)$. The corresponding weights are $w_0 = 4/9$ for the rest particle, $w_{1-4} = 1/9$ for the axis directions, and $w_{5-8} = 1/36$ for the diagonals, ensuring isotropy and correct recovery of hydrodynamic moments.

3.3 Implementation of point forces

For swimmers located at positions $\mathbf{r}_s$ and exerting forces $\mathbf{F}_s$, the continuum body force may be expressed as a superposition of Dirac delta functions:

$$\mathbf{F}(\mathbf{r}) = \sum_s \mathbf{F}_s \delta\left(\mathbf{r} - \mathbf{r}_s\right). \tag{3.36}$$

Since the swimmer positions $\mathbf{r}_s$ are generally not constrained to lattice nodes, the forces they exert on the fluid correspond to off-lattice singularities. In the conventional LB formulation, however, the force density is defined only at lattice nodes. It is therefore necessary to introduce a regularisation procedure that maps each off-lattice force onto the neighbouring nodes in a manner consistent with the LB discretisation [113].

For a cubic lattice with spacing ΔH, the Dirac delta function is replaced by a smooth kernel δ^P, which possesses compact support of size 2 [114]:

$$\delta^P(\mathbf{r} - \mathbf{r}_s) = \frac{1}{\Delta H^3} w\left(\frac{x - x_s}{\Delta H}\right) w\left(\frac{y - y_s}{\Delta H}\right) w\left(\frac{z - z_s}{\Delta H}\right), \tag{3.37}$$

where the weighting function $w(|\mathbf{r}|)$ is defined as

$$w(|\mathbf{r}|) = \begin{cases} \dfrac{3 - 2|\mathbf{r}| + \sqrt{1 + 4|\mathbf{r}| - 4|\mathbf{r}|^2}}{8}, & |\mathbf{r}| \leq 1, \\ \dfrac{5 - 2|\mathbf{r}| - \sqrt{-7 + 12|\mathbf{r}| - 4|\mathbf{r}|^2}}{8}, & 1 < |\mathbf{r}| \leq 2, \\ 0, & |\mathbf{r}| > 2. \end{cases} \tag{3.38}$$

In $2D$, the kernel δ^P becomes:

$$\delta^P(\mathbf{r} - \mathbf{r}_s) = \frac{1}{\Delta H^2} w\left(\frac{x - x_s}{\Delta H}\right) w\left(\frac{y - y_s}{\Delta H}\right). \tag{3.39}$$

The regularised, nodewise force density entering the forced LB equation at a lattice node $\mathbf{r}$ is then given by the kernel-weighted superposition of the surrounding swimmer forces:

$$\mathbf{F}(\mathbf{r}) = \sum_s \mathbf{F}_s \delta^P \left(\mathbf{r} - \mathbf{r}_s\right), \tag{3.40}$$

where the summation is effectively restricted to the compact support of the kernel, *i.e.*, to lattice nodes satisfying $|\mathbf{r}| < 2$ in each coordinate.

Using Eq. (3.40), a standard forced LB iteration is performed to obtain the updated fluid velocity field $\mathbf{u}(\mathbf{r})$ at each node. The fluid velocity at each off-lattice swimmer position is subsequently evaluated by interpolation from the surrounding lattice nodes using the same kernel:

$$\mathbf{u}(\mathbf{r}_s) = \sum_{\mathbf{r}} \mathbf{u}(\mathbf{r}) \delta^P \left(\mathbf{r}_s - \mathbf{r}\right). \tag{3.41}$$

The computational procedure can be summarized as follows:

- Regularise and interpolate the force singularities onto the neighbouring lattice nodes.
- Perform a lattice Boltzmann iteration to obtain the fluid velocity field $\mathbf{u}(\mathbf{r})$.
- Interpolate $\mathbf{u}(\mathbf{r})$ to the swimmer positions and update their translational and rotational dynamics according to Eqs. (2.2)–(3.2).

This scheme is computationally efficient due to the compact support of the interpolation kernel: each swimmer interacts only with a small local neighbourhood of lattice nodes, ensuring that the interpolation operations ($\mathrm{Swimmer} \rightarrow \mathrm{Node}$ and $\mathrm{Node} \rightarrow \mathrm{Swimmer}$) incur only $O(1)$ computational cost.

Chapter 4

Number Fluctuations and Nematic Order in Microswimmer Suspensions

In *dry* active matter (no momentum conservation), the coupling between quasi-long-range orientational order, either polar or nematic, and activity leads to anomalous number fluctuations [9],

$$\Delta N \sim \langle N \rangle^{\alpha}, \qquad \alpha > \tfrac{1}{2}. \tag{1}$$

These so-called giant number fluctuations (GNFs) are a generic feature in dry systems: they have been predicted theoretically in dry polar flocks and dry active nematics and observed both experimentally [12, 115] and in simulations [44, 116]. For dry systems with nematic order, linear theory yields an exponent that depends on the spatial dimensionality d,

$$\alpha = \frac{1}{2} + \frac{1}{d}. \tag{2}$$

When nonlinearities are accounted for, the exponent becomes nonuniversal and dependent on the microscopic model parameters [117].

By contrast, in *wet* (momentum-conserving) active matter, the nematically ordered state is *generically* unstable for finite wave numbers in the presence of flow fields [102]. As a result, the theoretical framework that works in dry media, linear stability analysis

of a stable ordered state followed by renormalisation of nonlinearities [42, 117], does not apply, and there are no predictions for GNFs in unbounded wet media. Nevertheless, several experiments and simulations have reported super-Gaussian number fluctuations in momentum-conserving systems, suggesting that some form of "giant fluctuations" can still emerge.

In *Paper I*, we simulate dilute three-dimensional suspensions of pusher microswimmers in an incompressible viscous fluid. We corroborate that super-Gaussian number fluctuations appear once the suspension crosses the onset of collective motion. Crucially, unlike in dry active matter, these enhancements are mesoscopic: they are confined to length scales shorter than the swimmer persistence length $\ell_p = v_s/\lambda$, beyond which Gaussian fluctuations are recovered. By quantifying orientational order in the collectively moving state, we show that the nematic length ξ (the linear size of locally nematic patches) is proportional to ℓ_p, directly linking nematic order to non-Gaussian density statistics. Taken together, these results indicate that *bona fide*, asymptotic GNFs are generically suppressed by fluid flows.

As discussed in Chapter 2, the transition to collective behaviour in unbounded suspensions of pusher microswimmers occurs for number densities n above [90, 92–94]

$$n > n_c^\infty = \frac{5\lambda}{\kappa}. \tag{4.1}$$

Here λ is the tumbling rate and κ the dipolar strength. In a finite cubic domain of size H, the most unstable mode is $k_{\min} = 2\pi/H$. This produces a finite-size shift [101],

$$n_c(H) = n_c^\infty \left(1 + \frac{3\pi}{5}\frac{\ell_p}{H} + \frac{4\pi^2}{5}\frac{\ell_p^2}{H^2}\right). \tag{4.2}$$

We simulate systems at various number densities n and persistence lengths ℓ_p in a cubic, periodic box of size H. We vary $\ell_p = v_s/\lambda$ by changing the swimming speed v_s at fixed λ, so n_c^∞ remains constant across runs.

We characterise the transition via the normalised fluid-velocity variance $\langle U^2\rangle/\langle U^2\rangle_0$ as a function of the rescaled density $\Delta = n/n_c(H)$, where $\langle U^2\rangle_0$ is the variance for non-interacting swimmers, for which fluid induced advection and rotation have been switched off. Fig. 4.1 a shows a sharp increase at the predicted onset of bacterial turbulence. The transition also sharpens with increasing ℓ_p: faster swimming suppresses pre-transitional correlations, narrowing the onset region without shifting the critical density [95]. In the turbulent regime ($\Delta = 2$), the flow fields display vortical structures whose characteristic sizes grow with ℓ_p (Fig. 4.1 b). For shakers ($\ell_p = 0$), these structures are spatially limited.

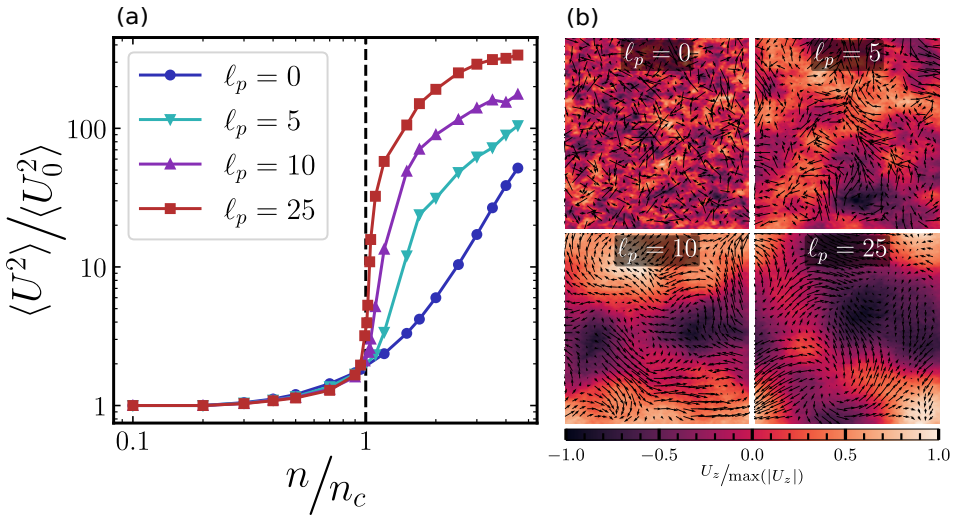


Figure 4.1: (a) Normalised fluid-velocity variance, $\langle U^2 \rangle / \langle U^2 \rangle_0$, for various persistence lengths ℓ_p. The swimmer density n is scaled by the critical density $n_c(H)$. Increasing swimming speed (larger ℓ_p) suppresses pre-transitional swimmer-swimmer correlations, sharpening the transition. (b) Snapshots of a $2d$ slice of the $3d$ flow field for $n/n_c = 2.0$ and $H = 250$, for the values of ℓ_p indicated. Arrows show the in-plane component and colours show the out-of-plane component of $\mathbf{U}$, scaled by its maximum value.

We quantify density fluctuations by sampling particle counts in randomly placed spherical windows of radius R and measuring the spatiotemporal mean $\langle N \rangle$ and standard deviation ΔN. Fig. 4.2 a shows a clear deviation from Gaussian fluctuations for $\ell_p > 0$ at the onset of bacterial turbulence: just above the instability threshold, ΔN grows faster than the baseline $\langle N \rangle^{1/2}$ over a finite range of sampling window radii. By contrast, for shakers ($\ell_p = 0$) no such deviation is observed; number fluctuations remain Gaussian even in the turbulent regime. This is consistent with preserved fore-aft symmetry and the resulting absence of local polar fluxes [46].

A non-monotonic dependence of ΔN on $\langle N \rangle$ is also observed for $\ell_p > 0$. While global particle conservation enforces a downturn at very large sampling windows, conservation alone does not explain the initial rise and subsequent return towards Gaussian scaling. To analyse this behaviour, we fit the $(\langle N \rangle, \Delta N)$ curves to high-order polynomial, and define the *apparent* exponent

$$\alpha_{\mathrm{app}}(R) \equiv \frac{\mathrm{d}\,\log \Delta N}{\mathrm{d}\,\log \langle N \rangle}, \tag{4.3}$$

which is the local slope of the log-log curve and thus an effective scaling exponent at sampling size R.

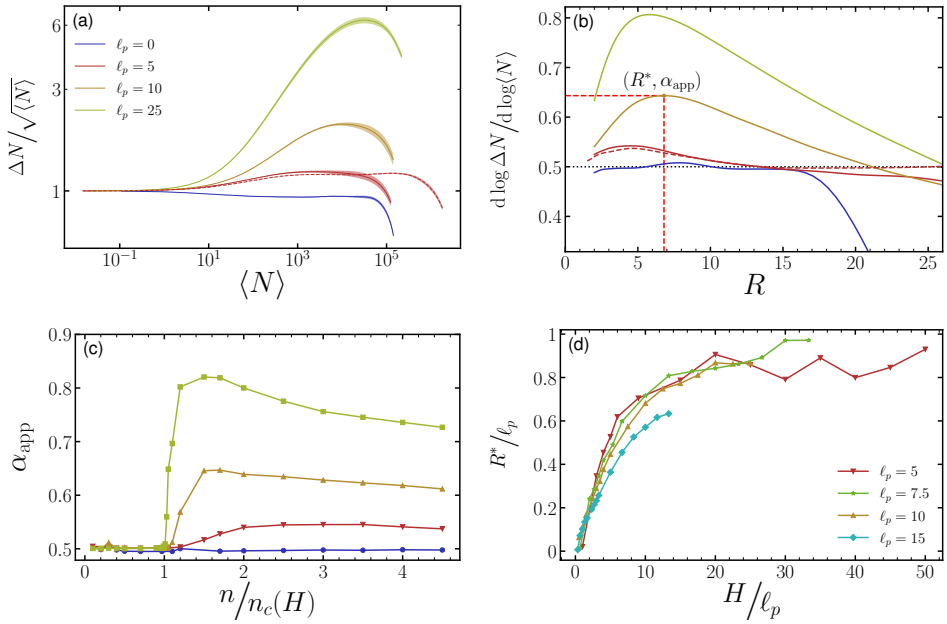


Figure 4.2: (a) Reduced standard deviation $\Delta N/\sqrt{\langle N\rangle}$ of the particle number N in spherical subvolumes versus the mean occupancy $\langle N\rangle$ at fixed $n/n_c = 2.0$ for several persistence lengths ℓ_p. The dashed curve for $\ell_p = 5$ corresponds to a larger system, $H = 250$, and shows an extended plateau of Gaussian fluctuations at intermediate scales. Shaded bands denote one standard deviation, obtained by averaging each run over ten subbatches. (b) Scale-dependent slope $\mathrm{d}\log(\Delta N)/\mathrm{d}\log\langle N\rangle$ as a function of sampling radius R. Its maximum defines the apparent GNF exponent $\alpha_{\rm app}$, and the radius where it occurs defines R^*. (c) $\alpha_{\rm app}$ versus reduced density n/n_c. Non-Gaussian fluctuations emerge at the onset of collective motion only for sufficiently large ℓ_p and are absent for shakers. (d) R^*/ℓ_p versus reduced system size H/ℓ_p for $n/n_c = 2.0$. The data collapse indicates that for $H \gg \ell_p$, R^* is controlled by ℓ_p; hence, the apparently non-Gaussian fluctuations are transient.

In Fig. 4.2 b and c, we observe that, for shakers, $\alpha_{\rm app}(R) \approx 1/2$ for all R, as expected. For $\ell_p > 0$, $\alpha_{\rm app}(R)$ develops a pronounced peak just above onset, exceeds $1/2$ over a bounded interval in R, and relaxes back towards $1/2$ at larger R. The height of this peak,

$$\alpha_{\rm app}^{\rm max} = \max_R \alpha_{\rm app}(R), \tag{4.4}$$

is **not universal**: it depends systematically on ℓ_p, the reduced density n/n_c, and the domain size H. The enhancement is strongest near onset and increases with ℓ_p, yet no single, system size independent exponent emerges.

We use the peak location to identify the sampling radius of strongest anomalous fluctuations,

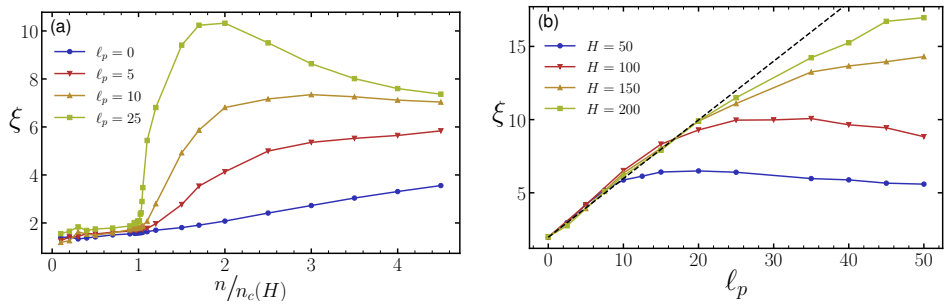


Figure 4.3: (a) Fitted nematic correlation length ξ versus reduced swimmer density $n/n_c(H)$ for $H = 100$ and several persistence lengths ℓ_p. (b) ξ as a function of ℓ_p at $n/n_c = 2.0$ for four system sizes H. The dashed line is a linear fit to the first six data points for $H = 200$.

$$R^\star \equiv \arg\max_R \, \alpha_{\rm app}(R). \tag{4.5}$$

Fig. 1 d illustrates two robust emergent features. First, $R^\star$ grows with H only up to a point and then saturates once the domain is sufficiently large,

$$R^\star(H) \to R^\star \quad \text{for} \quad H \gtrsim 20\,\ell_p. \tag{4.6}$$

Second, $R^\star$ tracks the persistence length with an approximately constant prefactor,

$$R^\star \approx c_\star\,\ell_p, \qquad c_\star \simeq 0.8, \tag{4.7}$$

over the parameter ranges explored. Thus the super-Gaussian fluctuations regions are persistence-length limited: strongest on mesoscopic windows set by the swimmers' persistence length ℓ_p and fading outside that window.

To investigate the origin of the observed anomalous fluctuations, we measure the distance-dependent nematic order parameter

$$S(r) \equiv \tfrac{1}{2}\big\langle 3\cos^2\theta - 1\big\rangle_r, \tag{4.8}$$

whose initial decay is fitted to a stretched exponential $S(r) = S_0 \exp\big(-(r/\xi)^\beta\big)$, where ξ is the characteristic nematic patch size. We find that ξ increases sharply at onset (for $\ell_p > 0$) and then grows approximately linearly with ℓ_p at fixed reduced density (Fig. 4.3 a and b), implying

$$\xi \sim \ell_p, \qquad R^\star \sim \xi. \tag{4.9}$$

Physically, these *pseudo-giant number fluctuations* arise within locally ordered nematic patches of size ξ as a result of the broken fore-aft symmetry ($\ell_p > 0$) leading to the emergence of local polar order that drive density fluctuations; beyond ξ, central-limit behaviour is restored so that $\alpha_{\rm app}(R) \to 1/2$. The eventual downturn of ΔN at the largest windows is enforced by global conservation of the number of particles, but the location and magnitude of the peak are controlled by the persistence-limited nematic patch size.

In summary, the $\alpha_{\rm app}(R)$ analysis shows that number fluctuations in wet, momentum-conserving suspensions are large but intrinsically **mesoscopic**: they peak at $R^\star \sim \ell_p$ within locally nematically aligned domains and relax to Gaussian statistics beyond that scale, thereby ruling out asymptotic, system-size independent GNFs in this setting.

A kinetic description written in terms of the number density c, the polarisation m_i, and the nematic tensor Q_{ij} (with i and j denoting the Cartesian components and the summation convention over repeated indices implied), coupled to the Stokes equations [118], admits a natural nondimensionalisation with time λ^{-1}, length $\ell_p = v_s/\lambda$, and velocity v_s. The (nondimensional) moment and Stokes equations used are:

$$(\partial_t + U_\alpha \nabla_\alpha)\, c = -\nabla_\alpha m_\alpha, \tag{4.10}$$

$$\begin{aligned}(\partial_t + U_\alpha \nabla_\alpha)\, m_i = &- m_i - \nabla_\alpha\big(Q_{i\alpha} + \tfrac{1}{3}\delta_{i\alpha} c\big) + \tfrac{4}{5}\, m_\alpha \nabla_\alpha U_i \\ &- \tfrac{1}{5}\, m_\alpha \nabla_i U_\alpha,\end{aligned} \tag{4.11}$$

$$\begin{aligned}(\partial_t + U_\alpha \nabla_\alpha)\, Q_{ij} = &- Q_{ij} - \tfrac{1}{5}\big(\nabla_j m_i + \nabla_i m_j - \tfrac{2}{3}\delta_{ij}\nabla_\alpha m_\alpha\big) \\ &+ \tfrac{1}{5}\, c\,(\nabla_j U_i + \nabla_i U_j) + \tfrac{1}{7}\nabla^2 Q_{ij} \\ &+ \tfrac{3}{35}\big(\nabla_i \nabla_\alpha Q_{\alpha j} + \nabla_j \nabla_\alpha Q_{\alpha i} - \tfrac{2}{3}\delta_{ij}\nabla_\alpha \nabla_\beta Q_{\alpha\beta}\big) \\ &- \tfrac{2}{7}\delta_{ij} Q_{\alpha\beta}\nabla_\alpha U_\beta + \tfrac{5}{7}(Q_{i\alpha}\nabla_\alpha U_j + Q_{j\alpha}\nabla_\alpha U_i) \\ &- \tfrac{2}{7}(Q_{i\alpha}\nabla_j U_\alpha + Q_{j\alpha}\nabla_i U_\alpha)\,,\end{aligned} \tag{4.12}$$

$$\nabla_i\big(p + \tfrac{1}{3}c\big) = \nabla^2 U_i - 5\,\frac{n}{n_c}\,\nabla_\beta Q_{i\beta}, \qquad \nabla_i U_i = 0. \tag{C4–C5}$$

In these units, the dynamics depend *only* on one parameter, namely the reduced density n/n_c. Hence, any emergent (non-dimensional) length scale $\tilde{\ell}$ should take the form

$$\tilde{\ell} = f(n/n_c), \tag{4.13}$$

where f in an unknown function. As a consequence, the corresponding dimensional length scale ℓ

$$\ell = \ell_p f(n/n_c), \tag{4.14}$$

this means that ℓ_p sets the scale of *all* emergent properties of the turbulent state, such as the size of the nematic patches ξ. Consistent with this expectation, our results show

$$\ell_p \sim R^\star \sim \xi. \tag{4.15}$$

Chapter 5

The role of dimensionality

Suspensions of rear-actuated pusher microswimmers display hydrodynamic instabilities that, beyond a threshold density, lead to bacterial turbulence, an archetypical realisation of emergent behaviour in wet active matter [46, 47, 119–121]. Despite fundamental differences between two- and three-dimensional Stokes flows (see Chapter 1), mean-field theory predicts closely related onset criteria for collective motion in unbounded systems (Eq. (2.13)):

$$n_c = \frac{(d+2)\lambda}{B\kappa}, \tag{5.1}$$

where $d \in \{2, 3\}$ is the dimension of the system [45, 90, 93, 94, 102]. Because mean-field approaches neglect correlations and noise, in *Paper II*, we revisit the role of dimensionality using large-scale, particle-resolved simulations in $2D$ that retain nonlinear hydrodynamics.

We consider a $2D$ suspension of pusher microswimmers of number density n in an incompressible viscous fluid of viscosity μ. Each swimmer is approximated as an extended force dipole generating the $2D$ far-field flow as introduced in Eq. (1.11)

$$\mathbf{u}(\mathbf{r}) = \frac{\kappa}{4\pi}\left(\frac{2\,(\mathbf{p}\cdot\mathbf{r})^2}{r^4} - \frac{1}{r^2}\right)\mathbf{r}, \tag{5.2}$$

which has four-fold in-plane symmetry and $1/r$ decay [94]. Swimmers dynamics are dictated by the same equations of motion Eqs (2.2)–(3.2) as their $3D$ counterpart, while the hydrodynamics are solved with a $D2Q9$ lattice Boltzmann implementation, coupling off-lattice point forces to the LB grid as discussed in Chapter 3.

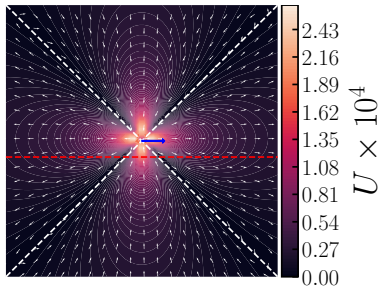


Figure 5.1: Flow field around a single pusher in $2D$ obtained using the LB algorithm. The red cut is used for the profiles illustrated in Fig. 5.2, while the white dashed lines illustrate regions where the velocity changes sign.

The simulated velocity field around a stationary pusher reproduces Eq. (1.11) without fitting parameters, including the prescribed κ. The far-field deviations, observed in Fig. 5.2 a, reflect periodic boundary conditions, more visible in $2D$ due to the slower $1/r$ decay, in contrast with $1/r^2$ in $3D$ (Fig. 5.2 b).

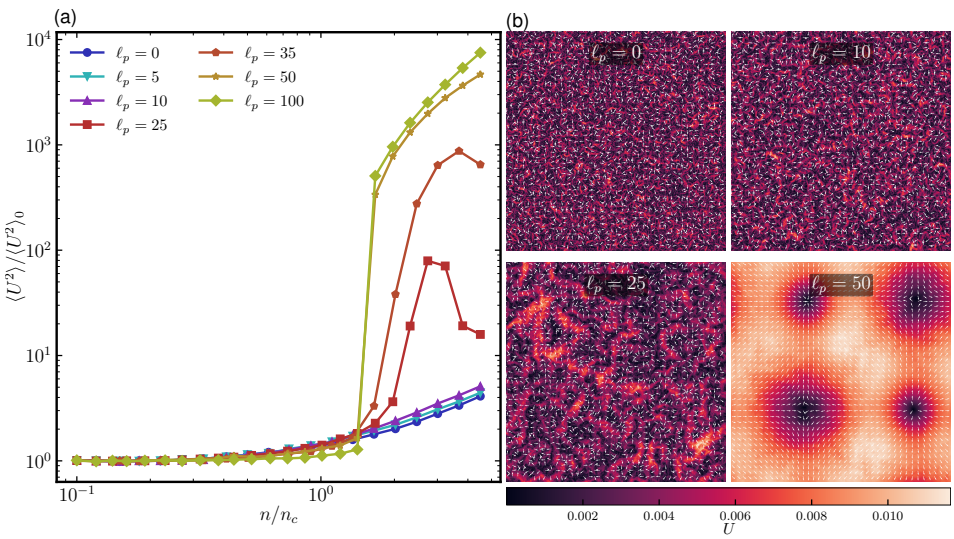


Figure 5.3: (a) Fluid velocity variance $\langle U^2 \rangle$, normalized by its corresponding value in a non-interacting suspension, $\langle U^2 \rangle_0$, as a function of the rescaled density n/n_c shown for various swimmer persistence lengths ℓ_p. (b) Snapshots of the steady-state flow field for $n/n_c = 4.5$ and four different values of ℓ_p, as indicated. The structures observed for $\ell_p = 25$ likely represent transient precursors to the system-spanning state seen at $\ell_p = 50$, which has not yet been reached due to the very slow relaxation dynamics.

We characterise the transition using the fluid-velocity variance $\langle U^2 \rangle / \langle U^2 \rangle_0$ as an order parameter and examine its dependence on the reduced density n/n_c for a range of persistence lengths $\ell_p = v_s/\lambda$, obtained by varying the swimming speed v_s at fixed tumbling rate λ. The density is normalised by the unbounded-system critical

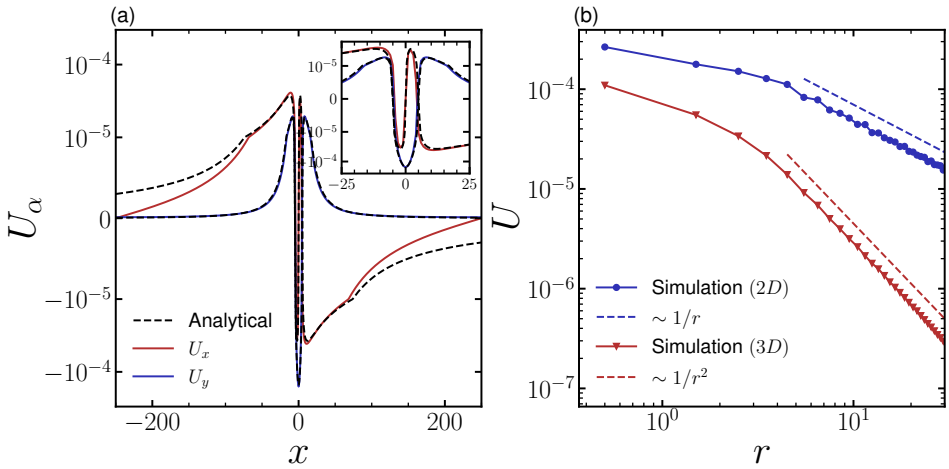


Figure 5.2: (a) Cartesian components U_α of the flow field produced by a stationary pusher dipole, evaluated in the x-direction along the cut illustrated in Fig. 5.1. The red and blue curve show the flow profile obtained from LB simulations, while the black dashed line indicates the theoretical prediction from Eq.(1.11) for the same value of κ as in the simulations. (b) The radial dependence of the velocity magnitude $U = |\mathbf{U}|$, demonstrating the expected far-field $1/r$ decay in $2D$ (blue line), in contrast to the $1/r^2$ behaviour characteristic of $3D$ flow (red line).

value in Eq. (2.13); since $H \gg \ell_p$ throughout, finite-size corrections to n_c are negligible, in contrast to $3D$, where large ℓ_p produces non-negligible shifts in the critical density [101, 106]. As ℓ_p increases, the order-parameter curve sharpens at $n = n_c$. Crucially, this sharpening is not continuous as in $3D$ [95, 106]; rather, the system exhibits a qualitative change from a gradual increase at small ℓ_p to a discontinuous jump for $\ell_p \geq 50$. Steady-state snapshots (Fig. 5.3b) further elucidate this change: at low ℓ_p the flow is disordered with only short-range correlations, whereas for $\ell_p = 50$ the system relaxes slowly from a homogeneous, isotropic initial condition to a stationary, system-spanning pattern, passing through transient configurations akin to those observed at $\ell_p = 25$. The configuration in Fig. 5.3 b for $\ell_p = 50$ resembles noiseless field-theory solutions [118, 122]: two large dipolar regions with extensile-like flows, consistent with nematically ordered pusher domains, coexist with smaller source- and sink-like regions located at the defects separating these domains.

Building on the order parameter analysis above, we now characterise the emergent temporal and spatial scales of the velocity field $\mathbf{U}(\mathbf{r})$ using the two-point correlation function $C(\mathbf{r}, t)$, decomposed into temporal and equal-time spatial components. Figure 5.4 a shows the temporal autocorrelation

$$C(t) \equiv \big\langle \mathbf{U}(\mathbf{r}_0, t_0) \cdot \mathbf{U}(\mathbf{r}_0, t_0 + t) \big\rangle$$

at $\ell_p = 50$, where angle brackets denote averaging over all positions $\mathbf{r}_0$ and time

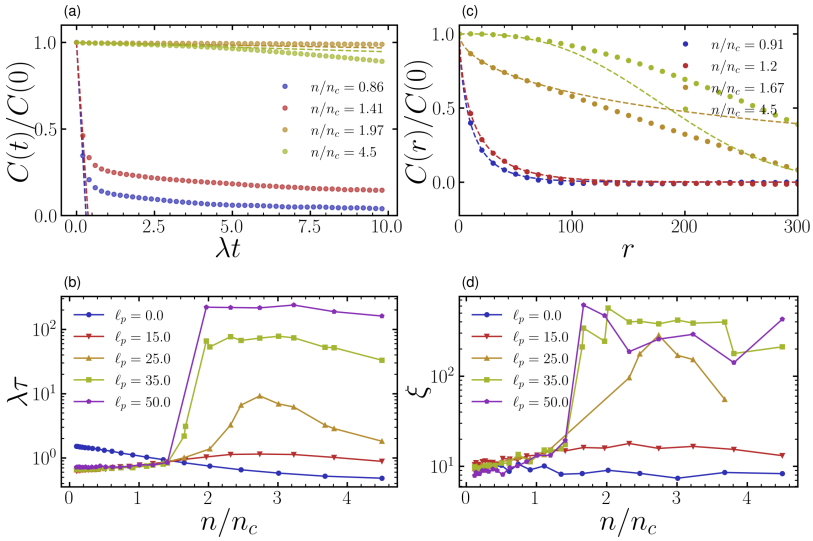


Figure 5.4: (a) Normalized temporal correlation function of the fluid velocity, $C(t)/C(0)$, shown for a persistence length $\ell_p = 50$ and several values of the rescaled density n/n_c. The dashed lines represent linear fits to the initial decay, and their intersection with the time axis provides an estimate of the characteristic relaxation time τ. (b) The relaxation time τ, obtained from panel (a), plotted as a function of the rescaled density n/n_c for different persistence lengths ℓ_p. (c) Normalized spatial correlation function of the fluid velocity, $C(r)/C(0)$, shown for a persistence length $\ell_p = 50$ and several values of the rescaled density n/n_c. The dashed lines correspond to fits of the initial decay using stretched exponential functions $C(r) \sim \exp[-(r/\xi)^\alpha]$ with $\alpha \geq 1$, from which the characteristic length scale ξ is determined. (d) The characteristic length ξ, extracted from panel (c), plotted as a function of the rescaled density n/n_c for different persistence lengths ℓ_p.

origins t_0 after the system has reached its steady state. The curves at $n/n_c = 1.97$ and 4.5 illustrate the stationary character of the high-density steady structures ($n \gg n_c$). In Fig. 5.4 b, we report the correlation time τ, defined operationally as the time t at which a linear fit to the initial decay extrapolates to zero. The abrupt increase of τ with n signals a discontinuous change from a rapidly relaxing flow to one that remains effectively stationary over very long times, consistent with the jump observed in Fig. 5.3 b.

The spatial counterpart is presented in Fig. 5.4 c, where we examine the equal-time correlation

$$C(r) \equiv \big\langle \mathbf{U}(\mathbf{r}_0, t_0) \cdot \mathbf{U}(\mathbf{r}_0 + \mathbf{r}, t_0) \big\rangle,$$

with averaging over all pairs $(\mathbf{r}, \mathbf{r}_0)$ such that $|\mathbf{r} - \mathbf{r}_0| = r$ (and over t_0 in steady state). Fig. 5.4 d reveals a sharp change from short-ranged to system-spanning correlations. Taken together, these results corroborate a discontinuous transition from disordered flows to highly ordered, stationary structures.

Part II

Single particle trajectory segmentation

Chapter 1

Theory and Method

Living cells are dynamic, crowded environments where molecules constantly encounter one another, bind and unbind, and interact with heterogeneous structures such as membranes. An example of such dynamics is when a ligand diffuses on the surface of a cell until it docks at a receptor [69–71]. These interactions can take the form of changes in effective diffusivity, confinement, or directed motion. Capturing those dynamics from experimental data is central to turning qualitative observations quantitative models of cellular function.

Super-high resolution microscopy [49–51] coupled to single-particle tracking (SPT) techniques [52–57] provides a direct access to these dynamics. In SPT, individual particles are localized over time, producing trajectories at camera frame rates. Because SPT observes the motion of single objects rather than ensemble averages, it is sensitive to heterogeneity in the system: the same particle can traverse multiple dynamical regimes within a single track. At the same time, SPT measurements are imperfect. Each recorded position is a time average over the camera exposure and is affected by localisation error, induced by the tracking method. The raw output is therefore a sequence of shutter-averaged, noisy positions from which we must infer when the underlying dynamics changed and which dynamical regime was active at each time.

This need leads to the segmentation problem: given a trajectory that may contain multiple latent states, identify the points in time where the state changes and label each interval with a consistent dynamical description. This step is fundamental as, without reliable segmentation, global statistics such as a single diffusion coefficient can be badly biased and binding lifetimes can be systematically wrong. Conversely, accurate segmentation enables an accurate estimation within intervals of the same state.

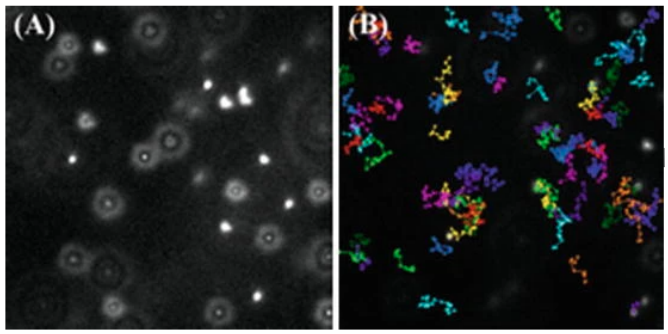


Figure 1.1: (a) Frame from a time-lapse fluorescence microscopy movie showing 200 nm fluorescent nanospheres diffusing in water. (b) Trajectories of individual nanospheres with ≥ 10 recorded positions, extracted from the movie. Reproduced from [123] with permission.

This chapter presents such a method. It is a probabilistic data-driven segmentation pipeline designed for driftless Brownian trajectories where the switches between diffusive states happen at random times. Once the segmentation is obtained, diffusion coefficients and other parameters are estimated.

The remainder of this chapter is organized to move from general principles to concrete procedure. We begin with the foundations of Brownian motion and their implications under finite exposure and localisation noise. Then, we describe the segmentation pipeline in full detail.

1.1 Synthetic trajectories

1.1.1 Brownian motion

At the scale of single particles in a viscous medium, the Reynolds number is vanishingly small and inertial terms relax on timescales much shorter than camera exposure. In this overdamped regime, the coarse-grained dynamics are well described by a balance of viscous drag and thermal fluctuations, leading to an effectively directionless, Brownian motion. The unique parameter fixing the rate of spatial spreading is the diffusion coefficient D. In a homogeneous Newtonian solvent, D relates to temperature T, viscosity η, and hydrodynamic radius a by the Stokes-Einstein formula in three dimensions [2],

$$D = \frac{k_B T}{6\pi\eta a}. \tag{1.1}$$

In *Paper III*, we work under the assumptions of (i) isotropy and homogeneity of the medium, and (ii) absence of systematic drift.

1.1.2 Overdamped Langevin equation

A model that reproduces faithfully the diffusion process described earlier is the overdamped Langevin equation

$$\mathrm{d}\mathbf{X}(t) \;=\; \sqrt{2D(\mathbf{X}(t),t)}\,\mathrm{d}\mathbf{W}(t), \tag{1.2}$$

where $\mathbf{W}(t)$ is a d-dimensional Wiener process, and where the diffusion coefficient D may depend on position and time. For systems with multiple diffusive states, Eq (1.2) becomes:

$$\mathrm{d}\mathbf{X}(t) \;=\; \sqrt{2D_{S(t)}}\,\mathrm{d}\mathbf{W}(t), \tag{1.3}$$

where $S(t) \in \{1,\ldots,K\}$ is a latent state and $D_{S(t)}$ is constant within each individual state segment. Within any interval where $S(t) = i$, the increments to each of the Cartesian components have variance $2D_i\Delta t$. In the piecewise-constant D, this simplifies to

$$\mathbf{X}_{j+1} \;=\; \mathbf{X}_j \;+\; \sqrt{2D_i\,\delta t}\;\mathbf{Z}_j \qquad \text{while } S(t_j) = i \tag{1.4}$$

where each of the components of $\mathbf{Z}_j$ is drawn from a normal distribution with zero mean and unit variance, and δt denotes the discretization step.

1.1.3 Poisson point process

We work under the assumption that the changes in diffusivity are memoryless, this can be done using a Poisson point process (PPP) for switching events. For trajectories with multiple diffusive regimes, we couple the Brownian diffusion to a continuous-time Markov chain $S(t) \in \{1,\ldots,K\}$ with generator $Q = (q_{ij})$. While in state i, the diffusion coefficient is D_i, and switches occur at total rate $\lambda_i = \sum_{j\neq i} q_{ij}$. Upon a change in diffusivity, the next state is j with probability q_{ij}/λ_i. The coupled dynamics are

$$\mathrm{d}\mathbf{X}(t) \;=\; \sqrt{2D_{S(t)}}\,\mathrm{d}\mathbf{W}(t), \tag{1.5}$$

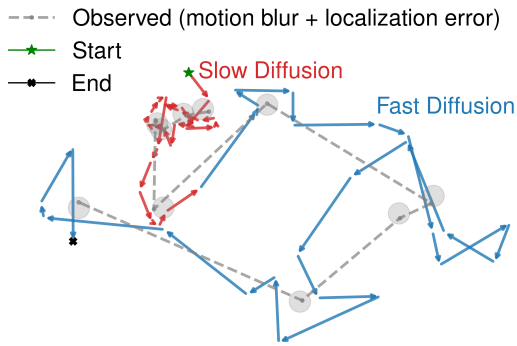


Figure 1.2: Example of a two-state Brownian trajectory. The true particle path is shown as colour-coded arrows, while the observed trajectory, depicted as a dashed grey line with circular markers, is influenced by motion blur ($n = 5$), where positions are averaged over non-overlapping time intervals. localisation error is introduced by adding Gaussian noise to the blurred positions, illustrated by circles surrounding each point to represent the positional uncertainty.

$$\mathbb{P}[S(t+\mathrm{d}t) = j \mid S(t) = i] \;=\; q_{ij}\,\mathrm{d}t \;+\; o(\mathrm{d}t)\;(j \neq i), \tag{1.6}$$

with $\mathbb{P}[S(t+\mathrm{d}t) = i \mid S(t) = i] = 1 - \lambda_i\,\mathrm{d}t + o(\mathrm{d}t)$. The mean lifetime in state i is $\tau_i = 1/\lambda_i$ [124].

1.1.4 Camera statistics: motion blur and localisation noise

In real systems, the trajectories obtained are shutter-averaged noisy positions due to the camera exposure time and the localisation error introduced by the tracking procedure. To mimic these effects, starting from the true trajectory $\mathbf{X}(t)$ obtained as discussed earlier, we introduce the observed position $\bar{\mathbf{X}}_j$

$$\bar{\mathbf{X}}_j \;=\; \int_{(j-1)\Delta t}^{j\Delta t} s[t' - (j-1)\Delta t]\,\mathbf{X}(t')\,\mathrm{d}t' \;+\; \boldsymbol{\xi}_j, \tag{1.7}$$

$$\boldsymbol{\xi}_j \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbb{1}_d) \text{ i.i.d.}, \tag{1.8}$$

where $\boldsymbol{\xi}_j$ is a zero-mean Gaussian noise with $\langle \boldsymbol{\xi}_i \boldsymbol{\xi}_j^{\mathsf{T}} \rangle = \sigma^2 \delta_{ij} \mathbb{1}_d$ representing the localisation error [125]. The integral represents the blurring due to the exposure time of the camera Δt with the shutter function $s(t)$ a non-negative function whose integral over Δt is equal to unity. The shutter function describes the illumination intensity

profile in a single frame. The measured displacements for the i-th Cartesian coordinate $\Delta\bar{X}_j^i = \bar{X}_{j+1}^i - \bar{X}_j^i$ becomes distributed according to a multivariate Gaussian distribution with [125]

$$\langle \Delta\bar{X}_j^i \rangle = 0 \tag{1.9}$$

$$\langle \Delta\bar{X}_j^i \Delta\bar{X}_l^i \rangle = \begin{cases} 2D\Delta t - 2(2DR\Delta t - \sigma^2), & j = l \\ 2DR\Delta t - \sigma^2, & j = l \pm 1, \\ 0 & \text{otherwise} \end{cases} \tag{1.10}$$

where R is the motion blur coefficient defined as

$$R = \frac{1}{\Delta t} \int_0^{\Delta t} S(t)(1 - S(t))\, \mathrm{d}t, \tag{1.11}$$

with $S(t)$ the fraction of illumination in the interval $[0, t]$ defined as

$$S(t) = \int_0^t s(t)\, \mathrm{d}t. \tag{1.12}$$

In our case, we consider constant illumination ($s(t) = 1/\Delta t$) which yields the motion blur coefficient $R = 1/6$ [125].

Practically, we first generate the true trajectory $\mathbf{X}(t)$ as described in Section 1.1.2. Then we mimic the shutter-averaging effect introduced by the camera acquisition considering constant illumination and the localisation error

$$\bar{\mathbf{X}}_j \;\approx\; \frac{1}{M} \sum_{k=0}^{n-1} \mathbf{X}(t_j + k\delta t) + \boldsymbol{\xi}_j, \qquad n = \Delta t/\delta t, \tag{1.13}$$

as illustrated in Fig. 1.2.

1.2 Segmentation of trajectories

Given the shutter-averaged, noisy positions $\{\bar{\mathbf{X}}_j\}_{j=0}^N$ generated as described in Eq. (1.13) and an unknown piecewise-constant state sequence $\{S_j\}$, the objective is to infer per-frame labels $\widehat{S}_j \in \{1, \ldots, K\}$. The method proceeds as follows (Fig. 1.3):

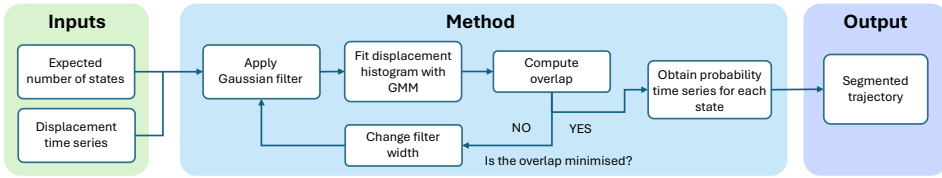


Figure 1.3: Schematic of the main steps in the method. Given a set of trajectories and a specified number of states, the overlap between the displacement distributions is minimised by applying an optimised Gaussian filter to the displacement time series. The resulting output provides a segmentation of each trajectory, identifying the instantaneous state of the particle at every time point.

1. Compute the d-dimensional displacement as the Euclidean distance.
2. Optimally denoise it with a Gaussian filter by finding the filter size that minimizes the overlap between the K components of their distribution.
3. Assign labels to each frame using the Bayesian decision rule.

Starting from the shutter-averaged, noisy positions $\{\bar{\mathbf{X}}_j\}_{j=0}^{N}$, we compute the displacement time series $\Delta r(t_j) = \|\bar{\mathbf{X}}_{j+1} - \bar{\mathbf{X}}_j\|$. For a 2-dimensional Brownian trajectory of K components, the distribution of the displacements is a Rayleigh mixture with the same number of components K where the scale of the i-th state is $\sigma_i' = \sqrt{2D_i\Delta t - 2(2D_iR\Delta t - \sigma^2)}$ [125–127]:

$$P(\Delta r) = \sum_{i=1}^{K} \pi_i \frac{\Delta r}{\left(\sigma_i'\right)^2} \exp\left(- \frac{r^2}{2(\sigma_i')^2} \right), \tag{1.14}$$

with:

$$\pi_i \geq 0, \quad \sum_{i=1}^{K} \pi_i = 1. \tag{1.15}$$

As is, the displacement time series $\Delta r(t_j)$ is highly noisy and no inference can be done to correctly label each displacement to the underlying diffusive state. The idea behind the method is to reduce the noise in $\Delta r(t_j)$ by applying an appropriate Gaussian filter that will reduce the noise without smearing the different diffusive states.

1.2.1 Gaussian filtering and noise reduction in time series

For a displacement time series $\Delta r(t_j)$, the filtered displacement $\widehat{\Delta r}(t_j)$ is defined as

$$\widehat{\Delta r}_j(f) \;=\; \frac{\sum_{m=0}^{N-1} \exp\Big(-\frac{(j-m)^2}{2f^2}\Big)\,\Delta r_m}{\sum_{m=0}^{N-1} \exp\Big(-\frac{(j-m)^2}{2f^2}\Big)}. \tag{1.16}$$

where f is the size of the Gaussian filter. Near boundaries, we truncate and renormalize the Gaussian kernel to keep weights summing to one.

On one hand, for $f \longrightarrow 0$, the Gaussian kernel acts as a Dirac δ function, too small values of f are not useful as the Gaussian filter does not remove enough noise to separate the different states in the time series. On the other hand, across a diffusive state boundary, large values of f smear the diffusive states making them indistinguishable. Hence, for a displacement times series $\Delta r(t_j)$ characterized by the diffusion coefficients D_i and the state lifetimes τ_i for $i \in \{1, \ldots, K\}$, a judicious choice of the filter size f has to be made to reduce the noise in $\Delta r(t_j)$ without blurring the transitions between states. Another point of interest, which is fundamental to the method, is that, for lifetimes τ_i larger than f, each of the filtered displacements $\{\widehat{\Delta r}_j\}_{j=0}^{N-1}$ should be dominated by contributions from a single state. Furthermore, the variance of each state is finite. These two conditions are sufficient for the Lindeberg-Feller Central Limit theorem [128] to hold, hence the filtered displacements $\{\widehat{\Delta r}_j\}_{j=0}^{N-1}$ are distributed according to a univariate Gaussian mixture with K components.

1.2.1.1 Data-driven identification of the optimal filter size

To determine the optimal filter size f^*, we define the cost function $\theta(f)$ as the total overlap of K-component Gaussian mixture fitted to the filtered displacements $\{\widehat{\Delta r}(f)\}$ [129, 130]:

$$\theta(f) \;=\; \sum_{i<j} \int \min\Big\{w_i\,\mathcal{N}(\widehat{\Delta r}\,|\,\mu_i, \sigma_i^2),\; w_j\,\mathcal{N}(\widehat{\Delta r}\,|\,\mu_j, \sigma_j^2)\Big\}\;\mathrm{d}\widehat{\Delta r}. \tag{1.17}$$

The rationale behind the choice of $\theta(f)$ as a cost function and $f^\star = \arg\min_f \theta(f)$ as the optimal filter is intuitive. Small filter sizes do not efficiently reduce noise in the displacement time series $\Delta(t_j)$ and the different diffusive states are hardly distinguishable leading to a large value of θ. Large filter sizes on the other lead to the smearing of the boundary between diffusive states which would also lead to a large value of θ. Hence, this suggests the existence of an optimal filter size that minimizes θ without smearing the states.

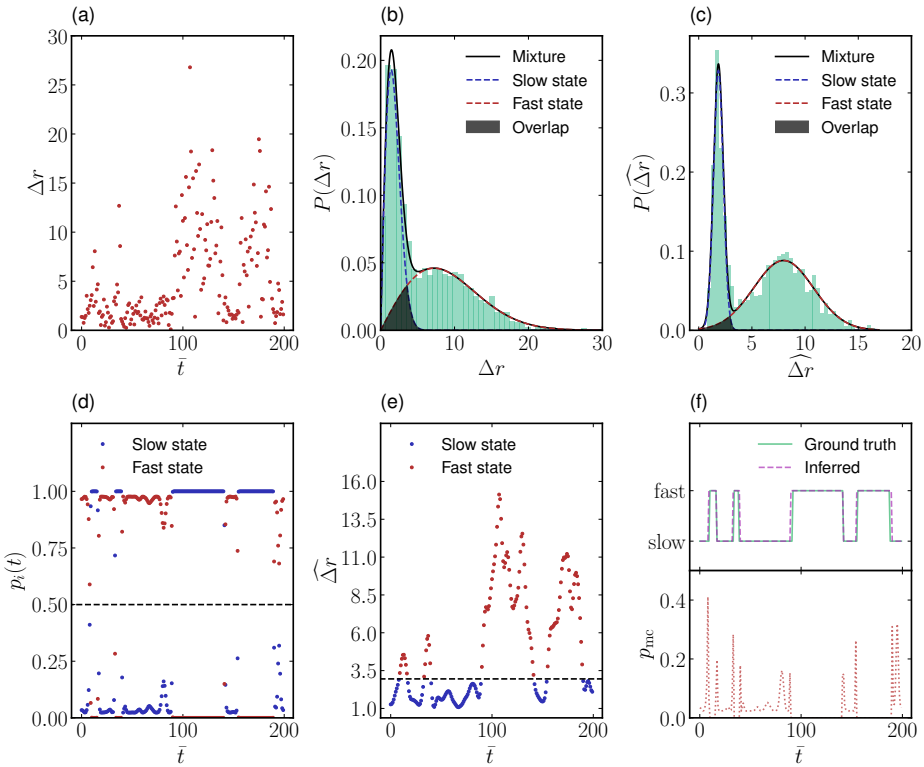


Figure 1.4: (a) Displacement time series of a two-state Brownian particle with lifetimes $\tau_1 = \tau_2 = 20$ and $D_2/D_1 = 5$. (b) Displacement distribution of the unfiltered trajectories, fitted to a Rayleigh mixture model as given in Eq. (1.14). The shaded regions indicate the overlap between the displacement distributions of the two states. (c) Displacement distribution of the same trajectories after applying the optimal filter with width f^*, fitted to a Gaussian mixture following Eq. (1.18). (d) Time series of the probability p_i (Eq. (1.20)) that the particle belongs to the diffusive state i, obtained from the filtered displacement data. The black dashed line marks the threshold corresponding to $p_i = 0.5$. (e) Filtered displacement time series from panel (a) using the optimal filter f^*. The dashed line indicates the optimized threshold separating states 1 and 2 (Eq. (1.25)), determined by the intersection of the two fitted Gaussians shown in panel (c). (f) Comparison between the ground-truth and inferred state time series of the trajectory (top panel) and the inferred probability of misclassification, $p_{\mathrm{mc}} = \min(p_1, p_2)$.

1.2.2 Gaussian mixture model

To find the optimal filter size f^* that minimizes the cost function θ, we perform a grid search for all values of $f \in [0, \min_{i\in\{1,\dots,K\}} \tau_i]$. For each value of f, we fit the filtered displacements distribution $P(\widehat{\Delta r})$ using a Gaussian mixture model (GMM) [129, 130] and compute θ. The interest of using GMMs is threefold: (i) GMMs do not require any binning, as the likelihood is directly maximized using the Expectation-Maximization (EM) algorithm, (ii) parameter constraints are handled natively, (iii) and most importantly, GMMs assign soft-labels to each filtered displacement which

gives the probability of each filtered displacement to be in a given diffusive state.

Starting from the filtered displacement probability density $P(\widehat{\Delta r})$ with parameters $\Gamma = \{(w_i, \mu_i, \sigma_i^2)\}_{i=1}^K$

$$P(\widehat{\Delta r}(f)) \;=\; \sum_{i=1}^{K} w_i \mathcal{N}(\widehat{\Delta r}(f) \mid \mu_i, \sigma_i^2), \qquad w_i > 0,\ \sum_i w_i = 1, \tag{1.18}$$

with w_i, μ_i and σ_i^2 respectively the weight, mean, and variance of the i-th component, the log-likelihood ℓ, given a set of parameters $\Gamma = \{(w_i, \mu_i, \sigma_i^2)\}_{i=1}^K$, is defined as

$$\ell(\Gamma) \;=\; \sum_{n=1}^{N-1} \log\left(\sum_{i=1}^{K} w_i \mathcal{N}(\widehat{\Delta r}(t_j) \mid \mu_i, \sigma_i^2)\right) \tag{1.19}$$

1.2.2.1 The Expectation-Maximization algorithm

The Expectation-Maximization (EM) algorithm is an iterative procedure to maximize the likelihood of the model parameters, namely the weight, mean, and variance of each component of the Gaussian mixture. As indicated by its name, the algorithm iterates between an "expectation step" at which the responsibilities γ_{ji} are computed, followed by a "maximization step" where the mixture parameters Γ are updated. As the likelihood is non decreasing under the EM procedure, the iterations are performed until the change in likelihood per iteration is lower than a predefined tolerance [129, 130].

Formally, for a set of parameters $\Gamma = \{(w_i, \mu_i, \sigma_i^2)\}_{i=1}^K$ characterizing a Gaussian mixture, the EM iterations are:

E-step (computing responsibilities) We compute the posterior membership probability (responsibility) γ_{ji} for each filtered displacement $\widehat{\Delta r}(t_j)$
and component i

$$\gamma_{ji} \;=\; \frac{w_i \mathcal{N}(\widehat{\Delta r}(t_j) \mid \mu_i, \sigma_i^2)}{\sum_{j=1}^{K} w_j \mathcal{N}(\widehat{\Delta r}(t_j) \mid \mu_j, \sigma_j^2)}. \tag{1.20}$$

M-step (updating parameters) We update the mixture parameters using the posterior membership probability computed in the E-step

$$w_i^{\text{new}} \leftarrow \frac{1}{N-1}\sum_{n=1}^{N-1}\gamma_{ji}, \tag{1.21}$$

$$\mu_i^{\text{new}} \leftarrow \frac{\sum_j \gamma_{ji}\widehat{\Delta r}(t_j)}{\sum_j \gamma_{ji}}, \tag{1.22}$$

$$(\sigma_i^2)^{\text{new}} \leftarrow \frac{\sum_j \gamma_{ji}(\widehat{\Delta r}(t_j) - \mu_i^{\text{new}})^2}{\sum_j \gamma_{ji}}. \tag{1.23}$$

This steps are iterated until the relative increase in

$$\ell(\Gamma) \;=\; \sum_{n=1}^{N-1}\log\left(\sum_{i=1}^{K} w_i\,\mathcal{N}(\widehat{\Delta r}(t_j)\mid \mu_i,\sigma_i^2)\right) \tag{1.24}$$

falls below a predefined tolerance [129, 130].

1.2.3 Framewise segmentation, decision boundaries, and uncertainty

Once the overlap is minimized, we use the resulting responsibilities obtained with the GMM fit. To each displacement Δr_j we associate a hard label $\widehat{S}_j = \arg\max_i \gamma_{ji}$ and soft labels $(\gamma_{j1},\ldots,\gamma_{jK})$. The hard labels time series $\widehat{S}(t_j)$ corresponds to the desired segmentation of the initial trajectory and the soft labels time series $p_i(t_j) = \gamma_{ji}$ represents the probability of a displacement Δr_j belonging to a diffusive state i. For $K = 2$, this is equivalent to simply choosing the decision boundary c that solves

$$w_1\,\mathcal{N}(c\mid\mu_1,\sigma_1^2) \;=\; w_2\,\mathcal{N}(c\mid\mu_2,\sigma_2^2), \tag{1.25}$$

and the rule "$\widehat{\Delta r_j} < c \Rightarrow$ state 1; $\widehat{\Delta r_j} \geq c \Rightarrow$ state 2".

1.3 Extraction of dynamical quantities

Once the segmentation $\widehat{S}_j$ of the trajectory is obtained, the diffusion coefficients D_i are estimated using Eq. (1.10) [125]:

$$D_i \;=\; \frac{\mathbb{E}\Big[(\Delta\bar{X}_j)^2 \mid \widehat{S}_j = i\Big] - 2\sigma^2}{2\Delta t\,(1-2R)}. \tag{1.26}$$

When the localisation error magnitude σ or the motion blur R are unknown, the off-diagonal terms of the displacements covariance matrix, defined in Eq. (1.10), can be used to estimate the diffusion coefficient [125]:

$$D_i \;=\; \frac{\mathbb{E}\Big[(\Delta\bar{X}_j)^2 \mid \widehat{S}_j = i\Big] + 2\,\mathbb{E}\Big[\Delta\bar{X}_j\Delta\bar{X}_{j+1} \mid \widehat{S}_j = i, \widehat{S}_{j+1} = i\Big]}{2\Delta t}. \tag{1.27}$$

For the lifetimes estimation, as discussed in Section 1.1.3, switching between diffusive states is modelled as a PPP. This leads to exponentially distributed lifetimes $\tau_i \sim \mathrm{Exp}(\lambda_i)$, therefore the lifetimes can be readily obtained by fit [131].

Chapter 2

Results

2.1 Overview

This chapter evaluates the segmentation method introduced earlier and quantifies its performance with respect to state separability, characterised by the diffusion coefficients of the different diffusive states, the state lifetimes, and measurement artefacts (motion blur and localisation error). The objectives are: (i) to identify conditions under which framewise state labels are reliable for synthetic data as described in Section 1.1; and (ii) to assess the accuracy of parameters inferred from the recovered segments, in particular diffusion coefficients and lifetimes. To do so, we examine the relevance of the overlap $\theta_{12}(f)$ as the cost function to minimise, report segmentation accuracy across diffusion-coefficient ratios and lifetimes, characterise robustness to localisation error and motion blur, and finally assess how segmentation quality propagates to the accuracy of diffusion-coefficient and lifetime estimates.

2.2 Experimental setup and evaluation parameters

Unless stated otherwise, results are based on simulated datasets with $N = 200$ trajectories, each of length $T = 2000$ frames, acquired at a sampling interval Δt. We consider constant illumination with motion-blur coefficient $R = 1/6$. State separability is explored by sweeping the diffusion coefficient ratio $\tilde{D} \equiv D_{\mathrm{fast}}/D_{\mathrm{slow}}$ and the shorter mean lifetime $\tilde{\tau} \equiv \min(\tau_{\mathrm{fast}}, \tau_{\mathrm{slow}})$. Robustness is assessed against the rescaled localisation error $\tilde{\sigma} = \sigma/\sqrt{2D_{\mathrm{slow}}\Delta t}$ and against synthetic motion blur emulated by varying $\Delta t = n\,\delta t$.

The performance of the segmentation is quantified by the framewise accuracy

$$\text{Accuracy}(f) \;=\; \frac{1}{T-1}\sum_{t=1}^{T-1} \mathbb{1}\big[\,\hat{S}_t(f) = S_t\,\big]. \tag{2.1}$$

Here, $\hat{S}_t(f)$ is the inferred diffusive state at time t, S_t is the ground truth, and $\mathbb{1}[\cdot]$ is the indicator function, equal to unity if the prediction is correct, to zero otherwise.

2.3 Segmentation performance across parameter space

2.3.1 Cost function and optimisation of the filter size

We justify choosing the overlap $\theta_{12}(f)$ as the cost function by analysing, as a function of the filter size f, the behaviour of both $\theta_{12}(f)$ and the framewise accuracy in Eq. (2.1). Empirically, $\theta_{12}(f)$ is convex in f, making the optimisation well-posed, and the accuracy maxima coincide with the overlap minima across $\tilde{D}$ and $\tilde{\tau}$. Hence, minimising overlap is an effective proxy for maximising segmentation accuracy (Figs 2.1 b and c).

Having established that $\theta_{12}(f)$ reliably guides the choice of f, we next quantify segmentation accuracy across $(\tilde{D}, \tilde{\tau})$ and under common measurement artefacts, using the setup in Section 2.2.

2.3.2 Accuracy across parameter space

Mapping accuracy over $(\tilde{D}, \tilde{\tau})$ in the absence of localisation error and motion blur shows that the accuracy exceeds 90% across a broad region but declines when states are weakly separated ($\tilde{D} \lesssim 3$) or when $\tilde{\tau}$ is short ($\tilde{\tau} \lesssim 20$ frames). A clear trade-off emerges: smaller $\tilde{\tau}$ requires larger $\tilde{D}$ to maintain separability after filtering, whereas larger $\tilde{\tau}$ relaxes this requirement, as illustrated in Fig 2.1 a.

We examine the effect of per-trajectory length T while holding the total number of frames fixed across the dataset at $N \times T = 20\,000$. Segmentation accuracy is essentially independent of T, which is practically important because single-particle trajectories are often short due to photobleaching or tracking losses (Fig. 2.1 d). This property holds provided that most trajectories contain at least one state switch .

Finally, we assess the robustness to measurement artefacts, localisation error and motion blur. Varying the rescaled localisation error (Fig 2.1 c), $\tilde{\sigma} = \sigma/\sqrt{2D_{\text{slow}}\Delta t}$,

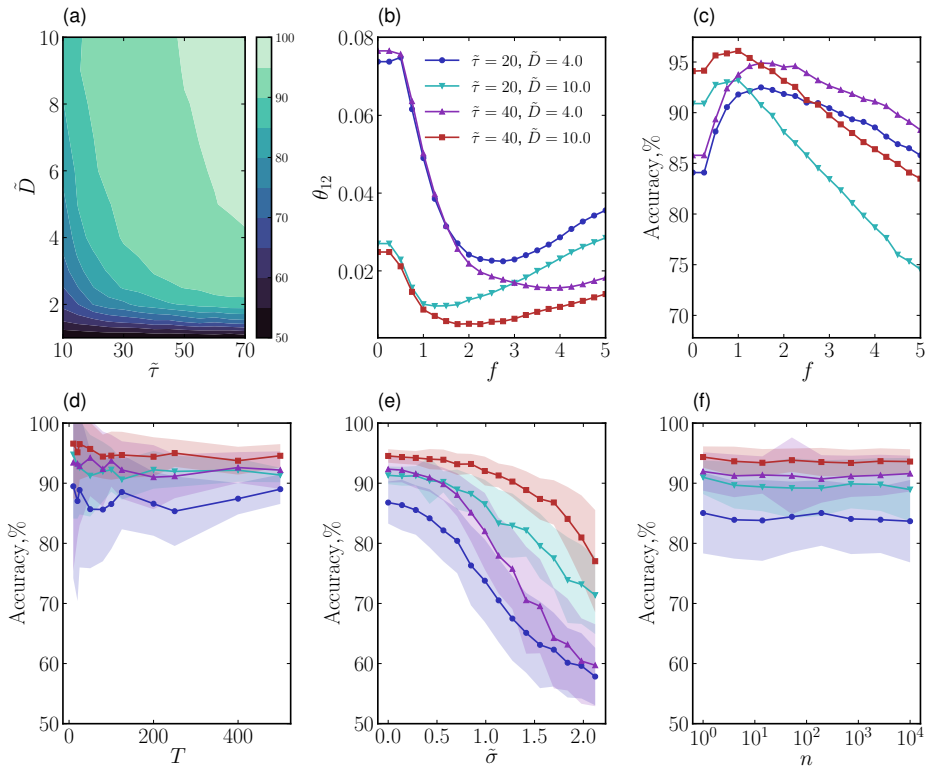


Figure 2.1: (a) Contour plot showing the segmentation accuracy as a function of the ratio of diffusion coefficients $\tilde{D}$ and shortest mean state duration $\tilde{\tau}$, averaged over $N = 100$ trajectories of length $T = 2000$. (b) Overlap between the displacement distributions of the two states, denoted θ_{12}, and (c) segmentation accuracy, both shown as functions of the filter width f for the same values of $\tilde{D}$ and $\tilde{\tau}$ as in the other panels. (d) Segmentation accuracy as a function of the trajectory length T for different $\tilde{D}$ and $\tilde{\tau}$, with the total number of sampled displacements fixed to $N \times T = 20000$. (e) Accuracy as a function of the localisation error $\tilde{\sigma}$ for the same parameters $\tilde{D}$ and $\tilde{\tau}$ as in panel (b). (f) Accuracy as a function of motion-blur intensity, simulated by varying n in Eq. (3). In this case, the "ground truth" used to assess segmentation accuracy was determined by a majority vote over all n points involved in the blurring process. Shaded regions, where present, indicate one standard deviation of the accuracy across $N = 100$ trajectories.

shows that accuracy is stable for $\tilde{\sigma} < 1$ and decays gradually thereafter, consistent with increasing overlap between the filtered state distributions; this indicates that the filter suppresses the noise introduced by the localisation error effectively until the noise scale approaches the mean displacement of the slowest state. Motion blur is emulated by averaging positions over the exposure interval $\Delta t = n\,\delta t$, and segmentation accuracy remains essentially unchanged over the tested range of n (Fig 2.1 f). In the next section, we examine how segmentation accuracy propagates to parameter inference.

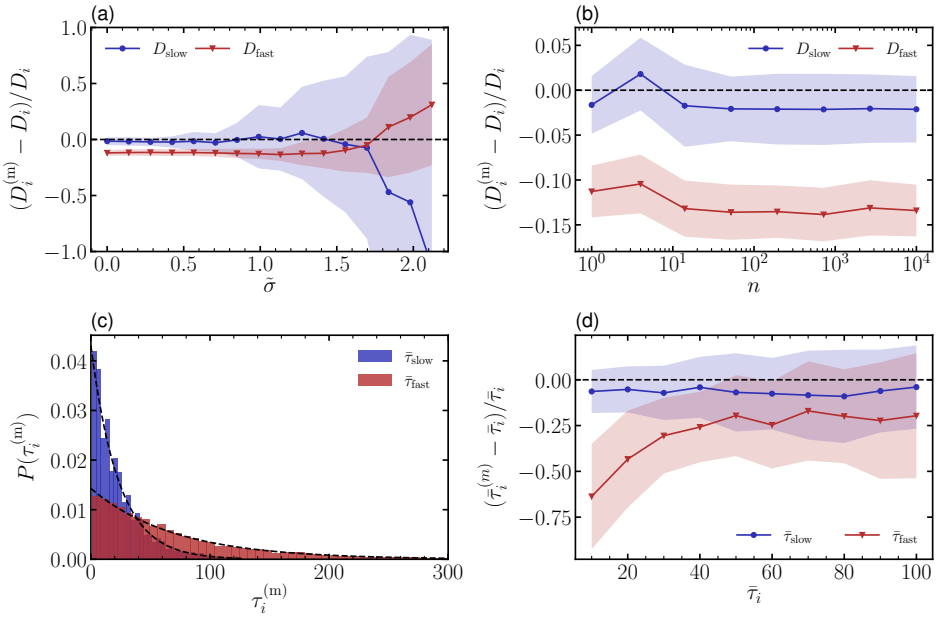


Figure 2.2: (a,b) Mean relative error in the inferred diffusion coefficient $D_i^{(m)}$ as a function of (a) the rescaled localisation error $\tilde{\sigma}$ and (b) the blurring strength n. In both cases, $\tilde{D} = 10$ and $\bar{\tau}_1 = \bar{\tau}_2 = 20$ are used. (c) Probability distribution of the measured lifetimes $\tau_i^{(m)}$. The distributions are fitted with exponential functions (dashed lines), yielding estimated average lifetimes of $\bar{\tau}_1^{(m)} = 23.3$ and $\bar{\tau}_2^{(m)} = 70.1$, compared to the true values $\bar{\tau}_1 = 20$ and $\bar{\tau}_2 = 60$. This corresponds to a relative error of approximately 15%. (d) Relative mean error of the measured average lifetimes $\bar{\tau}_{\mathrm{fast}}^{(m)}$ and $\bar{\tau}_{\mathrm{slow}}^{(m)}$ as a function of the true lifetime $\bar{\tau}_i$, with $\bar{\tau}_{\mathrm{fast}} = \bar{\tau}_1$ and $\bar{\tau}_{\mathrm{slow}} = \bar{\tau}_2$. Shaded regions represent one standard deviation of the estimate obtained from $N = 200$ trajectories.

2.3.3 Inference of dynamical parameters from segments

Diffusion coefficients. After trajectory segmentation, diffusion coefficients are estimated via the Berglund correction (Eq. (1.10)) under constant illumination ($R = 1/6$). With shutter and localisation-error corrections in place, Figs 2.2 a and b show that the estimates remain robust to moderate localisation error, and are essentially insensitive to motion blur. A small systematic effect persists for the fast state, where D_{fast} tends to be underestimated by about 10% due to asymmetrical misclassification near transitions, where slow-to-fast errors are slightly more frequent than the converse.

Lifetimes. State lifetimes are inferred by fitting the lifetimes measured between inferred switches to an exponential distribution, as shown in Fig 2.2 c. Fig 2.2 d shows that this procedure is more sensitive to segmentation errors than diffusion estimation, since a single mislabelled frame within a lifetime creates a spurious transition and bi-

ases lifetimes downwards. Nevertheless, for $\tau \gtrsim 30$ frames and with sufficiently long trajectories, relative errors typically remain within $\pm 25\%$.

References

[1] R. Brown, Philos. Mag. **4**, 161 (1828).

[2] A. Einstein, Ann. Phys. **322**, 549 (1905).

[3] M. von Smoluchowski, Ann. Phys. **326**, 756 (1906).

[4] J. Happel and H. Brenner, *Low Reynolds number hydrodynamics* (Kluwer Academic, 1983).

[5] D. S. Lemons and A. Gythiel, Am. J. Phys. **65**, 1079 (1997).

[6] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).

[7] H. Risken, *The Fokker-Planck equation*, 2nd ed. (Springer, 1996).

[8] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).

[9] S. Ramaswamy, Annu. Rev. Condens. Matter Phys. **1**, 323 (2010).

[10] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Rev. Mod. Phys. **85**, 1143 (2013).

[11] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Rev. Mod. Phys. **88** (2016).

[12] J. Deseigne, O. Dauchot, and H. Chaté, Phys. Rev. Lett. **105**, 098001 (2010).

[13] C. Dombrowski, L. Cisneros, S. Chatkaew, R. E. Goldstein, and J. O. Kessler, Phys. Rev. Lett. **93**, 098103 (2004).

[14] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Löwen, and J. M. Yeomans, Proc. Natl. Acad. Sci. U. S. A. **109**, 14308 (2012).

[15] J. Dunkel, S. Heidenreich, K. Drescher, H. H. Wensink, M. Bär, and R. E. Goldstein, Phys. Rev. Lett. **110**, 228102 (2013).

[16] T. Sanchez, D. T. N. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic, Nature **491**, 431 (2012).

[17] F. C. Keber, E. Loiseau, T. Sanchez, S. J. DeCamp, L. Giomi, M. J. Bowick, M. C. Marchetti, Z. Dogic, and A. R. Bausch, Science **345**, 1135 (2014).

[18] S. J. DeCamp, G. S. Redner, A. Baskaran, M. F. Hagan, and Z. Dogic, Nat. Mater. **14**, 1110 (2015).

[19] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, I, and O. Shochet, Phys. Rev. Lett. **75**, 1226 (1995).

[20] J. Toner and Y. Tu, Phys. Rev. E **58**, 4828 (1998).

[21] J. Toner, Y. Tu, and S. Ramaswamy, Ann. Phys. **318**, 170 (2005).

[22] A. Doostmohammadi, J. Ignés-Mullol, J. M. Yeomans, and F. Sagués, Nat. Commun. **9**, 3246 (2018).

[23] M. E. Cates and J. Tailleur, Annu. Rev. Condens. Matter Phys. **6**, 219 (2015).

[24] Y. Fily and M. C. Marchetti, Phys. Rev. Lett. **108**, 235702 (2012).

[25] F. Höfling and T. Franosch, Rep. Prog. Phys. **76**, 046602 (2013).

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

[27] B. Wang, S. M. Anthony, S. C. Bae, and S. Granick, Proc. Natl. Acad. Sci. U. S. A. **106**, 15160 (2009).

[28] M. V. Chubynsky and G. W. Slater, Phys. Rev. Lett. **113**, 098302 (2014).

[29] A. V. Chechkin, F. Seno, R. Metzler, and I. M. Sokolov, Phys. Rev. X. **7** (2017).

[30] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser, Nat. Methods **5**, 695 (2008).

[31] F. Persson, M. Lindén, C. Unoson, and J. Elf, Nat. Methods **10**, 265 (2013).

[32] C. L. Vestergaard, P. C. Blainey, and H. Flyvbjerg, Phys. Rev. E **89**, 022726 (2014).

[33] N. Monnier, Z. Barry, H. Y. Park, K.-C. Su, Z. Katz, B. P. English, A. Dey, K. Pan, I. M. Cheeseman, R. H. Singer, and M. Bathe, Nat. Methods **12**, 838 (2015).

[34] J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, and K. W. Eliceiri, Methods **115**, 80 (2017).

[35] Azonesa and Radomianin, Aerial view of a flock of sheep with a shepherd in china (derivative work), `https://commons.wikimedia.org/wiki/File:Herding_sheep353_(edited).jpg` (2015), accessed: 2025-10-28. Licensed under Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

[36] Bluesbby, A school of fish, `https://commons.wikimedia.org/wiki/File:A_school_of_fish_(29882069786).jpg` (n.d.), accessed: 2025-10-28. Licensed under Creative Commons Attribution 2.0 Generic (CC BY 2.0).

[37] H. Chaté, Annu. Rev. Condens. Matter Phys. **11**, 189 (2020).

[38] A. Bricard, J.-B. Caussin, N. Desreumaux, O. Dauchot, and D. Bartolo, Nature **503**, 95 (2013).

[39] A. Bricard, J.-B. Caussin, D. Das, C. Savoie, V. Chikkadi, K. Shitara, O. Chepizhko, F. Peruani, D. Saintillan, and D. Bartolo, Nat. Commun. **6**, 7470 (2015).

[40] A. Morin and D. Bartolo, Phys. Rev. X. **8** (2018).

[41] L. Giomi, Phys. Rev. X. **5** (2015).

[42] S. Ramaswamy, R. A. Simha, and J. Toner, EPL **62**, 196 (2003).

[43] V. Narayan, S. Ramaswamy, and N. Menon, Science **317**, 105 (2007).

[44] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud, Phys. Rev. E 77, 046113 (2008).

[45] D. Saintillan and M. J. Shelley, Phys. Rev. Lett. **99**, 058102 (2007).

[46] D. Saintillan and M. J. Shelley, Theory of active suspensions, in *Complex Fluids in Biological Systems* (Springer New York, New York, NY, 2015) pp. 319–355.

[47] A. Sokolov and I. S. Aranson, Phys. Rev. Lett. **109**, 248109 (2012).

[48] Z. Liu, W. Zeng, X. Ma, and X. Cheng, Soft Matter **17**, 10806 (2021).

[49] W. E. Moerner and L. Kador, Phys. Rev. Lett. **62**, 2535 (1989).

[50] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, Proc. Natl. Acad. Sci. U. S. A. **97**, 8206 (2000).

[51] W. E. Moerner and D. P. Fromm, Rev. Sci. Instrum. **74**, 3597 (2003).

[52] G. Seisenberger, M. U. Ried, T. Endress, H. Büning, M. Hallek, and C. Bräuchle, Science **294**, 1929 (2001).

[53] C. Manzo and M. F. Garcia-Parajo, Rep. Prog. Phys. **78**, 124601 (2015).

[54] F. Simon, L. E. Weiss, and S. van Teeffelen, Nat. Rev. Methods Primers **4** (2024).

[55] M. Orrit and J. Bernard, Phys. Rev. Lett. **65**, 2716 (1990).

[56] H. Shen, L. J. Tauzin, R. Baiyasi, W. Wang, N. Moringo, B. Shuang, and C. F. Landes, Chem. Rev. **117**, 7331 (2017).

[57] J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, and K. W. Eliceiri, Methods **115**, 80 (2017).

[58] M. Lelek, M. T. Gyparaki, G. Beliu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, Nat. Rev. Methods Primers **2** (2022).

[59] W. M. Śmigiel, L. Mantovanelli, D. S. Linnik, M. Punter, J. Silberberg, L. Xiang, K. Xu, and B. Poolman, Sci. Adv. **8**, eab05387 (2022).

[60] M. de Brabander, R. Nuydens, A. Ishihara, B. Holifield, K. Jacobson, and H. Geerts, J. Cell Biol. **112**, 111 (1991).

[61] N. Ruthardt, D. C. Lamb, and C. Bräuchle, Mol. Ther. **19**, 1199 (2011).

[62] J. A. Varela, C. Åberg, J. C. Simpson, and K. A. Dawson, Small **11**, 2026 (2015).

[63] T. Germier, S. Kocanova, N. Walther, A. Bancaud, H. A. Shaban, H. Sellou, A. Z. Politi, J. Ellenberg, F. Gallardo, and K. Bystricky, Biophys. J. **113**, 1383 (2017).

[64] E. Amselem, B. Broadwater, T. Hävermark, M. Johansson, and J. Elf, Nat. Commun. **14**, 1336 (2023).

[65] R. Simson, E. D. Sheets, and K. Jacobson, Biophys. J. **69**, 989 (1995).

[66] S. Huet, E. Karatekin, V. S. Tran, I. Fanget, S. Cribier, and J.-P. Henry, Biophys. J. **91**, 3542 (2006).

[67] T. Morisaki, W. G. Müller, N. Golob, D. Mazza, and J. G. McNally, Nat. Commun. **5**, 4456 (2014).

[68] J.-B. Masson, P. Dionne, C. Salvatico, M. Renner, C. G. Specht, A. Triller, and M. Dahan, Biophys. J. **106**, 74 (2014).

[69] M. Chouliara, V. Junghans, T. Dam, A. M. Santos, S. J. Davis, and P. Jönsson, Biophys. J. **120**, 5032 (2021).

[70] L. Veya, J. Piguet, and H. Vogel, J. Biol. Chem. **290**, 27723 (2015).

[71] V. T. Chang, R. A. Fernandes, K. A. Ganzinger, S. F. Lee, C. Siebold, J. McColl, P. Jönsson, M. Palayret, K. Harlos, C. H. Coles, E. Y. Jones, Y. Lui, E. Huang, R. J. C. Gilbert, D. Klenerman, A. R. Aricescu, and S. J. Davis, Nat. Immunol. **17**, 574 (2016).

[72] N. Granik, L. E. Weiss, E. Nehme, M. Levin, M. Chein, E. Perlson, Y. Roichman, and Y. Shechtman, Biophys. J. **117**, 185 (2019).

[73] G. Muñoz-Gil, G. Volpe, M. A. Garcia-March, E. Aghion, A. Argun, C. B. Hong, T. Bland, S. Bo, J. A. Conejero, N. Firbas, Ò. Garibo I Orts, A. Gentili, Z. Huang, J.-H. Jeon, H. Kabbech, Y. Kim, P. Kowalek, D. Krapf, H. Loch-Olszewska, M. A. Lomholt, J.-B. Masson, P. G. Meyer, S. Park, B. Requena, I. Smal, T. Song, J. Szwabiński, S. Thapa, H. Verdier, G. Volpe, A. Widera, M. Lewenstein, R. Metzler, and C. Manzo, Nat. Commun. **12**, 6253 (2021).

[74] L. D. Landau, E. M. Lifshits, and E. M. Lifshitz, *Fluid Mechanics: Vol 6* (Butterworth-Heinemann, 1987).

[75] G. K. Batchelor, *An introduction to fluid dynamics* (Cambridge University Press, 2000).

[76] E. Lauga and T. R. Powers, Rep. Prog. Phys. **72**, 096601 (2009).

[77] E. Guazzelli and J. F. Morris, *A physical introduction to suspension dynamics* (Cambridge University Press, 2012).

[78] J. M. Yeomans, D. O. Pushkin, and H. Shum, Eur. Phys. J. Spec. Top. **223**, 1771 (2014).

[79] E. M. Purcell, Am. J. Phys. **45**, 3 (1977).

[80] A. Shapere and F. Wilczek, J. Fluid Mech. **198**, 557 (1989).

[81] H. C. Berg, *E. coli in Motion* (Springer, 2004).

[82] J. D. Jackson, Classical electrodynamics (Wiley-VCH Verlag GmbH & Co. KGaA, 2003).

[83] S. Kim and S. J. Karrila, *Microhydrodynamics* (Butterworth-Heinemann, 1991).

[84] G. J. Hancock, Proc. R. Soc. Lond. **217**, 96 (1953).

[85] G. K. Batchelor, J. Fluid Mech. **41**, 545 (1970).

[86] A. T. Chwang and T. Y.-T. Wu, J. Fluid Mech. **67**, 787 (1975).

[87] K. Drescher, J. Dunkel, L. H. Cisneros, S. Ganguly, and R. E. Goldstein, Proc. Natl. Acad. Sci. U. S. A. **108**, 10940 (2011).

[88] D. C. Guell, H. Brenner, R. B. Frankel, and H. Hartman, J. Theor. Biol. **135**, 525 (1988).

[89] G. B. Jeffery, Proc. R. Soc. Lond. A Math. Phys. Sci. **102**, 161 (1922).

[90] G. Subramanian and D. L. Koch, J. Fluid Mech. **632**, 359 (2009).

[91] D. Saintillan and M. J. Shelley, Phys. Rev. Lett. **100**, 178103 (2008).

[92] C. Hohenegger and M. J. Shelley, Phys. Rev. E **81**, 046311 (2010).

[93] J. Stenhammar, C. Nardini, R. W. Nash, D. Marenduzzo, and A. Morozov, Phys. Rev. Lett. **119** (2017).

[94] V. Škultéty, D. Bárdfalvy, J. Stenhammar, C. Nardini, and A. Morozov, J. Fluid Mech. **980** (2024).

[95] V. Škultéty, C. Nardini, J. Stenhammar, D. Marenduzzo, and A. Morozov, Phys. Rev. X. **10** (2020).

[96] N. N. Bogoliubov, Journal of Physics (USSR) **10**, 265 (1946).

[97] M. Born and H. S. Green, Proc. R. Soc. Lond. A Math. Phys. Sci. **188**, 10 (1946).

[98] J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946).

[99] J. G. Kirkwood, J. Chem. Phys. **15**, 72 (1947).

[100] Y. Jacques and R. Yves, *La théorie statistique des fluides et l'équation d'état* (Hermann Cie, éditeurs, 1935).

[101] V. A. Martinez, E. Clément, J. Arlt, C. Douarche, A. Dawson, J. Schwarz-Linek, A. K. Creppy, V. Škultéty, A. N. Morozov, H. Auradou, and W. C. K. Poon, Proc. Natl. Acad. Sci. U. S. A. **117**, 2326 (2020).

[102] R. Aditi Simha and S. Ramaswamy, Phys. Rev. Lett. **89**, 058101 (2002).

[103] P. G. de Gennes, *The physics of liquid crystals*, 2nd ed. (Clarendon Press, 1994).

[104] A. N. Beris and B. J. Edwards, *Thermodynamics of flowing systems: With internal microstructure* (Oxford University Press, 1993).

[105] S. P. Thampi and J. M. Yeomans, Eur. Phys. J. Spec. Top. **225**, 651 (2016).

[106] I. El Korde, D. Bárdfalvy, J. M. Lewis, A. Morozov, C. Nardini, and J. Stenhammar, Phys. Rev. Lett. **135**, 098302 (2025).

[107] M. C. Sukop and D. T. Thorne, Jr, *Lattice Boltzmann modeling* (Springer Berlin Heidelberg, 2006).

[108] P. L. Bhatnagar, E. P. Gross, and M. Krook, Phys. Rev. **94**, 511 (1954).

[109] T. Krüger, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva, and E. M. Viggen, \emph \bibinfo title The Lattice Boltzmann Method: Principles and Practice (Springer International Publishing, Cham, 2017).

[110] X. Shan and H. Chen, Phys. Rev. E **47**, 1815 (1993).

[111] K. N. Premnath and J. Abraham, Phys. Rev. E **71**, 056706 (2005).

[112] Z. Guo, C. Zheng, and B. Shi, Phys. Rev. E **65**, 046308 (2002).

[113] R. W. Nash, R. Adhikari, and M. E. Cates, Phys. Rev. E 77, 026709 (2008).

[114] C. S. Peskin, Acta Numer. **11**, 479 (2002).

[115] N. Kumar, H. Soni, S. Ramaswamy, and A. K. Sood, Nat. Commun. **5**, 4688 (2014).

[116] H. Chaté, F. Ginelli, and R. Montagne, Phys. Rev. Lett. **96**, 180602 (2006).

[117] S. Shankar, S. Ramaswamy, and M. C. Marchetti, Phys. Rev. E. **97** (2018).

[118] V. Škultéty, *Collective motion in microswimmer suspensions*, \bibinfo type PhD Thesis, The University of Edinburgh (2023).

[119] J. Gachelin, A. Rousselet, A. Lindner, and E. Clement, New J. Phys. **16**, 025003 (2014).

[120] H. C. Berg, Curr. Biol. **15**, R599 (2005).

[121] A. Opathalage, M. M. Norton, M. P. N. Juniper, B. Langeslay, S. A. Aghvami, S. Fraden, and Z. Dogic, Proc. Natl. Acad. Sci. U. S. A. **116**, 4788 (2019).

[122] L. Ohm and M. J. Shelley, J. Fluid Mech. **942** (2022).

[123] H. Deschout and K. Braeckmans, Single particle tracking, in \emph \bibinfo booktitle Encyclopedia of Biophysics, edited by G. C. K. Roberts (Springer Berlin Heidelberg, 2013) pp. 2326–2327.

[124] G. G. Yin and Q. Zhang, *Continuous-time Markov chains and applications*, 2nd ed. (Springer, 2012).

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

[126] E. Amselem, B. Broadwater, T. Hävermark, M. Johansson, and J. Elf, Nat. Commun. **14**, 1336 (2023).

[127] Y. Matsuda, I. Hanasaki, R. Iwao, H. Yamaguchi, and T. Niimi, Phys. Chem. Chem. Phys. **20**, 24099 (2018).

[128] P. Billingsley, *Probability and Measure*, 2nd ed. (John Wiley & Sons, 1986).

[129] C. M., *Pattern Recognition and Machine Learning*, 1st ed. (Springer, 2006).

[130] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, 2nd ed. (Springer, 2009).

[131] L. Schrangl, V. Mühlgrabner, R. Platzer, F. Kellner, J. Wieland, R. Obst, J. L. Toca-Herrera, J. B. Huppa, G. J. Schütz, and J. Göhring, Biomolecules **14**, 1001 (2024).