

## REVIEW • OPEN ACCESS

# Learning dynamical models of single and collective cell migration: a review

To cite this article: David B Brückner and Chase P Broedersz 2024 *Rep. Prog. Phys.* **87** 056601

View the [article online](#) for updates and enhancements.

## You may also like

- [Substrate curvature regulates cell migration](#)  
Xiaoxiu He and Yi Jiang

- [Spatio-temporal analysis of collective migration \*in vivo\* by particle image velocimetry](#)  
María F Sampedro, Gastón L Miño, Carolina D Galetto et al.

- [3D-printed insert-array and 3D-coculture-array for high-throughput screening of cell migration and application to study molecular and cellular influences](#)  
Brian E Grottkau, Zhixin Hui, Chuan Ye et al.

The advertisement features a red header with the company's website and email. The main title 'Instruments for Advanced Science' is displayed over a background image of various scientific instruments and a satellite in space. Below the title, a subtitle mentions mass spectrometers for vacuum, gas, plasma and surface science. Three main sections are shown: 'Residual Gas Analysis' with an image of RGA components and a Mars-like planet; 'Thin Film Surface Analysis' with an image of a SIMS system; and 'Plasma Characterisation' with an image of plasma processing equipment and a red sphere. Each section includes a brief description and a detailed paragraph about the specific technology.

www.hidenanalytical.com  
info@hiden.co.uk

## HIDEN ANALYTICAL

# Instruments for Advanced Science

Mass spectrometers for vacuum, gas, plasma and surface science

### Residual Gas Analysis

Perform RGA at UHV/XHV. Our RGA configurations include systems for UHV science applications including temperature-programmed desorption and electron/photon stimulated desorption.

### Thin Film Surface Analysis

Conduct both static and dynamic SIMS analysis with a choice of primary ions for full chemical composition and depth profiling. Our SIMS solutions include complete workstations and bolt-on modules.

### Plasma Characterisation

Fully characterise a range of plasmas: RF, DC, ECR and pulsed plasmas, including neutrals and neutral radicals. Extend your analyses to atmospheric pressure processes using the HPR-60, with time-resolved mass/energy analysis.

[www.HidenAnalytical.com](http://www.HidenAnalytical.com)   [info@hiden.co.uk](mailto:info@hiden.co.uk)

## Review

# Learning dynamical models of single and collective cell migration: a review

David B Brückner<sup>1,\*</sup>  and Chase P Broedersz<sup>2,3,\*</sup> 

<sup>1</sup> Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria

<sup>2</sup> Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

<sup>3</sup> Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department of Physics, Ludwig-Maximilian-University Munich, Theresienstr. 37, D-80333 Munich, Germany

E-mail: [david.brueckner@ist.ac.at](mailto:david.brueckner@ist.ac.at) and [c.p.broedersz@vu.nl](mailto:c.p.broedersz@vu.nl)

Received 7 October 2023, revised 14 February 2024

Accepted for publication 22 March 2024

Published 4 April 2024

Corresponding editor: Dr Erwin Frey



## Abstract

Single and collective cell migration are fundamental processes critical for physiological phenomena ranging from embryonic development and immune response to wound healing and cancer metastasis. To understand cell migration from a physical perspective, a broad variety of models for the underlying physical mechanisms that govern cell motility have been developed. A key challenge in the development of such models is how to connect them to experimental observations, which often exhibit complex stochastic behaviours. In this review, we discuss recent advances in data-driven theoretical approaches that directly connect with experimental data to infer dynamical models of stochastic cell migration. Leveraging advances in nanofabrication, image analysis, and tracking technology, experimental studies now provide unprecedented large datasets on cellular dynamics. In parallel, theoretical efforts have been directed towards integrating such datasets into physical models from the single cell to the tissue scale with the aim of conceptualising the emergent behaviour of cells. We first review how this inference problem has been addressed in both freely migrating and confined cells. Next, we discuss why these dynamics typically take the form of underdamped stochastic equations of motion, and how such equations can be inferred from data. We then review applications of data-driven inference and machine learning approaches to heterogeneity in cell behaviour, subcellular degrees of freedom, and to the collective dynamics of multicellular systems. Across these applications, we emphasise how data-driven methods can be integrated with physical

\* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

active matter models of migrating cells, and help reveal how underlying molecular mechanisms control cell behaviour. Together, these data-driven approaches are a promising avenue for building physical models of cell migration directly from experimental data, and for providing conceptual links between different length-scales of description.

**Keywords:** cell migration, inference, data-driven models, machine learning, active matter, stochastic dynamics, collective phenomena

## Contents

1. Introduction	2
2. Top-down and bottom-up models of cell migration	3
3. Learning the stochastic dynamics of single cell migration	5
3.1. An equation of motion for freely migrating cells	5
3.2. Experimental approaches for cell migration in structured environments	8
3.3. Dynamical models of confined cell migration	10
3.4. Why do cell migration dynamics appear to be underdamped?	11
4. Learning equations of motion from stochastic trajectories	12
4.1. General principles	13
4.2. Sources of inference error	14
5. Inferring heterogeneity in cell behaviour	16
5.1. Quantifying temporal and cell-to-cell variability in behaviour	17
5.2. Identifying sources of environmental heterogeneity	19
6. Connecting cell dynamics to mechanisms	19
6.1. Bottom-up models for cell migration	20
6.2. Inference from cellular features	22
7. Learning the collective dynamics of multicellular systems	23
7.1. Inference approaches for interacting active systems	24
7.2. From cell pairs to collective migration	25
8. Outlook	28
Data availability statement	29
Acknowledgments	29
References	29

## 1. Introduction

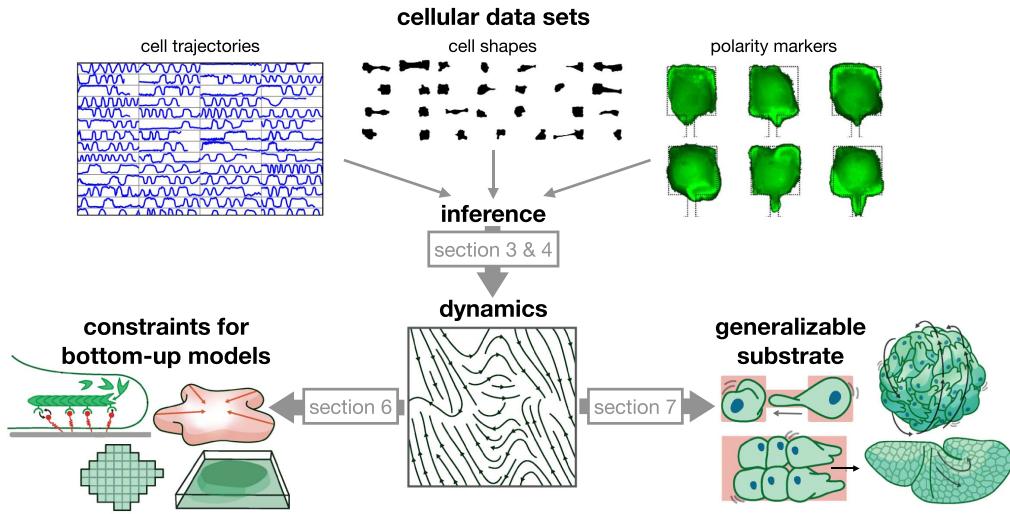
The vast majority of cells in our body do not move around—but when they do, it is for an important reason: migrating cells shape you, they can protect you, but also harm or even kill you. In development, cells actively migrate to be at the right place at the right time, shaping the early embryo [1, 2]. Later on, while most cells become sedentary, immune cells have a remarkable ability to migrate through the tightest pores to hunt down pathogens, protecting you from diseases [3]. Furthermore, all cells retain the ability to switch to a migratory mode, allowing

them to efficiently close wounds [4, 5]. However, this ability is hijacked by cancer cells, which migrate during metastasis to spread to other organs [6–8].

The underlying processes required to make a cell move are determined by a broad variety of physical phenomena: the polymer physics of cytoskeletal filaments [9–12], the reaction–diffusion dynamics of signalling molecules [13, 14], and the active mechanics of acto-myosin contraction [15–17]. The cellular motility machinery integrates these physical processes to push forward the cell membrane, giving rise to overall motion of the cell. Much of this machinery is highly conserved across organisms and tissues [18], giving hope that understanding the physics of these processes will lead to a general understanding of cell motility. However, while much progress has been made in understanding each of these biophysical aspects, how these integrate to generate behaviours at the scale of the cell as a whole remains the subject of current research.

An exciting perspective is therefore whether physics can go beyond explaining the physical components of cellular systems and provide conceptual and predictive frameworks to describe the emergent behaviour of cells as a whole. To accomplish this, we need to connect physical modelling approaches across scales and understand how they interplay at the system level. To achieve such connections in systems with such daunting inherent complexity, data-driven theoretical approaches that connect directly to experimental data are emerging as a fruitful and promising avenue. Put simply, such data-driven approaches aim to solve the inverse problem of determining an effective physical description of a system from data. Indeed, in recent years, a number of studies have started developing data-driven approaches to learn dynamical models of stochastic cell migration directly from experimental data. This includes a wide variety of inference approaches using stochastic inference, machine learning and dimensional reduction to infer how cells interact with their environment and with each other. This field is currently at a unique crossroad: due to advances in nanofabrication, image analysis and tracking technology, experimental studies now yield unprecedented large data sets on cellular phenotypes; and at the same time, there is an increasing pivot among theoreticians to interact directly with experimental data and apply tools such as machine learning and physics-guided inference approaches to learn from data.

In this article, we take stock of these recent advances and the outstanding challenges in learning dynamical models of the stochastic behaviour of single and collective cell migration



**Figure 1.** Conceptual approach of learning data-driven models from cellular data sets. Cellular data sets such as cell trajectories, cell shapes or intracellular markers serve as input to model inference (here shown for the example of a confined cell [19, 20]). This provides a dynamical systems representation of behaviour, providing constraints for bottom-up models, and a generalisable basis for more complex systems.

directly from experimental data (figure 1). A key challenge for these approaches is to connect to more classical biophysics models of cell migration, including soft matter, hydrodynamic, and mechanical theories. These include mechanistic models at the single cell level (active gel theory, phase field models, cellular Potts models); and active matter models at the collective scale (active hydrodynamics, active particle models, vertex models). We discuss how inference from data can be connected and integrated with these physical approaches, and how it may provide bridges to connect diverse modelling approaches into a coherent framework for cell migration. As the diversity, accuracy, dimensionality, and size of these cellular datasets is rapidly increasing, we expect such data-driven approaches to play an increasingly important role in building physical models of single and collective cell behaviour.

First, we provide a perspective of how we envision learning cell dynamics at the behavioural level will advance our understanding of cell migration (section 2). We then review data-driven approaches to learn the dynamics of single migrating cells (section 3), and summarise the technical aspects of performing stochastic inference from cell trajectories (section 4). Furthermore, we will review how these approaches have been extended to give insight into the variability of cell behaviours in time and across individuals (section 5). In section 6, we provide a perspective on how data-driven approaches to emergent cell dynamics can be connected to underlying molecular mechanisms. Zooming out from the single cell level, we then review data-driven approaches to describe the interactions between cells (section 7). These examples demonstrate how combining advances in physical modelling, inference methods, and high-throughput experimental approaches can help reveal the underlying physics of what makes a cell move.

## 2. Top-down and bottom-up models of cell migration

In multicellular organisms, individual cells migrate to execute functional tasks. Thus, cells are programmed to perform certain behaviours, including for example net motion through an external environment (migration), changes in cell shape (morphodynamics), exerting forces on the extra-cellular environment (traction forces), adaptation to external signals (stimulus response), or the degradation of surrounding matrix polymers (proteolysis). What all these examples have in common, is that they are performed at the length-scale of the whole cell and often take place on long time-scales. Here, we refer to ‘long time-scales’ as those time intervals on which a cell-scale behaviour takes place (for many cell types on the time-scales of hours [19]), which are typically much larger than the typical time-scales of intra-cellular processes (often on the time-scales of seconds to minutes), such as the polymerisation rate of a single actin filament ( $\approx 0.2 \mu\text{m s}^{-1}$  [21]) or the lifetime of a single focal adhesion ( $\approx 20 \text{ min}$  [22]). On these cell- and behaviour-level length- and time-scales, cell behaviour emerges as a consequence of a large number of intra-cellular processes operating simultaneously. From the point of view of physical models of cell migration, this complexity means that molecularly reductionistic approaches are very challenging [23]: precise knowledge of one or several particular signalling processes and all the associated parameters may not be predictive for the whole-cell behaviour. This is because whole-cell behaviours integrate many processes, and developing quantitative models for all of these processes at once is unfeasible.

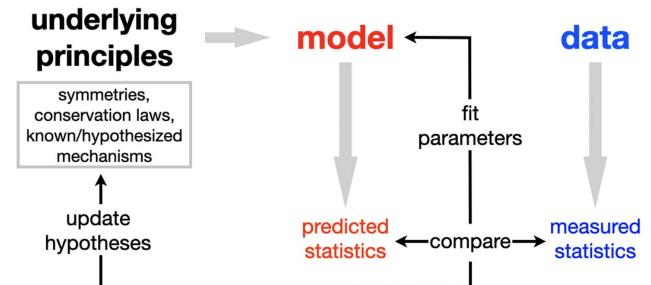
To circumvent this problem, minimal physical models are often employed, which seek to identify the key mechanisms

at play and integrate them into quantitative, predictive model. We refer to these approaches as **bottom-up models** in this review, since these approaches start by postulating a set of rules to describe the various components of a cell and then seek to predict the emerging behaviour (figure 2(A)) [24]. These models are typically compared to data by fitting specific parameters suggested by the model by comparing predicted and experimentally measured statistics of the process. Examples for such bottom-up models are cellular Potts models [25–33], phase field models [34–39], as well as the molecular clutch model [40, 41], active gel theories [42–50] and models coupling actin flow, polarity cues, and focal adhesion dynamics [13, 14, 51–58]. We will review these approaches in more detail in section 6. However, applying these types of models directly to experimental observations is challenging: depending on the implementation, these models may have many parameters that are difficult to constrain based on experimental data. To avoid this, models are frequently tailored to capture a particular aspect of the data, but it has often remained difficult to capture the full long time-scale dynamics of the cells, or how these dynamics adapt to external inputs.

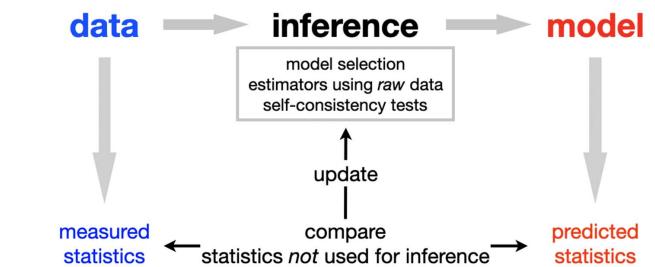
An alternative to mechanistic models are data-driven **top-down approaches**, which systematically constrain model candidates using experimental data (figure 2(B)) [24]. Naturally, top-down approaches tend to provide a more phenomenological description of the system since they are typically based on experimental data at the cellular or tissue scale rather than the molecular scale. An example of a top-down approach are models inferred directly from measured cell migration trajectories. The resulting phenomenological description based on such data therefore effectively coarse-grains over the molecular detail. Generally speaking, phenomenological theories in physics have often generated conceptual understanding that remained elusive in the reductionist approach, an idea that was famously articulated by Phil Anderson in his essay ‘More is different’ [23]. Indeed, different levels of description can be relevant at different time- and length-scales, suggesting that the molecularly reductionistic approach is not the only way of modelling a system, but phenomenological descriptions could be very helpful at the large time- and length-scales of cellular behaviours. Following this philosophy, we argue that top-down approaches are a promising direction to develop quantitative frameworks for cell behaviour. We argue that these top-down approaches should ideally have the following properties:

- **Data-driven:** to provide a phenomenological description of a cellular system without reference to specific molecular processes, top-down approaches need to be constrained by experimental data. An important aspect of a data-driven approach is the degree of ‘rawness’ of the data that is used for inference. An approach that uses essentially the full range of spatial and temporal information that characterises a process—such as the full trajectory of a migrating cell—will be more powerful at distinguishing models than an approach that uses processed statistics derived from the data—such as fitting a correlation function. Such inference from raw data requires both a sufficiently general inference

## A Bottom-up models



## B Top-down approaches



**Figure 2.** Bottom-up vs top-down approaches for model development. (A) Bottom-up approaches start from postulated underlying physical principles, such as symmetries, conservation laws or specific known or hypothesised mechanisms. Based on these principles, a mathematical model is formulated, which predicts certain features of systems quantified by statistics such as the mean-square-displacement (MSD) of a cell. To fit the model parameters, a fitting procedure that compares predicted and experimentally measured statistics is used. (B) Top-down approaches use experimental data as a starting point. These data are passed to an inference or learning algorithm, which performs model selection and estimation of dynamical terms directly from experimental data. The resulting inferred model is then tested by comparing measured and predicted statistics that have not been used in the inference procedure. Reproduced with permission from [24].

techniques as well as high-quality quantitative datasets of cell behaviours.

- **Unbiased and model-agnostic:** a central idea in top-down approaches is that they should be agnostic with respect to the underlying molecular or mechanistic basis of the behaviour. Thus, the starting point of an inference approach should be a sufficiently general class of different models, and the inference should select the most likely model based on the available data. This aspect distinguishes inference or learning approaches from parameter fitting of a bottom-up model. In parameter fitting, the quantitative values of parameters of a single model candidate are determined. Inference approaches should be able to distinguish between multiple qualitatively different possible models.
- **Predictive:** while a given model may be constrained using data, it should then also be able to predict new observations beyond the data that were used to constrain it. Tests of predictive power have two distinct roles: firstly, making predictions for the same experimental data set used to constrain the model, but for statistics that were not explicitly

used in the inference, allows testing whether the model provides a meaningful representation of the cellular behaviour. Secondly, performing predictions for new experiments tests the usefulness of the model to provide a generalisable basis for new systems.

There has been a recent surge of activity in developing such data-driven, unbiased approaches in a number of other biological systems across scales, including protein folding [59, 60], chromosome organisation [61–63] and dynamics [64–66], neural systems [67–70] and animal behaviour [71–75]. In recent years, the abundance of data-sets of dynamical systems across the disciplines has been rapidly increasing. This has led to the development of data-driven methods to infer the underlying dynamics of complex systems directly from experimental data, including deterministic [76–80] and stochastic trajectories [75, 81–91], spatially extended fields [92, 93], and morphological dynamics [72–74]. We argue that there are four key challenges that inference approaches for cell behaviour could help address, based on which we organise the structure of this review (figure 1):

- (1) Owing to the intrinsic stochasticity and variability of cell behaviours, a key challenge is to identify what constitutes a ‘typical’ behaviour. Data-driven approaches could provide **analysis tools** for unbiased, quantitative characterisation, classification, and observation of cellular behaviours. Examples for cellular readouts are cell persistence in two-dimensional (2D) migration [13, 94], transition times and occupancy probabilities in confined migration [19, 20], movement biases in directional migration [95], and collision outcomes in cell–cell interactions [96–98]. We discuss how data-driven approaches can provide readouts of typical behaviours (section 3) and of their variability (section 5).
- (2) Due to the emergent nature and underlying complexity of cell behaviours, it is often unclear what the right quantitative concepts are to describe a particular observed behaviour. Data-driven approaches could yield **conceptual frameworks** to think about cell behaviours by identifying underlying quantitative concepts that can be used to describe cell dynamics. Examples for such concepts in the context of freely migrating cells on 2D substrates are the persistent random motion model [94, 99, 100], Lévy flights [101], and intermittent dynamics [13]. We will review these models and their biological implications in section 3, and discuss methods for model inference more generally in section 4.
- (3) Phenomenological models which are constrained in an unbiased and data-driven manner could furthermore yield strong **constraints for bottom-up models** for the underlying mechanistic basis of the behaviour. These mechanistic models come in different flavours, from minimal mechanical models to active polar gel theories and complex computational implementations. A central difficulty in connecting these models to experiments is that they are frequently under-constrained and over-parametrised. Phenomenological descriptions could provide much more

precise ‘targets’ for mechanistic approaches by introducing stronger constraints. Furthermore, they could be used to test conceptual modelling assumptions or approximations, and thus give insight into the key biological processes in a given system. We will discuss this connection in section 6.

- (4) Finally, data-driven frameworks may provide **systematic frameworks** to address increasingly complex questions, making it possible to add complexity step-by-step. For example, to describe the dynamics of interacting cells, it may be useful to have a theory for the dynamics of single migrating cells. We will discuss how data-driven approaches can help quantify the behavioural variability of migrating cells in section 5 and identify models for cell–cell interactions section 7.

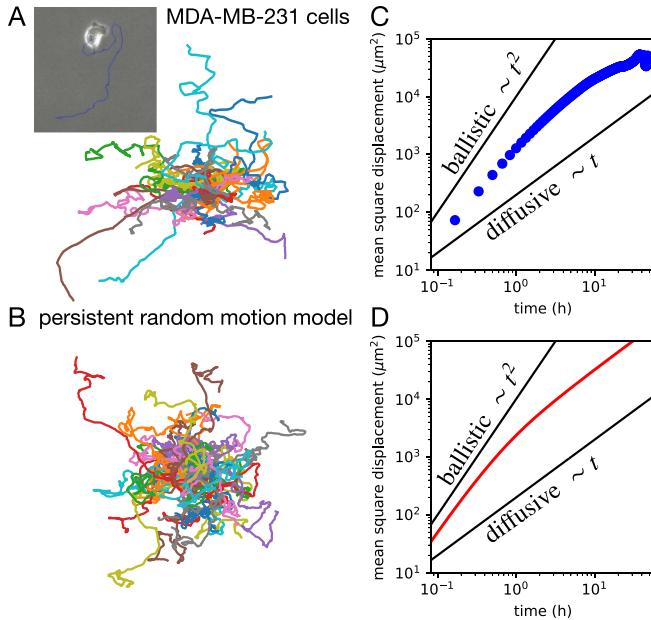
### 3. Learning the stochastic dynamics of single cell migration

#### 3.1. An equation of motion for freely migrating cells

The simplest possible experiment that could teach us something about cell migration behaviour is perhaps the motion of isolated single cells on a uniform 2D substrate. This is of course not a common setting in physiological processes, in which cells typically encounter heterogeneous, confining three-dimensional (3D) environments—yet it is the archetypal cell migration experiment that has taught us much of what we know about migrating cells. We will turn our attention towards the description of systems that include spatial structures in the next section. Here, we will review what we have learnt from 2D cell migration, and how this may provide a generalisable basis to describe more complex systems.

Even in the simple environment of a uniform 2D substrate, the migration of single cells is non-trivial, as it is powered by a complex cytoskeletal assembly. To study migrating cells, a natural avenue is to focus on the underlying biochemical and biophysical mechanisms, and the molecular pathways underlying them. For this endeavour, the simple scenario of free 2D cell migration was key, and the insights gained have been reviewed elsewhere [102]. An alternative, however, is to zoom out from the molecular level to the behaviour at the cellular scale and to measure the overall motion of the cell. Characterising these system-level dynamics could then teach us about typical behaviours of cells, which may eventually help understand how such emergent behaviours are generated by the underlying molecular players.

A simple way to quantify the dynamics of migrating cells is a reduction to a single variable: the position of the cell as a function of time, i.e. the trajectory  $\mathbf{x}(t)$  of its nucleus or centroid (figure 3(A)). The first cell tracking experiments were performed over a century ago [103, 104] (see [100] for a review). At this level, all other putative cellular degrees of freedom (DOFs), such as the cell shape, cytoskeletal organisation, and traction forces, remain unobserved. The trajectory of the cell is thus a minimal representation of a behaviour: it is observed at the cellular scale, and over long time-periods compared to the time-scales of the internal dynamics. One of



**Figure 3.** Free 2D cell migration. (A), (B) 2D trajectories of single migrating MDA-MB-231 cells and of simulated cells based on the persistent random motion model, respectively. Inset in (A): brightfield microscopy image of a migrating MDA-MB-231 cell. Data from [19]. (C), (D) mean square displacement curves calculated from experiment and the persistent random motion model (equation (3)), respectively. Black lines indicate the limits of ballistic and diffusive motion.

the simplest statistics often derived from stochastic trajectories is the mean-square-displacement (MSD), which in many scenarios follows a power-law:

$$\langle [\mathbf{x}(t) - \mathbf{x}(0)]^2 \rangle \propto t^\alpha. \quad (1)$$

The exponent qualifies the type of random motion observed:  $\alpha < 1$  corresponds to subdiffusive,  $\alpha = 1$  to diffusive,  $1 < \alpha < 2$  to superdiffusive and  $\alpha = 2$  to ballistic motion. Some of the earliest cell tracking experiments by Fürth in 1917 revealed that the active motion of protozoa exhibit a mix of deterministic trends, visible as persistent segments, and seemingly random, stochastic components [104]. By considering the continuum limit of a minimal model of a persistent random walker on a lattice (figure 3(B)), he showed that such behaviour predicts an MSD of the form [100, 105]

$$\langle [\mathbf{x}(t) - \mathbf{x}(0)]^2 \rangle = A \left( t/\tau_p + e^{-t/\tau_p} - 1 \right). \quad (2)$$

Thus, the MSD exhibits signatures of ballistic motion ( $\alpha = 2$ ) at short time-scales and diffusive motion ( $\alpha = 1$ ) at long time-scales (figure 3(D)). Indeed, equation (2) provides a good first approximation to the MSD of cell migration trajectories (figure 3(C)) and has been used to fit a broad range of cell types [99, 105, 106]. More recent measurements indicate more complex behaviours including superdiffusive behaviour with fractional diffusion exponents  $1 < \alpha < 2$  [101, 107–111], which we discuss further below.

By measuring the MSD, one can recover two key parameters that characterise the behaviour: the persistence time

$\tau_p$ , which quantifies the time over which correlations in the cell velocity decay, and the diffusion coefficient  $D = A/2d\tau_p$ , where  $d$  is the dimensionality. These parameters are frequently used to quantify cell migration, for example to determine the effect of pharmacological treatments of cells, or to contrast different cell types. However, only measuring the MSD is not sufficient to determine a model of cell migration: many underlying mechanisms can give rise to the same MSD [111], suggesting that additional statistics are required to determine a model from data. Ideally, we would like to obtain an equation of motion of the cell that predicts all features of the trajectories. To constrain such an equation of motion model, we must therefore use additional information contained in the trajectories  $\mathbf{x}(t)$  than just the MSD.

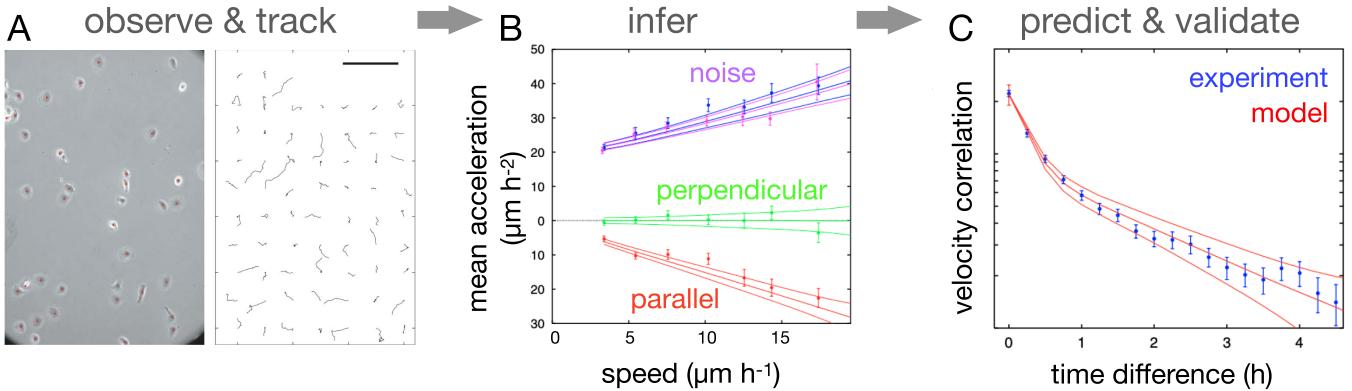
The appropriate mathematical framework to think about such equations of motion are stochastic differential equations. A simple model that predicts an MSD of the form of equation (2) is an equation of motion for the cell velocity  $\mathbf{v} = d\mathbf{x}/dt$ , the *persistent random motion model*:

$$\frac{d\mathbf{v}}{dt} = -\frac{1}{\tau_p} \mathbf{v} + \sigma \boldsymbol{\eta}(t) \quad (3)$$

where  $\boldsymbol{\eta}(t)$  is a 2D Gaussian white noise, with  $\langle \boldsymbol{\eta}_i(t) \rangle = 0$  and  $\langle \boldsymbol{\eta}_i(t) \boldsymbol{\eta}_j(t') \rangle = \delta_{ij}\delta(t-t')$ , where  $\delta_{ij}$  is the Kronecker delta, and  $\delta(t-t')$  is the Dirac delta function. This equation of motion predicts the cell acceleration as a function of its velocity and generates trajectories similar to those observed in experiments (figures 3(A) and (B)). It consists of two components: a deterministic contribution (first term on the right-hand side), which accounts for the cell persistence, and a Gaussian white noise term (second term on the right-hand side), which accounts for the stochasticity of the motion. This equation predicts the MSD in equation (2) with  $A = 2\sigma^2\tau_p^2$ . However, equation (3) also predicts many other features of the trajectory dynamics. Specifically, it predicts a Gaussian steady state probability distribution of velocities  $p(\mathbf{v})$  with a variance  $\tau_p\sigma^2/2$ , and a velocity auto-correlation function  $\langle \mathbf{v}(t)\mathbf{v}(t') \rangle$  that decays as a single exponential with a time-scale  $\tau_p$ . Furthermore, it makes a specific prediction about the *conditional average* of the observed cellular accelerations, i.e. the average of the instantaneous acceleration for each observed instantaneous velocity:

$$\left\langle \frac{\Delta \mathbf{v}}{\Delta t} \middle| v \right\rangle \approx -\frac{1}{\tau_p} v. \quad (4)$$

This relation is not exact, but approximate, since it neglects the effects of discretisation: the derivatives  $v = \Delta x/\Delta t$  and  $\Delta v/\Delta t$  typically cannot be measured exactly, but are estimated through numerical differentiation of the position trajectories  $x(t)$ . This leads to non-trivial discretisation effects [86, 112], which we neglect in equation (4) and discuss in detail in section 4 (see equation (14)). These additional statistics provided by equation (4) can thus be used to systematically constrain models for 2D cell migration in a data-driven manner. For example, calculating the conditional average on the left-hand side in equation (4) can constrain the deterministic term of the description: in principle, the dependence



**Figure 4.** Inference of a dynamical model of 2D cell migration. (A) Observation and tracking step: microscopy image of human dermal keratinocytes (HaCaT) and corresponding nucleus trajectories (scale bar: 200  $\mu\text{m}$ ). (B) Inference step: mean and standard deviation as function of speed of the following quantities: average components (calculated using the left hand side of equation (4)) parallel (red) and perpendicular (green) to the direction of motion, and stochastic components parallel (blue) and perpendicular (magenta), providing an estimate of  $\sigma(v)$ . Solid curves show the same quantities, plus/minus one standard deviation, calculated from the inferred model (equation (5)). (C) Prediction and validation step: experimental (blue) and predicted (red) velocity auto-correlation function. Reproduced with permission from [100]. Copyright © 2008, EDP Sciences/Società Italiana di Fisica/Springer-Verlag. CC BY-NC 2.0.

of acceleration on velocity could be non-linear, and this analysis would reveal such an effect in a model-independent manner. Similarly, the magnitude of the stochastic noise term  $\sigma$  can be inferred from the variance of the fluctuations in the trajectories.

As our measurements of cell trajectories have become increasingly accurate and computer-based tracking has allowed generating large sets of such data, a number of statistical features that are not predicted by the persistent random motion model (equation (3)) have been identified. Specifically, the velocity distributions of cells are typically not Gaussian, but exhibit exponential tails [113, 114] and the velocity auto-correlation is not exponential, but typically bi-exponential [94]. To build a model of free 2D migration that captures these anomalous features, a data-driven approach to learn an equation of cell motion directly from data was proposed by Selmeczi *et al* [94] (figure 4). For this, the conditional average of the acceleration (equation (4)) provides a strong constraint on the model (figure 4(B)). Based on this, the authors determined the simplest model consistent with all the observed statistics, which contains an additional memory term in the velocities. We therefore refer to it as the *persistent memory model*. Specifically, the authors identified the following equation of motion based on the data:

$$\frac{d\mathbf{v}}{dt} = -\beta(v)\mathbf{v} + \alpha^2 \int_{-\infty}^t dt' e^{-\gamma(t-t')} \mathbf{v}(t') + \sigma(v) \boldsymbol{\eta}(t) \quad (5)$$

where the multiplicative noise  $\sigma(v)$  is interpreted in the Itô sense [115]. Here, the first term provides a (speed-dependent) time-scale  $\tau = \beta^{-1}(v)$  on which the velocity fluctuates around zero, like in the persistent random motion model (equation (3)). The second term is a memory kernel, which depends on past velocities with a memory time-scale  $\gamma^{-1}$ . These two time-scales then give rise to a bi-exponential velocity auto-correlation, as observed experimentally (figure 4(C)). Furthermore, this inferred model captures

various other anomalous statistics, including the non-Gaussian speed distribution. Similar results were subsequently also found in 2D migration of the amoeba *Dictyostelium* [100, 116–118] and breast cancer cells [19]. Notably, these various studies showed that while the overall form of equation (5) is conserved across cell types, the functions  $\beta(v)$  and  $\sigma(v)$  had qualitatively different shapes for different cell types.

The most remarkable feature of the persistent random motion and the persistent memory models for free 2D cell migration is the drastic reduction in complexity achieved. Small and fast dynamics of the cell contour appear as dynamical noise (not to be confused with technical noise or measurement error), and only a small number of parameters are necessary to accurately capture cell motion at the level of trajectories. The data-driven development of these models therefore formalises concepts such as persistence and cellular fluctuations. Indeed, an important step in the inference procedure was to disentangle the deterministic (average) and stochastic (fluctuating) components of the dynamics. Decomposing these two contributions is a key advantage of learning the stochastic equation of motion of the system, and then allows interpretation of each component.

Such data-driven, quantitative frameworks for 2D cell migration are useful in several ways. First, they provide a benchmark for characterising the behaviours of different cell types and determining the effects of drug treatments or genetic perturbations. Secondly, the structure of the inferred model can give insight into the underlying cell dynamics. Importantly, the memory kernel indicates that knowing the current state of motion (determined by the velocity  $\mathbf{v}(t)$  at time  $t$ ) is not enough to predict future cell motion, but the history of the process (up to a time-scale given by  $\gamma^{-1}$ ) also needs to be considered. This memory is presumably encoded in the polar structure of the cell, corresponding to unobserved associated variables that render the dynamics of cell position and velocity non-Markovian (see section 3.4). Importantly, determining equation (5) from the data yields a quantitative

description of how these latent variables affect cell motion. Thus, this description can now provide constraints for bottom-up models that seek to connect mechanisms to overall motion. We will discuss this avenue in more detail in section 6.

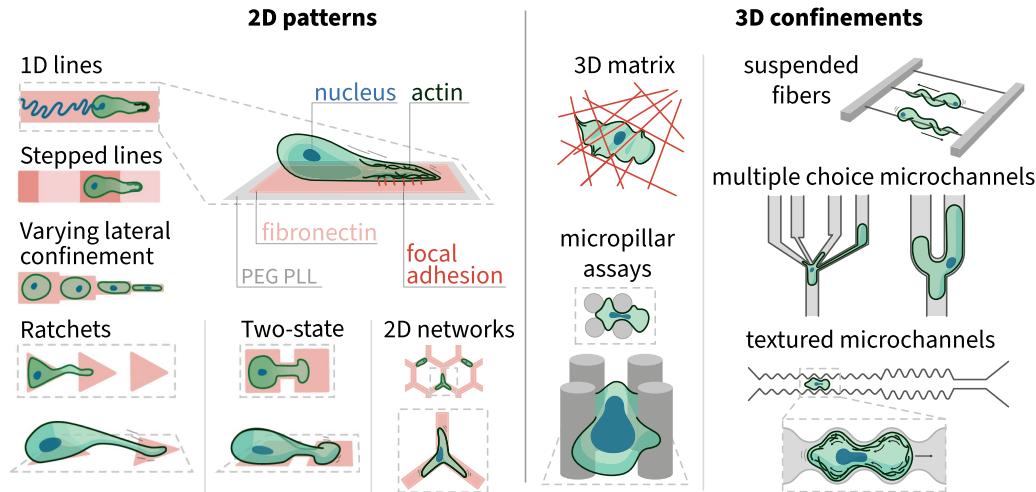
Alternatives to this persistent random memory framework also exist in the literature. These have primarily been motivated by the observation of fractional diffusion exponents  $1 < \alpha < 2$  of the MSD (equation (1)), indicating superdiffusive behaviour [101, 107–111]. Multiple alternative explanations for superdiffusive exponents have been put forward [111]: (1) cell-to-cell variability of cell motility parameters, leading to a distribution of cross-over time-scales in the ensemble-averaged MSD [109, 119, 120] (see section 5.1 for a detailed discussion). (2) Lévy walk models, i.e. run-and-tumble models with a power-law distribution of step lengths [121], which were shown to capture the dynamics of T-cell migration [101, 110]. (3) Fractional diffusion equations, i.e. replacing the Fokker–Planck formulation of equation (3) with a fractional Klein–Kramers equation containing fractional time-derivatives [107]. This highlights that the type of description may vary depending on the cell type, but also that multiple descriptions of the same data set may be possible, raising the need to explore connections between these descriptions and for principled inference approaches that can distinguish alternative scenarios. In the context of superdiffusive motion, an approach to distinguish hypotheses (1) and (2) has been proposed, which concluded that superdiffusive motion of fibroblasts was caused by cell-to-cell variability rather than Lévy walk behaviour [111]. Finally, stochastic equation of motion frameworks have also been extended to capture biased random walks for directional cell motion such as chemotaxis [122–124].

While the persistent random motion framework is intuitive and is frequently used to describe cell migration, the aim of the approach outlined in this section was to determine a dynamical model for single cell migration without such prior intuition, directly from data. More specifically, the aim was to learn an equation of motion from the stochastic cell trajectories. This places this work into a general class of inverse problems where the aim is to derive a physical description from data in an unbiased manner. This inference principle is a key technique whose full power becomes apparent when used with more general inference methods and on complex data sets. We are by no means constrained to infer cell acceleration as a function of velocity: what if the migration takes place in a complex structured environment? Then, other DOFs, such as the cell position, can be used as conditioning variables. We can therefore infer how cellular responses (measured in accelerations) depend upon the local geometry or structure of the environment (measured by position). We will discuss such an approach in section 3.2. Furthermore, we can imagine tracking other DOFs of the cell beyond its position, for example protrusions and retractions, or even spatially extended variables such as shape or internal concentration fields. Deriving the equations of motion of these DOFs, and their coupling to each other and to the environment, could yield key insights into cell behaviour. This approach could provide more direct

connections with mechanistic models (see section 6). Finally, new inference techniques also allow for inference in high-dimensional and interacting systems [85, 86], which could be used to learn the dynamics of interacting cells in collective migration (section 7). The data-driven persistent random motion framework introduced in the previous section establishes a conceptual basis to understand these other approaches, which become increasingly complex when we go beyond this simple stochastic process. Inferring an equation of cell motion based on experimental trajectories has helped to elevate persistent random cell motion from a concept into a theory, meaning that we progress from a somewhat fuzzy intuition to a mathematical equation that makes falsifiable predictions that can be tested on the data. We will highlight avenues for achieving something similar for more complex systems. For this, we will first turn to the example of a single cell migrating in a standardised structured environment, allowing inference of its interaction with external features. To enable going through such an example in detail, we will discuss a biased selection of the literature and focus on our work of learning the equation of motion of a cell confined in a two-state micro-pattern [19]. In the following sections, we will then discuss the much broader literature on inferring cell-to-cell variability, connecting to bottom-up models, and learning models of collective migration.

### 3.2. Experimental approaches for cell migration in structured environments

Cell migration on unstructured 2D substrates provides an important benchmark for how to think about cell migration dynamics, and its simplicity has allowed significant theoretical progress. However, in physiological processes, cells do not encounter such unstructured environments: they navigate extra-cellular environments that are complex, structured, and confining. These include collagen matrices, bone marrow, or blood vessel linings [6]. Thus, if we want to understand cellular dynamics in physiological processes, we need to study confined cell migration. Cell migration in 3D extra-cellular matrices (ECMs) has been studied extensively (see reviews in [125–127]). However, these matrices are spatially heterogeneous, and thus single cells will only rarely encounter the same obstacle twice. While some studies have made progress on quantifying cell trajectories through ECM [128] as well as bacterial motion through heterogeneous porous media [129, 130], it is in general difficult to gather sufficient statistics to understand how the local microstructure determines the cell behaviour. A popular approach to study confined migration while keeping the extra-cellular environment as simple as possible, are *in vitro* artificial confining geometries. Such geometrical confinements can be implemented using micropatterning, 3D printing, or microfluidics, and can be designed to expose cells to challenges such as overcoming a constriction or navigating a maze. Overcoming such challenges is an inherent feature in *in vivo* contexts, and is clearly an aspect that is missed by studying cells in featureless 2D surfaces. In



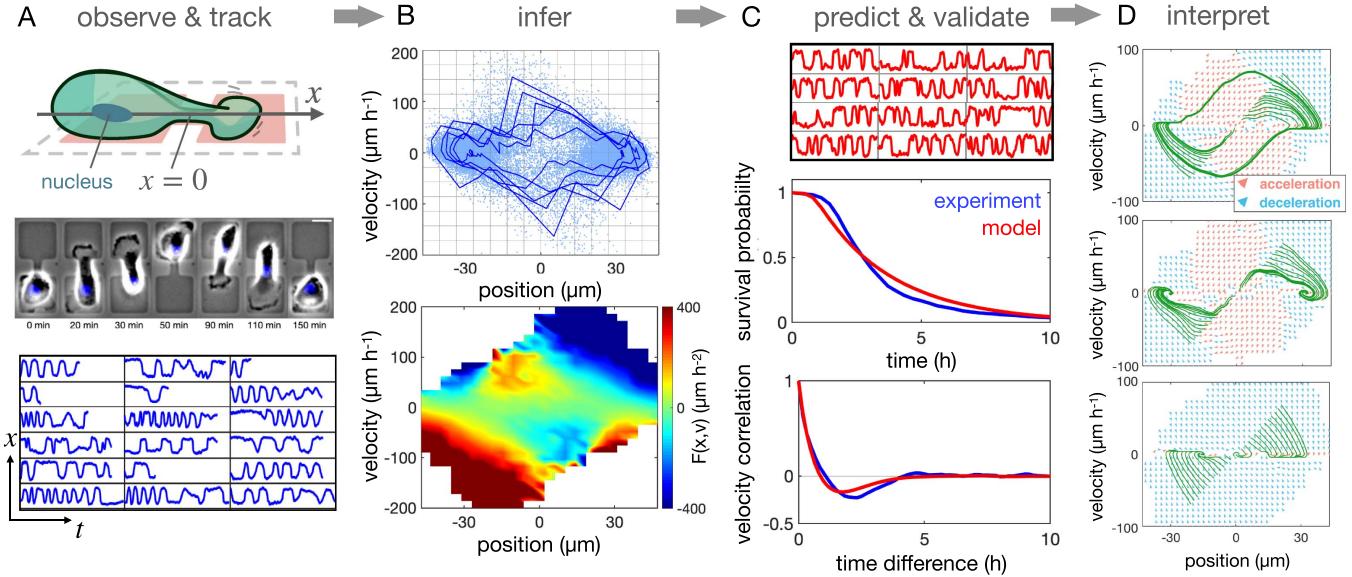
**Figure 5.** Experimental approaches for studying confined cell migration. 2D confining geometries of cells are typically designed using micropatterning, in which a region of defined geometry is coated with a cell-adhesive protein, fibronectin, while the surroundings are passivated with cell-repellent PEG-PLL polymers [135, 136]. To study cell migration, such micropatterns have been used in the shape of 1D lines [145, 146], stepped lines with varying protein coating density [58], varying lateral confinement [148], series of triangles in a ratchet-like arrangement [95, 149, 150], two-state micropatterns [19, 20], and 2D networks of 1D lines [151]. 3D confinements to study cell migration include 3D extracellular matrices, micropillar arrays, suspended fibres [142–144], multiple-choice microchannels [138, 139, 152], as well as textured microchannels [140]. In these 3D confining systems, there is not only basal, but also lateral confinement, causing among other things deformation of the cell nucleus when cells migrate through constrictions.

this section, we will give a brief overview over the key experimental approaches to study confined cells *in vitro*, before turning to inference approaches for confined migration. For in depth discussions of the experimental and technical aspects of artificial cellular confinements, please refer to reviews in [131–134].

Artificial systems to study confined migration include 2D micropatterns [135, 136], microfluidic devices [137], 3D confinements [138–140], micropillar arrays [141], and suspended nanofibres [142–144] (figure 5). These systems allow monitoring of large numbers of cells migrating in identical, standardised structured environments, yielding unprecedented large data sets on cell behaviour. Micropatterning provides a simple way to confine cells: using differential surface coatings, one can define areas to which cells can adhere, surrounded by cell-repellent regions. With this technique, confinements of arbitrary geometrical shape can be produced, giving access to a wide variety of systems. One of the simplest migration experiments using micropatterns is confinement to narrow stripes [145]. In such effectively one-dimensional (1D) confinements, cells typically perform persistent random motion in 1D [146]. This 1D mode of migration has been proposed as a model for aspects of cell migration in 3D ECMs: in 3D matrices, cells frequently encounter narrow channels through which they migrate, reminiscent of an effective 1D confinement [143, 146, 147]. Indeed, the morphology of cells on narrow 1D lines is highly stretched, similar to morphologies observed in 3D, which do not feature the broad fan-like lamellipodia observed on 2D substrates [143, 147, 153]. To understand decision making along multiple possible paths, networks of such micropatterned 1D lines have been introduced [151]. However, unlike 1D lines, physiological extra-cellular environments are structured, for example through the presence of

thin constrictions through which cells need to squeeze during migration [8, 154–156]. To study the response to such constrictions, micropatterned lines with periodic modulations, or gaps, which cells need to overcome have been developed. For example, ratchet-like confinement geometries were found to rectify the direction of motion of cells [95, 149, 150, 157], a process termed ratchetaxis (see [158] for a review). Using a microfluidic confinement with walls featuring similar ratchet-like modulations, a novel mode of migration relying on friction with the local topography of the walls was revealed [140]. Increasing the complexity of the environments even more, experimental systems have been developed to study how cells make decisions at junctions featuring either two symmetric [152] or several constrictions of varying widths [138, 139], which revealed the intra-cellular processes involved in cellular decision making in such systems. Finally, another approach to study cells overcoming constrictions is to consider geometries where the boundaries on both sides are closed, meaning that the cell has to turn around and make transitions back and forth across the same constriction. This was done using two-state micropatterns, which have the advantage that long trajectories of subsequent transitions can be obtained [19, 20].

These experimental approaches using standardised confinements have given insight into intra-cellular processes [138, 139] and have yielded quantitative cellular readouts, for example the degree of directionality in ratchetaxis [95], switching rates between run and rest states on 1D lines [159], or transition rates in two-state micropatterns as a function of the geometry [19, 20]. Based on our discussion of free 2D cell migration, a key challenge to go beyond cellular readouts from confined migration experiments is to develop an equation of cell motion that accounts for structured environments. In this case, the terms of the equation of motion will depend on both



**Figure 6.** Inferring an equation of confined cell motion. (A) Observation and tracking step: human breast cancer cells (MDA-MB-231) are confined to two-state micropatterns and imaged at 10 min time intervals (scale bar: 25  $\mu\text{m}$ ). Bottom: nucleus trajectories as a function of time, plotted for a (0, 50 h) interval. (B) Inference step: single-cell trajectory in  $xv$ -space (blue line) and recorded data points from a large data-set of cells (lightblue points, top). Averaged together, this gives the deterministic term  $F(x, v)$  (bottom). (C) Prediction and validation step: trajectories predicted based on the inferred model (top); experimental (blue) and predicted (red) survival probability  $S(t)$ , measuring the probability that a transition across  $x = 0$  has not occurred after time  $t$  (middle); normalised velocity auto-correlation functions (bottom). (D) Interpretation of the inference results: trajectories (green) of the deterministic dynamics for a number of different initial conditions. The flow field is shown by arrowheads, where acceleration is orange and deceleration is blue; shown for MDA-MB-231 cells (top), MCF10A cells (middle); and MDA-MB-231 cells migrating in a system without constriction.

the position and velocity of the cell. As these cells solve the challenge of navigating their confining environment, the terms of the equation of motion give insight into how cells dynamically solve this problem and thus encode how it responds to the structures in its environment, which we will discuss in the next section.

### 3.3. Dynamical models of confined cell migration

Learning a data-driven model of confined cell migration requires large data sets of trajectories, which can be obtained using minimal *in vitro* confinements. In previous work, we used two-state micropatterns as a minimal system to study how cells overcome thin constrictions in confining environments [19]. To provide a pedagogical example of how one can learn an equation of motion from confined cell migration data, we will discuss this example here in more detail. These micropatterns consist of two square adhesive islands connected by a thin adhesive bridge (figure 6(A)). This setup leads to repeated stochastic transitions of the cells between these two islands, with large variability both over time and across cells. Based on the trajectories of these cells, we then developed a generalisation of the persistent random motion model (equation (3)) to the problem of confined migration. An important assumption in the persistent random motion model is the uniformity and isotropicity of space: the cellular dynamics are assumed to be independent of position, and the same in all directions. Clearly, these assumptions are no longer valid in structured systems. This suggests a more general formulation of an equation of

cell motion for confined migration, in which the dynamics can also depend on the position  $x$  of the cell, which we refer to as an *equation of confined cell motion*:

$$\frac{dv}{dt} = F(x, v) + \sigma(x, v) \eta(t) \quad (6)$$

where  $F(x, v)$  is a generalised version of the deterministic term in equation (3), and  $\sigma(x, v)$  is the amplitude of the stochastic fluctuations. Note that in the presence of state-dependent noise, meaning that  $\sigma(x, v)$  is not a constant, the inferred deterministic term depends on the chosen noise-convention [115]. Here and throughout the text, this equation is interpreted in the Itô sense, but note that the inferred deterministic term  $F$  would differ in the Stratonovich convention if the noise is  $v$ -dependent. Put simply,  $F(x, v)$  is the average acceleration of the cell as a function of its position  $x$  and its velocity  $v$ . Importantly, other descriptions for the dynamics are in principle possible, and this postulated equation could be incorrect. Thus, once a model of this form has been inferred, one has to test its predictive power and contrast it with that of alternative descriptions, which we discuss below. Note that in this case, the dynamical description is 1D, as the lateral dimensions are highly constrained by the pattern. Furthermore, we here start with a memory-less description, which is simpler than the memory kernel equation of motion for 2D migration (equation (5)). Thus, the inference procedure starts with the simplest model which is only modified when the data demands it. The aim is now to determine the structure of the dynamical terms  $F$  and  $\sigma$  in a completely data-driven method

based on the experimental trajectories. Specifically, to a first approximation, the deterministic term of this equation can be inferred using a conditional average of the observed cellular accelerations:

$$F(x, v) \approx \left\langle \frac{\Delta v}{\Delta t} \middle| x, v \right\rangle \quad (7)$$

which is the generalised formulation of equation (4) for an equation of motion with positional dependence. As in equation (4), this relation is approximate as discretisation effects are neglected here for simplicity [86]; see equation (14) for a more technical discussion on how these effects can be removed. The simple grid-based binning approach suggested by equation (7) works as follows: the trajectories are represented in the position–velocity phase space, which is split into bins using a regular grid (figure 6(B), top). In each bin, the average acceleration is measured (equation (7)), giving the deterministic term  $F(x, v)$  (figure 6(B), bottom). Similarly, by calculating the standard deviation of fluctuations, the stochastic term  $\sigma(x, v)$  can be inferred. Note that a more data-efficient approach using on a set of smooth basis function such as polynomials or Fourier components can also be used, which we discuss in section 4.1.

Importantly, while the experimental data is used to constrain the shape and parameters of the deterministic dynamics  $F(x, v)$ , there is no guarantee that this approach yields an adequate representation of the dynamics of the system over a broad range of time-scales: the inference approach relies on the assumption that the dynamics of the system can in fact be described by the equation of motion equation (6), which could fail in many ways.

On the one hand, the dynamics could be more complex and could require additional memory terms [94], a time-dependent description [109], or an explicit description of the cell-to-cell variability [128]. To test the validity of this description, we therefore need to perform a test of predictive power. Specifically, to perform the inference, we constrained the equation of motion solely based on the short time-scale information provided by the experimental trajectories, including the velocities and accelerations of the cell. Thus, as an independent test of the model [19, 94], we predict statistics quantifying the cell behaviour on long time-scales, for example the distribution of transition times or the velocity auto-correlation function, which all capture the experimentally observed statistics (figure 6(C)).

On the other hand, the dynamics could also be simpler and we have to ensure that we identified the simplest model consistent with the data. To address this, we increased the complexity of inferred models step-by-step and ruled out the possible simpler models. For instance, an alternative inference based on a first order equation of motion (for  $dx/dt$  as opposed to  $dv/dt$ ) was unable to capture the data. Furthermore, simplifications of the general non-linear term  $F(x, v)$  into a separable form  $F(x, v) = F_x(x) + F_v(v)$ , as would be the case for a conservative potential  $V(x)$ , such that  $F_x(x) = -\partial_x V(x)$ , were inconsistent with the data. Based on this, we concluded that equation (6) was the simplest model that could capture the

data. These examples already show how exploring models that do not describe the data can be very instructive, as they allow to rule out simple hypotheses.

In the example of the confined cell problem, we found that an insightful representation of the system can be achieved by examining the deterministic dynamics of the system in a phase-portrait of position and velocity (figure 6(D)). Intuitively, one might expect that the hopping behaviour across the thin constriction placed by the micropattern could be generated by a noisy cellular activity competing with an effective energy barrier placed by the constriction. Strikingly, however, the inferred map of the deterministic accelerations reveals that cells have a tendency to accelerate into the constriction. In fact, the flow field of the deterministic dynamics exhibits an excitable flow, where a small noise-driven perturbation leads to a large excursion in the phase space due to a deterministic amplification of the cell speed. This amplification is observed in both cancerous (MDA-MB-231) and non-cancerous (MCF10A) cells, suggesting that it may be a generic cellular response to thin constrictions. Indeed, in systems in which the constriction is removed, the amplification vanishes (figure 6(D), bottom). This approach also reveals that the non-linear dynamics are poised close to a bifurcation between a limit cycle and a bistable system. Interestingly, different cell lines exhibit behaviours on both sides of this transition: MDA-MB-231 cells exhibit a limit cycle, while MCF10A cells show excitable bistable dynamics. Thus, the deterministic phase-portrait implies that the cancerous cells have a stronger tendency to overcome the constriction, while the non-cancerous cells rely on stochastic fluctuations to perform transitions. Interestingly, cancerous cells treated with tumour suppressor microRNA 200c were recently shown to undergo the same transition [160], highlighting how this data-driven approach can be used as a read-out of cell behaviour.

In the next section, we will discuss how we can use these insights to quantify and characterise the striking variability in the observed cell behaviours, which are already apparent at the level of the cell trajectories. Moreover, this approach could help advance our understanding of locomotion at the molecular level by providing constraints for bottom-up models that connect microscopic rules to the system-level dynamics of cells. Finally, the insights gained based on this framework could provide a generalisable basis to investigate the dynamics of assemblies of interacting cells. We will discuss both of these aspects in the following sections.

### 3.4. Why do cell migration dynamics appear to be underdamped?

The equation for 2D persistent random motion (equation (3)) and the equation of motion for confined cell migration (equation (6)) share a key feature: both are stochastic differential equations that are second-order in time, and therefore a manifestation of the *underdamped* Langevin equation. These equations predict the acceleration as a function of position and velocity. This is in contrast to first-order stochastic equations of motion which are frequently used to describe the motion of overdamped Brownian systems subject to thermal noise [161].

For such overdamped Brownian systems, the effects of inertia can be neglected at time-scales larger than the velocity relaxation time  $m/\zeta$ , where  $\zeta$  is the friction coefficient and  $m$  is the mass of the particle. Therefore, friction is directly equated with the sum of thermal and external forces, yielding a first-order, *overdamped* Langevin equation. However, the same physical argument applies to migrating cells: the forces acting on cells, including frictional forces, are much larger than the inertial term  $m\dot{v}$ , and thus we can take  $m \approx 0$  to a very good approximation. Why then are cell migration dynamics described by underdamped equations of motion?

An underdamped equation describes a process in which velocities have temporal correlations, and do not just follow a white noise process as in overdamped systems. Physical inertia is one way of introducing temporal correlations, as the inertia prohibits instantaneous reversals of direction, and instead introduces a characteristic time scale to adjust velocities. Similarly, cells do not instantaneously change their direction if they are in a polarised state, meaning that polarisation gives rise to a kind of ‘effective inertia’. To be precise, the cell’s propulsive forces constitute a stochastic process with correlation time-scales similar to the migration time-scales, and therefore introduce correlations in the cell velocities.

This idea can be demonstrated with a very simple model of the overdamped dynamics of a confined migrating cell that is driven by a self-propulsive cell polarity  $P(t)$  [162],

$$\dot{x} = f(x) + P(t) \quad (8)$$

$$\dot{P} = g(x, P) + \sigma\eta(t) \quad (9)$$

where  $f(x)$  are the forces acting on the cell in a confining environment, and  $g(x, P)$  is a general formulation of polarity dynamics that may depend on both the current polarity and the position of the cell. Here,  $P(t)$  subsumes all of the subcellular processes mentioned above that determine the direction of self-propulsion of the cell. Then, taking the derivative of equation (8) and substituting equation (9), we obtain:

$$\dot{v} = \underbrace{f'(x)v + g(x, v - f(x))}_{F(x, v)} + \sigma\eta(t). \quad (10)$$

This shows how an overdamped particle that is driven by underlying time-correlated polarity dynamics exhibits effective underdamped stochastic dynamics. The deterministic term  $F(x, v)$  is determined by a non-trivial combination of the confinement forces  $f(x)$  acting on the cell and the polarity dynamics  $g(x, P)$ . Importantly, this also means that we should not think of the deterministic term  $F(x, v)$  in equation (6) (and equivalently the term  $-\tau_p^{-1}\mathbf{v}$  in equation (3)) as physical force fields, but as an acceleration field that is determined by the underlying time-correlated machinery of the cell [162].

The underlying molecular processes that determine the cell polarity  $P(t)$  are complex, but can be understood as an interplay of actin flows and various polarity-mediating molecular factors. Importantly, these propulsive forces should not be confused with the traction forces exerted by the cell onto the substrate. Indeed, cellular tractions are typically much larger than

the forces needed to migrate [41, 163]. For instance, in keratocytes, traction forces are up to tens of nN [164], while the propulsive force of the leading edge was recently measured to be of the order of 1 nN [148]. Instead, the polarity is related to the intracellular concentrations of polarity cues and the actin flows, together determining the cell speed. Specifically, for a given actin polymerisation rate, the speed of a migrating cell is determined by the retrograde flow of actin, which is being polymerised at the leading edge, and depolymerised at the trailing edge: the slower the flow, the faster the cell [13, 165]. Note, however, that slower retrograde flow leads to higher traction, and thus there is an indirect correlation between traction force magnitude and cell speed [41]. The directionality of the actin flow is in turn determined by the concentration profiles of internal signalling cues within the cell, which reorient on long time-scales [14] (described in our example by equation (9)). Reorientations of these polarity fields lead to changes of the cell velocity vector, i.e. accelerations. Therefore, to understand the origin of the emergent cell migration dynamics, quantified by  $F(x, v)$ , we should consider how internal DOFs of the cell, including the cell shape, protrusion formation, and polarity determine the net movement of the cell, and how these DOFs couple to the external environment.

Contrasting the overdamped formulation (equations (8) and (9)) with the underdamped one (equation (10)) suggests an important conceptual insight into how inferred cell migration models can be connected to more mechanistically interpretable models. Clearly, the overdamped dynamics are physically more interpretable, as they connect directly to the known physics of self-propelled active particles [166] and the individual terms have a physical interpretation. However, inferring such overdamped equations for position and polarity from experimental data is currently an open challenge. To infer such equations from data, one would need trajectories of the cell polarity  $P(t)$ . However, there is no unique molecular marker of cell polarity, and for candidate markers of cell polarity, such as Rho GTPase localisation, it is experimentally challenging to collect large data sets of cell migration trajectories with motion and polarity tracked simultaneously (see [167] and section 6.2 for a more detailed discussion). In contrast, the underdamped formulation requires only tracking of the cell nucleus, from which the velocity DOF can be obtained through numerical differentiation. Thus, learning the underdamped dynamics of migrating cells from data can provide a key step towards understanding more mechanistic aspects. Indeed, the mapping from overdamped to underdamped dynamics suggested by equation (10) could provide a way to link mechanistic and inferred models more directly, by comparing the predicted  $F(x, v)$  of postulated active particle models to the inferred underdamped equation of motion.

#### 4. Learning equations of motion from stochastic trajectories

In the previous section, we discussed how inferring equations of cell motion gives insight into free and confined cell migration. In this section, we discuss the technical aspects of

performing stochastic inference. Please note that this section is not essential to understand the remainder of the review, and can therefore be skipped. Inferring equations from experimental data is a general problem that has been applied to a broad variety of physical and biological systems, ranging from dust particles in a plasma [168] to protein diffusion [83, 84], animal [73, 74] and robotic [169] behaviour, and neural dynamics [69]. There is a long history of inferring dynamical systems from trajectories of deterministic systems [76–80]. Such inverse problems are notoriously harder in stochastic systems such as migrating cells: it requires disentangling the stochastic from the deterministic contributions, both of which contribute to shape the trajectory. Importantly, however, fluctuations can also help to make a data set more informative about the system: in low-noise systems, the trajectory may only sample a very narrow region of the phase space, making it difficult to estimate the underlying dynamical system. Thus, successful inference typically requires a data set with sufficient diversity, which may pose a problem in highly stereotyped behaviours such as in morphogenesis.

A number of methods are now available to perform such equation inference in stochastic overdamped (first-order) equations [81–85, 87, 89–91] as well as underdamped (second-order) systems [86, 88]. Note that in addition to dealing with the intrinsic stochasticity of the system, realistic experimental data sets are also invariably subject to measurement error, which can have a major impact on numerical derivatives, and requires specialised estimators that are robust to such errors [85, 86]. In this section, we will first lay out the general principles of stochastic inference. Then, we focus on the specific case of performing inference of underdamped equations of motion which is relevant to cell migration trajectories.

#### 4.1. General principles

The overarching idea of equation of motion inference from a complex biological system is to derive a simple physical description of a small number of DOFs that does not require knowledge of all the microscopic details of the system. Thus, the idea is to identify the important DOFs that may follow relatively simple dynamics, that are slow compared to the time-scales of the microscopic processes. Developing an equation of motion model from experimental data in general involves five key steps, which were already illustrated in figures 4 and 6 using the examples of free and confined cell migration, respectively. Here, we will discuss these steps in a more general context, and illustrate them with the example case of underdamped equations of motion, as these are used to describe cell trajectories (see section 3.4), although the key points are equally relevant for overdamped stochastic systems [81–85].

**(1) Observation** In the first step, the important DOFs of the system have to be identified and observed. These DOFs have to be experimentally accessible and trackable over time to yield the trajectories  $x(t)$ . Furthermore, to enable inference and

interpretation of the model, this set of DOFs should ideally be low-dimensional and therefore provide a minimal representation of the system. In general, there is no principle to determine which DOFs should be tracked, and to some degree it is a choice that is made based on intuition and technical feasibility. The key objective is to arrive at a set of DOFs that allow construction of a predictive model (see point 3). In the examples of free and confined cell migration, this was done by simply measuring the trajectories of the cell nucleus (figures 4(A) and 6(A)). Identifying the relevant DOFs become even more challenging in collective multicellular settings, as discussed in section 7. In general, if the inference procedure proves to be difficult in the later steps, a different set of DOFs may need to be chosen.

**(2) Inference** The second step is the inference of a model from the observed trajectories. In this step, a general formulation of a stochastic dynamical system for the tracked DOFs should be postulated, which can then be systematically constrained using the data. To go from the data all the way to the inferred equation, three key steps need to be considered:

**(2.1) Equation selection** The first step is to select the structure of the equation of motion to be inferred from the data. In practise, this selection can often be done based on physical intuition. More principled approaches include searching for maximum predictability from delay embeddings [170], testing of Markovianity from data [171], or determining the scaling of increments with time [172]. For cell migration experiments where the polarity remains unobserved, the appropriate equations are typically underdamped equations of motion for the dynamics of the cell velocity (see section 3.4).

**(2.2) Basis selection** To infer the equation of motion an appropriate representation of the dynamical terms must be chosen. In the confined cell example, this corresponds to choosing how to approximate the functions  $F(x, v)$  and  $\sigma(x, v)$  by a set of basis functions. In this case, the dynamical terms are represented as a truncated basis expansion

$$F(x, v) \approx \sum_{\alpha=0}^{N_b} F_\alpha c_\alpha(x, v) \quad (11)$$

where  $\{c_\alpha(x, v)\}$  is the set of basis functions and  $N_b$  is the number of functions. Note that this expansion is written for a 1D system, but all expressions generalise straightforwardly to multidimensional systems [86]. Thus, the problem of inferring the equation of motion is reduced to estimating the parameters  $F_\alpha$ . If the noise is state-dependent, a similar expression can be written for the stochastic term  $\sigma^2(x, v)$ . The key problem is then to select the set of basis functions  $\{c_\alpha(x, v)\}$  that is appropriate for the problem at hand. These can be constrained by taking into account the symmetries of the system [85, 86], by applying Bayesian approaches to the fit-complexity trade-off [173], or by applying sparsity constraints to detect the relevant terms, such as SINDy [78, 79], which has recently been generalised to stochastic systems [87, 89–91]. Based on such

principles, one can then determine, for instance, if the noise in the system is state-dependent, or if it should be fitted by a constant amplitude.

(2.3) *Estimators* Finally, to perform the inference of the parameters specified by the selected basis, one should use error-corrected estimators. Here, by error-corrected we mean that the estimator converges to the correct result for simulated data sets without additional error terms (as shown either by mathematical proofs or numerical convergence tests [85, 86]). Such error terms can be caused by the stochasticity of the system, the inevitable discreteness of the sampled trajectories, and the presence of measurement errors (see section 4.2). These estimators then allow inference of the dynamical terms (figures 4(B) and 6(B)).

(3) **Self-consistency** Before testing the predictive power of the model, there are two tests of self-consistency that should be performed.

(3.1) *Noise correlation* A key assumption of stochastic inference approaches is that deterministic and stochastic contributions can be separated, which relies on the assumption of white noise of the stochastic term  $\eta(t)$ , such as in equation (6), meaning that  $\eta(t)$  is uncorrelated in time,  $\langle \eta(t)\eta(t') \rangle = \delta(t-t')$ . To test the self-consistency of this assumption, one can calculate the trajectories of the noise increments  $\Delta W(t) = \int_t^{t+\Delta t} \eta(s) ds$ . Specifically, an empirical estimator for  $\Delta W(t)$  is [19, 73, 94, 98]:

$$\Delta W(t) \approx \frac{\Delta t}{\sigma} \left[ \frac{\Delta v(t)}{\Delta t} - F(x(t), v(t)) \right] \quad (12)$$

where  $F$  and  $\sigma$  are the inferred deterministic and stochastic terms, respectively, and the deterministic term is evaluated using the position and velocity recorded at time  $t$ . Then, the auto-correlation  $\langle \Delta W(t)\Delta W(t') \rangle$  can be calculated, which should decay to zero within a single time-step if the white noise assumption was correct. Note that at the first time-step  $|t-t'| = \Delta t$ , a weak negative correlation can appear due to the presence of measurement errors [19, 112]. If this criterion is not satisfied, one typically has to revisit point 2.1 and consider a different class of models. For example, if a first-order equation inference is applied to cell migration trajectories, long time-scale noise correlations will appear, since a second-order equation is required.

(3.2) *Re-inference* A second criterion for a self-consistent model is that when new trajectories are simulated based on the inferred model, applying the same inference procedure to these simulated trajectories should yield a consistent result [86, 98]. An important aspect of this is that inferred equation can turn out to be unstable, meaning they fit the data locally in time, but diverge for long time intervals. This needs to be checked and avoided. If these criteria is not satisfied, it is likely that points

2.2 or 2.3 should be revisited, or an insufficient amount of data was used in the inference.

**(4) Validation and prediction** Steps 1 and 2 of the inference make assumptions about the system that could be incorrect, and thus the predictive power of the inferred model must be tested to validate it. The stochastic inference approaches described in sections 3.3 and 4.2 use as input only the short time-scale information of the trajectories, through the increments corresponding to velocities and accelerations. A key test of the model is then to predict long time-scale statistics that were not used in the inference. Which statistics are suitable for such prediction depends on the system at hand. For instance, for both free and confined migration the velocity auto-correlation function was a natural prediction target (figures 4(C) and 6(C)). If the predictive power of the model is low, one typically has to revisit points 2.1–2.3 to consider if the correct equation, a reasonable basis, and valid estimators have been used. Note that depending on the method and the amount of hyperparameter tuning in the inference, one may also want to consider splitting the data into training, validation and testing data sets, using standard approaches to such issues. Finally, to further challenge the model, one may want to test its predictive power on other experiments not used for training, such as mutants, perturbations, or other environmental conditions (such as a different micropattern geometry).

**(5) Interpretation** Having determined a valid model for the observed dynamics, this model can be interpreted to gain insight into the system. This last step is of course very much system-dependent. An important aspect of the stochastic inference approach is the decomposition of the dynamics into deterministic and stochastic components, i.e.  $F$  and  $\sigma$ . Based on this decomposition, these components can be interpreted separately, and their respective contributions to the dynamics can be conceptualised. For example, this decomposition revealed distinct classes of non-linear dynamical systems in the position–velocity phase space of confined migrating cells (figure 6).

#### 4.2. Sources of inference error

When following the sequence of steps laid out in the previous section, there are multiple sources of inference error, which can lead to deviation of the observed and the predicted dynamics. Here, we discuss three primary sources of error: finite data, imperfect data, and incomplete basis functions. The resulting errors can be minimised by adapting steps 2.2 and 2.3 of the scheme above. Additional sources of error can include the recording of unrepresentative DOFs or selecting the wrong equation, which we do not discuss further here.

**(1) Finite data** Realistic data sets consist of a finite number of trajectories of finite length, with a total length of all trajectories that we call  $T$ . The presence of noise and the potentially only partially explored phase space leads to sampling errors, which are random errors, and therefore vanish for  $T \rightarrow \infty$ .

**(2) Incomplete basis functions** In step 2.2 of the inference procedure, a set of  $N_b$  basis functions  $\{c_\alpha(x, v)\}$  needs to be chosen to perform the inference. If the basis is not sufficient to accurately represent the underlying model, then even with perfect, infinite data, there will be a systematic error, i.e. a representation error.

To deal with problems 1 and 2, there is a basic trade-off: as the number of parameters of the basis  $N_b$  increases, the representation error decreases, but the sampling error increases. Indeed, the mean-square error (MSE) due to finite data in the estimate of the deterministic term grows linearly with the number of parameters  $N_b$ , for both under- and overdamped dynamics [85, 86]:

$$\text{MSE} \propto N_b/T. \quad (13)$$

On the other hand, the representation error decreases with  $N_b$  in a way that depends on the underlying model and the set of basis functions considered. Thus, for a given amount of data  $T$ , there is an optimal basis size  $N_b^*(T)$  that can be inferred. To connect this discussion to the inference approach for confined migrating cells described in section 3.3, we point out that the grid-based binning approach effectively corresponds to a basis of top-hat functions at regularly spaced locations in the phase-space (figure 6(b)). The parameters  $F_\alpha$  then correspond to the average acceleration  $\dot{v}$  at that location in phase-space. This approach requires a large number of fitting parameters  $N_b$  and would therefore not perform well in high-dimensional systems such as interacting cells or experimental data sets with low statistics. In the confined cell example, we had access to a large data set of 1D nucleus trajectories (figure 6(A)), and thus the binning approach was feasible in this case. A generally better approach is to use a set of *smooth* basis functions, such as polynomials or Fourier components. In this case, fewer parameters are required and additional constraints such as symmetries of the system can be taken into account to further restrict the choice of basis functions. We will discuss how this approach allowed inferring an equation of motion for the more complex case of interacting cells in section 7.2.

Importantly, while equation (13) provides an argument for how many parameters  $F_\alpha$  can be learned from the observed trajectories, it does not tell us which parameters are relevant to describe the dynamics. For this, sparsity-enforcing methods, such as such as SINDy [78, 79], are required. Briefly, these work by augmenting the cost function that is minimised by the inference by penalising non-zero parameters. Such sparsity constraints have recently been generalised to stochastic systems [87, 89–91], but have not yet been applied to experimental data sets.

**(3) Imperfect data** Experimental trajectories are inevitably subject to measurement error, and are recorded at finite time intervals  $\Delta t$ . Both of these in principle separate effects—discreteness and measurement error—lead to systematic errors in the inference results, which rely on the numerical derivatives of the trajectories. To address these issues, error-corrected estimators have been developed.

To illustrate this, we first focus on how to deal with discreteness in an underdamped system [86]. This poses a non-trivial problem, since if only the trajectory  $x(t)$  is observed at discrete time steps, this means that one of the dynamical variables, the velocity, is not observed, but has to be estimated as well. Importantly, this leads to systematic errors to the inference result, which persist even in the limit as  $\Delta t \rightarrow 0$ , and do not average away even in the limit of infinite amounts of data. For instance, in the simplest case of an underdamped equation of motion with a linear damping term, i.e. the persistent random motion model with persistence time  $\tau$  (equation (3)), the conditional average of the accelerations does not converge to  $-v/\tau$ , but to  $-\frac{2}{3}v/\tau$ , as first pointed out in [112]. This is why we wrote equation (4) as an approximation rather than an equality. However, this systematic error can be removed by including an error-correction term in the estimator, derived by inverting the stochastic Ito–Taylor expansion of the equation of motion [174]. The estimator for the deterministic term that is robust against discretisation effects then reads [86]

$$F_\alpha = \langle \dot{v} c_\alpha(x, v) \rangle - \frac{1}{6} \langle \sigma^2 \partial_v c_\alpha(x, v) \rangle. \quad (14)$$

Here, the first term is the conditional average of the accelerations (equivalent to equation (3)), while the second term is a correction term that arises due to the projection of an estimated second derivative (acceleration) onto an estimated first derivative (velocity) of the DOF. Note that the correction term vanishes for deterministic systems ( $\sigma = 0$ ) and depends on the derivative of the basis function, indicating that smooth basis functions should be used. This correction can therefore not be applied to a binning inference. For the stochastic noise amplitude the estimator

$$\sigma_\alpha^2 = \frac{3\Delta t}{2} \langle \dot{v}^2 c_\alpha(x, v) \rangle \quad (15)$$

which depends on the square of the accelerations, akin to a standard deviation of the accelerations, since at short time-scales, the accelerations are dominated by the noise. Another way to address the error due to discretisation is a Bayesian maximum likelihood approach [88].

A conceptually similar approach can be used for error-corrected estimators that are robust to measurement errors. In this case, the assumption is that measurement errors are uncorrelated in time, meaning that multi-point estimates of the numerical derivatives can be combined in such a way that all systematic error terms vanish by construction [85, 86]. The importance of this measurement error-correction depends on multiple factors: the magnitude of the measurement error  $|e|$ , the measurement time interval  $\Delta t$ , and the dynamics of the process. Specifically non-corrected estimators (equation (7)) only provide accurate results in the regime  $|e| \ll \sigma \Delta t^{3/2}$ , where  $\sigma$  is the magnitude of the stochastic noise; while the error-corrected estimators can provide accurate results up to errors of the size of a typical step, i.e.  $|e| \lesssim \langle |v| \rangle \Delta t$ , where  $\langle |v| \rangle$  is the average speed [86].

Importantly, these error-corrections do not easily generalise to grid-based binning approaches that were discussed in

section 3. Binning effectively corresponds to using basis functions  $c_\alpha(x, v)$  that correspond to non-smooth top-hat functions placed at regular intervals. This means that equation (14), which includes the derivative of the basis functions, is not easily applicable, suggesting that smooth basis functions should be used, such as polynomials or Fourier components [86]. Binning can therefore be used to gain initial intuition for a data set, under the condition that the inference errors due to discretisation and measurement error are sufficiently small that a self-consistent and predictive model can be obtained (as was the case in the examples in section 3). However, in general, the use of an expansion of smooth basis functions with error-corrected estimators [85, 86] is strongly recommended. As we will discuss in section 7.1, smooth basis functions have the additional advantage of being much more data-efficient and therefore a better choice for high-dimensional or interacting systems of multiple cells.

In summary, stochastic inference approaches allow learning the governing dynamical systems from observed experimental trajectories in a way that trades-off finite data limitations and representation complexity, and is robust to measurement errors and discretisation errors. That said, there are many open challenges in terms of method development that could aid future applications of stochastic inference to cell migration data. For instance, there is currently no principled approach to automatically identify the most representative DOFs directly from an experimental movie (step 1 in section 4.1), although first approaches towards learning dynamics from movies have been made [175]. Furthermore, principled approaches for selecting the class of equation (step 2.1 in section 4.1) or developing models with unobserved hidden variables is a central challenge. These problems have been recently addressed in low-noise, near-deterministic systems either leveraging Taken's embedding theorem [170] or ranked-choice model selection [176], but a generalisation to stochastic systems is outstanding. This could also provide a way to address the potential role of unobserved hidden variables in the dynamics, such as the cell polarity. Finally, while approaches for basis selection and sparsity constraints have been applied to stochastic systems [87, 89–91], they have not yet been combined with error-corrected stochastic estimators, and are therefore not yet applicable to underdamped stochastic systems.

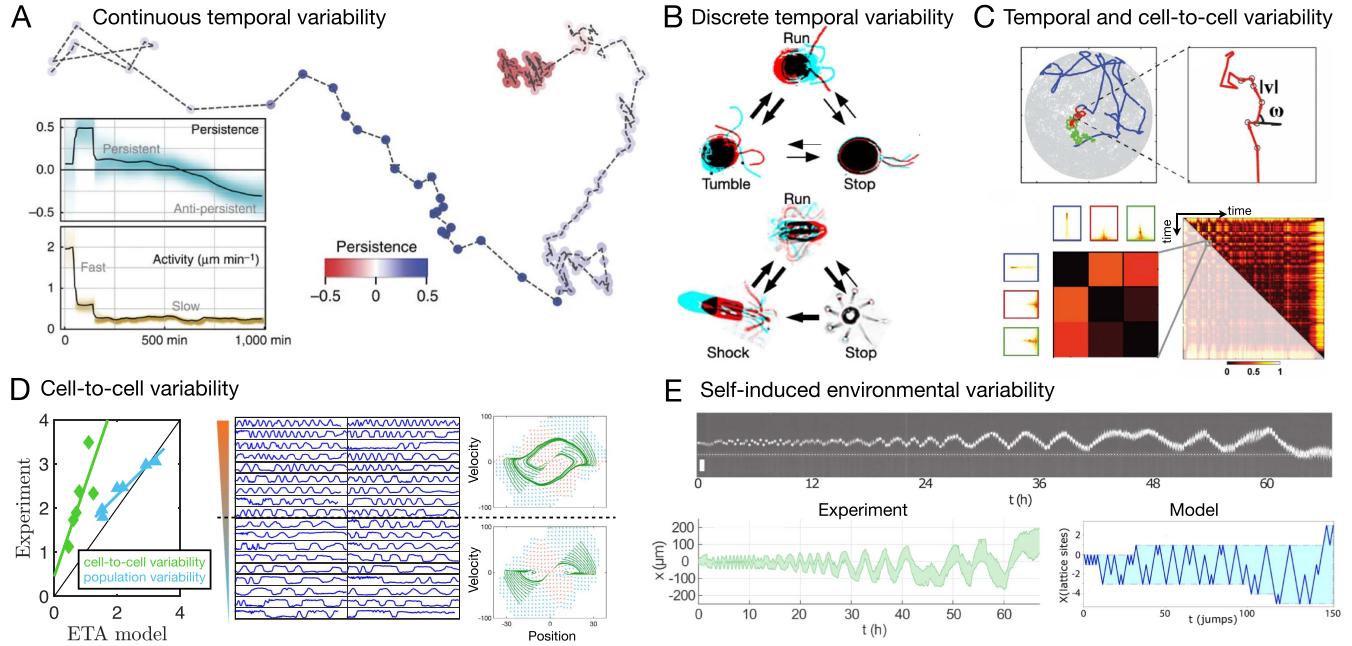
## 5. Inferring heterogeneity in cell behaviour

A key feature of migrating cells is the large variability of the observed behaviours within a cell population. A typical set of cell migration trajectories exhibits large variations in behaviour between different individual trajectories, but also over time within a single trajectory. Tracing the origins of such variability is an open challenge for which data-driven approaches are ideal, since it naturally relies on the analysis of large ensembles of observations [172]. Here, we propose to distinguish four distinct contributions to the behavioural variability of migrating cells:

- **Intrinsic stochasticity:** the intra-cellular machineries driving cell behaviours operate at the molecular level, and are thus subject to intrinsic noise. At larger scales, where these molecular DOFs remain unobserved, this intrinsic noise leads to seemingly random patterns in behaviour, which we refer to as the intrinsic variability of cell behaviour.
- **Cell-to-cell variability:** even in populations of cells with identical genomes, the stochasticity of intra-cellular processes such as gene expression, cytoskeletal rearrangement and protein localisation can lead to large differences in the proteomes of individual cells [177–181]. At the cellular scale, this diversity can lead to variations in cell behaviour, which is also referred to as phenotypic or population heterogeneity. The connection from molecular to behavioural heterogeneity has been demonstrated in cellular processes ranging from growth rate and drug response to morphology [182–185], and has been suggested to play an important role in collective cell migration [186, 187].
- **Temporal variability:** the behaviour of cells may also exhibit variations over time: as cells undergo the cell cycle, they grow, which may also affect other behaviours, including cell migration [188]. Furthermore, cells may switch between qualitatively distinct modes of behaviour, meaning that separate models for each behaviour, as well as a model for the switching itself, must be considered [189–191].
- **Environmental variability:** potentially unobserved changes in the extra-cellular environment may cause changes in behaviour, which could be mistaken for other types of variability. This can occur, for example, in cell migration experiments in environments with unobserved structures, including porous 3D matrices [128].

Clearly, these different sources of variability are hard to disentangle, and sometimes it may not be clear, even in principle, what should be counted as intrinsic noise vs temporal variability. Here, we argue that the key property is the time-scale of deviations of the process from the ensemble- and time-average (ETA). The definition of intrinsic noise is that it is stationary in time, implying no long-lasting deviations from the ETA. Cell-to-cell variability, in contrast, by definition leads to deviations on a time-scale equal to the lifetime of a single cell: these are differences in behaviour of a single cell relative to the population average that persist for the entire life of the cell (and may in principle be passed on to daughter cells). Temporal variability on the other hand leads to deviations on time-scales similar to the time-scales of the behaviour of interest, but shorter than the lifetime of the cell.

The variability of cell behaviour can make model inference challenging. Firstly, significant variability can mean that a large data set of trajectories is required to sufficiently sample the distribution of behaviours to infer a reasonable ETA model. Secondly, in some cases, ETA models may not be predictive of behavioural statistics that are sensitive to the variability. In this section, we will focus on inference approaches that are specifically tailored to quantify and characterise the variability of observed cellular behaviours in single-cell contexts.



**Figure 7.** Experimental and theoretical approaches for detecting variability in cell behaviour. (A) For cell behaviour with a variability that evolves continuously in time, the super-statistical random walk analysis infers the persistence and activity of the cell as a function of time. Inset shows the inferred time-dependence for the trajectory on the right (MDA-MB-231 breast cancer cell migrating on uncoated plastic). Reproduced from [109]. CC BY 4.0. (B) Single-cell ciliate organisms often switch between discrete states of motion, as shown here for *Chlamydomonas reinhardtii* (top, 2 cilia) and *Pyramimonas octopus* (bottom, 8 cilia). Reproduced from [191]. CC BY 4.0. (C) Trajectories of *Tetrahymena* cells (hundreds of cilia), that are quantified by speed  $|v|$  and angular velocity  $\omega$  (top). Changeability matrix quantifying the similarity of the joint distribution  $P(|v|, \omega)$  over time and across individuals. Reproduced with permission from [190]. (D) Comparison of experimental statistics to ensemble- and time-averaged (ETA) model predictions reveals the cell-to-cell variability in cell hopping behaviours in two-state micropatterns (left). Green data points: variance of average hopping times across cells; blue data points: variance across the population. Sorting cells by hopping activity reveals distinct underlying deterministic motility patterns (right; MDA-MB-231 breast cancer cells). (E) Self-induced variability through protein deposition of a migrating cell exhibiting oscillations of increasing amplitude (top and left; MDCK cells on fibronectin coated 1D lines). A model of the cell as a persistent self-attracting random walk captures the behaviour (right). Reproduced from [167]. CC BY 4.0.

Gaining insights into how these distinct contributions determine the overall variability of cell migration could be important for understanding physiological migration processes, as well as the mechanistic basis of the behaviour. However, disentangling these different contributions to the behavioural variability based on an observed data set can pose a formidable challenge. To this end, several data-driven approaches have been developed in the context of cell migration, which we will discuss here.

### 5.1. Quantifying temporal and cell-to-cell variability in behaviour

Models for cell migration are typically formulated as stochastic equations of motion, which is a natural way to capture processes exhibiting fluctuations. In the equations of cell motion introduced in previous sections (equations (3) and (6)), the stochastic white noise term ensures that no two trajectories look alike. This is a model for the *intrinsic stochasticity* of the migration process. To determine the structure and parameters of cell migration models, the dynamics are typically averaged across different cells and over time, yielding ETA stochastic models that describe the average member of a cell

population. Therefore, these approaches fail to capture cell-to-cell and temporal variability. Similarly, bottom-up models for cell motility typically assume that all cells in a population can be described by a common set of parameters that are constant in time. To demonstrate how data-driven methods can help quantify temporal and cell-to-cell variability, we will first discuss two examples of quantifying temporal variability, before discussing another approach to disentangle the contributions of temporal and cell-to-cell variability.

A paradigmatic example of temporal variability in behaviour is the switching between discrete modes of movement, such as run-and-tumble motion in bacteria [192]. Similar switching behaviours have been suggested for mammalian cell migration [189]. A principled inference approach to identify such temporal variability in cell migration was developed by Metzner *et al* [109]. They considered a generalisation of the persistent random motion framework, which allows for time-dependent migration parameters (figure 7(A)), without imposing *a priori* the existence of discrete behaviours. In this ‘super-statistical’ approach, both the persistence  $\tau_p$  and the noise amplitude  $\sigma$  in equation (3) become functions of time. The values of the parameters are inferred from experimental trajectories using a Bayesian approach. With this method, the

local persistence and activity of migrating cells could be identified as a function of time, revealing pronounced phases of ‘run’ and ‘rest’ states in trajectories of individual cancer cells (figure 7(A)). Such switching between behaviours has been suggested previously to be due to distinct transient intra-cellular organisations [13]. Interestingly, similar discrete switching between persistent and anti-persistent motion was recently identified *in vivo* in *Drosophila* haemocyte migration using a data-driven machine learning approach [193]. In the future, connecting such data-driven identification of temporal variability to live imaging of intra-cellular features could provide a way to link cellular behaviour to the underlying mechanisms, and how these control switching between subclasses of motility behaviours.

Links between morphological features and behaviour are clearer in the behaviour of swimming protozoans, whose flagella can organise into distinct states corresponding to behaviours such as run, tumble and stopping [194, 195] (figure 7(B)). These distinct morphological states already suggest the existence of discrete behaviours, which is in contrast to the more continuous state space of eukaryotic cells which undergo much more complex shape changes. This discreteness already suggests temporal variability with switching between different modes of behaviour as a natural framework for protozoan motility. Based on trajectory data of such swimming protozoans, quantitative frameworks for migration variability have been developed that characterise the time-dependent motion in sliding time windows to access the joint probability distribution of speed and turning moment [190] (figure 7(C)). By quantifying the change in this distribution over time and between individuals gives rise to a changeability matrix allowing pairwise comparison between any two time points and between individuals. Using clustering and dimensional reduction, this matrix leads to a low-dimensional behaviour space revealing two-state ‘roaming and dwelling’ model of swimming behaviour for multi-ciliate *Tetrahymena* cells. Conceptually similar approaches have been developed for other organisms and their interactions with confining boundaries [191], revealing different types of discrete cell states including run–tumble–stop behaviour in biflagellate vs run–shock–stop behaviours in octoflagellate protozoans (figure 7(B)). These analysis frameworks could have potential also for eukaryotic cell migration, provided that a sufficient time-resolution can be achieved experimentally, which is key for a sufficient sampling of the sliding time windows of such an approach. Indeed, we tested the changeability approach [190] on confined cell migration data of cancer cells [19], but found that the frame rate was insufficient to properly sample the changeability matrix. Conceptually similar frameworks have also been invoked in the literature on animal behaviour, including fitting of locally linear dynamical systems to motility data of the nematode *Caenorhabditis elegans* [75]. Since the data analysis problems in animal behaviour are often very similar to those in cell migration, potentially connecting these approaches to cellular data could be an interesting perspective. In sum, these works provide computational tools to systematically identify and characterise cell-to-cell

and temporal variability in migration behaviours from trajectories alone.

While these frameworks provide a way to quantify and characterise cell migration variability over time and between individuals, they do not provide a method to determine whether such variability exists in the first place. Indeed, if the observed trajectories are short, as is often the case in cell migration experiments, they may appear variable simply due to the randomness introduced by intrinsic stochasticity. How then can real variability be distinguished from apparent variability due to the intrinsic stochasticity? This question has previously been raised in the context of collective cell migration [196], where it was suggested to compare the observed variability to an appropriate ‘null-model’. Specifically, this means performing a direct comparison of variability-sensitive experimental observables, such as population variances, to the predictions by a parameter-optimised model without variability. Deviations from the variability-free model can then provide an indicator for real variability. A difficulty in applying this approach is that it requires both a large ensemble of migration trajectories in a standardised setting, and an appropriate theoretical framework to provide a null-model.

To demonstrate how such a null-model approach can work in practice, we will show how to use the inferred equation of motion for confined cell migration [19] (section 3.3) as a benchmark to identify behavioural variability [197]. The inference of this equation of motion was based on the assumption that there is no variability between cells or in time, such that we used an ETA inference approach. This ETA equation of motion model correctly captures the ETA statistics of the experiment, such as correlation functions [197]. Thus, this equation of motion provides a null-model to predict the amount of variability between individual (short) trajectories based on only intrinsic noise, to which we can compare the experiment. We found that the variance in behaviour between individual cells was larger in the experiment than that measured in an ensemble of trajectories of similar length predicted by the null-model (figure 7(D), left). This indicated that there is real cell-to-cell variability in the system, beyond the random variations expected from a single, ergodic and stationary process. Interestingly, our analysis further revealed that within the cell population, there are qualitative differences in the class of dynamical systems describing the migration of individual cells (figure 7(D), right). Faster cells exhibited limit-cycle dynamics, while slower cells exhibited bistability, with two stable fixed points. The coexistence of distinct dynamical system states within a population of migrating cells has been suggested to originate from a heterogeneity in microscopic migration parameters [55]. Specifically, it was suggested that tuning the elasticity and adhesiveness of cells could lead to distinct dynamical behaviours, including smooth migration, stick-slip migration, as well as bistability between these two modes.

Taken together, these results demonstrate how combining systematic inference tools that account for cell-to-cell variability with mechanistic models could in the future lead to novel insights into the behavioural variability of cell populations. An exciting approach in this respect would be to correlate

variability at the molecular scale with variability at the behavioural scale, which could give insight into how molecular organisation correlates with behaviour without relying on artificial perturbations of the system [198–201].

### 5.2. Identifying sources of environmental heterogeneity

Conceptually, we think of temporal and cell-to-cell variability to have their underlying cause in cell-intrinsic properties that change over time or between cells, such as protein concentrations or localisation. However, in addition to this, migrating cells also encounter variability in their environment including heterogeneous ECMs [202, 203] or contact with other cells [204, 205]. Indeed, apparent cell-to-cell variability in collective systems has in many cases been shown to be caused by environmental factors, including local cell density, cell–cell contacts and relative location in a cell cluster [182, 206–208]. In the context of single cell migration, this was nicely demonstrated by applying the ‘super-statistical’ approach introduced in the previous section to cells migrating through series of constrictions, showing how cell activity and persistence adapt to the local structure of an external confinement [109]. If these external features were not observed, then these responses would appear as strong temporal variability, even though cellular responses to confinements can be explained through an ensemble averaged model that takes into account the cell position within the confinement (as shown in section 3.3 and [19]). An interesting special case of such extrinsic variability is *self-induced environmental heterogeneity*, where the cell itself causes changes to its environment which in turn affect its behaviour. These are by nature harder to observe experimentally, and can therefore be mistaken for temporal variability. Here, we discuss two examples of this case in which quantitative frameworks for such self-induced environmental changes were developed.

First, in 3D migration through a matrix, some cells perform proteolysis, which is a mechanism that allows cells to locally digest the surrounding matrix to create a migration path. This behaviour was shown to lead to asymmetries in the preferred direction of motion of cells: cells were more likely to turn around by 180° than expected based on persistent random motion, thus backtracking on their previous path [128]. As a model for this process, the *anisotropic persistent random walk* model was proposed, which includes spatially anisotropic parameters (parallel vs orthogonal to the direction of motion) and thereby account for this effect. Such proteolytic behaviour was also shown to lead to directional random walks in the presence of global strain applied to the matrix [209].

In the second example, it was found that migrating cells deposit material on the surface on which they migrate, causing them to behave differently when they return to a location that they previously visited [167] (figure 7(E)). Specifically, cells were observed to preferentially occupy previously visited areas. In this work, data-driven inference was used to generalise the phase-space analysis introduced in section 3.3 to the problem of self-attracting migration on a 1D line. This approach revealed that cells deterministically accelerate away from the boundaries of previously explored space. This

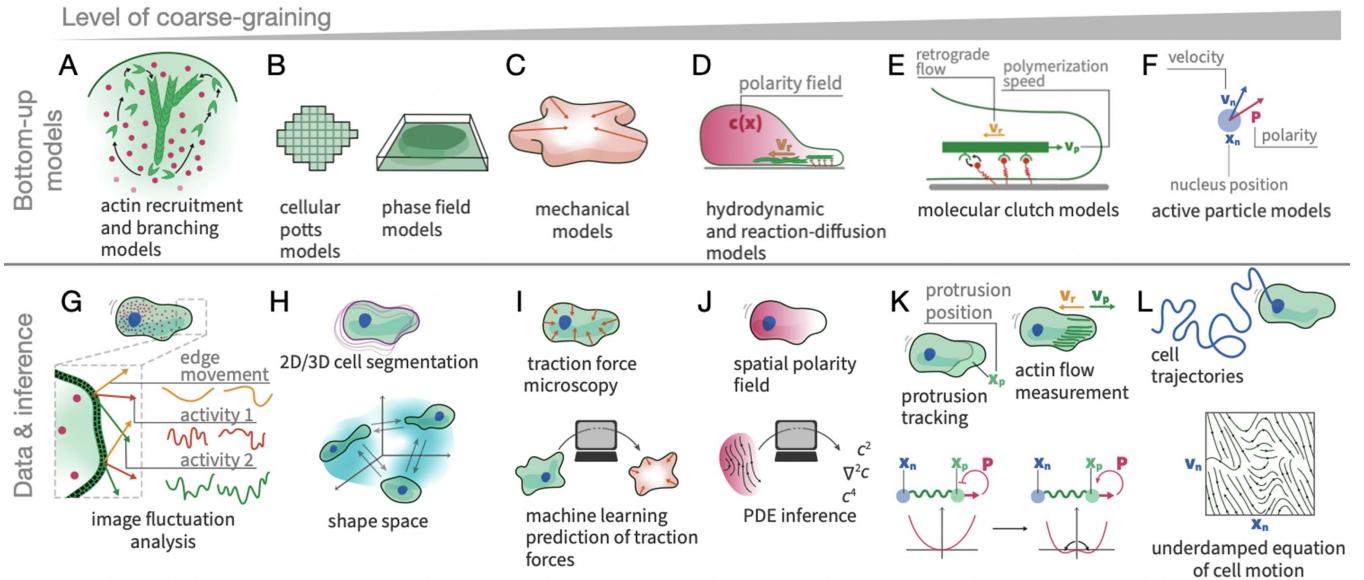
observation motivated a quantitative description using a *persistent self-attracting walk* model, which quantifies the relative probabilities of turning back vs. exploring new areas. This effect leads to long-lived spatial memory in the migration, which can have dramatic consequences for the ways in which cells search and explore space. Indeed, a phase-field model approach modelling the interaction of cells with their secreted footprint predicted that such memory lets cells switch between confined, oscillatory, and exploratory migration when they explore 2D spaces [210].

In conclusion, these approaches identified important cell migration mechanisms using data-driven analysis of the migration trajectories which exhibited striking variability. The analysis revealed that the observed variability is in fact due to extrinsic effects, albeit regulated by the cell itself. These findings are particularly interesting in the broader picture of regulated cell-to-cell variability proposed in [208], where it was suggested that deterministic, regulated variability could have functional importance in cell population, which is in contrast to cell-to-cell variability caused by random fluctuations of intra-cellular processes.

## 6. Connecting cell dynamics to mechanisms

In the previous sections, we have discussed how quantitative frameworks for cell migration can provide data analysis tools and yield conceptual frameworks to think about cell behaviour. A third important contribution such frameworks could make to the field is by providing constraints for mechanistic cell migration models. We refer to models as ‘mechanistic’ if they are based on a bottom-up approach in which the model is postulated based on known cellular processes and their simplified physical description. This is in contrast to the data-driven, top-down approaches that we have focused on so far in this review.

There is a long history of mechanistic biophysical modelling of cell migration (see e.g. [102] for a review). Here, we focus on how combining bottom-up models with top-down data-driven approaches can help address some of the key challenges in the field: (1) constraining mechanistic models that make predictions for the long time-scale behavioural dynamics of cells. (2) Understanding how cell dynamics may respond to external inputs, and how this could be included in physical models. (3) Connecting different classes of mechanistic models across scales into a coherent theoretical framework for cell migration. We first provide a brief overview the key types of mechanistic models for single cell migration, and discuss how they may be connected to inference approaches to address these challenges (section 6.1). Next, we discuss how performing data-driven inference on more complex cellular features beyond cell trajectories, such as cell shapes and protein localisation, could provide a bridge between top-down and mechanistic models (section 6.2). We review these approaches by systematically increasing the level of coarse-graining and length scale of the models and observables, and specifically highlight how models and data can be compared at each length scale (figure 8).



**Figure 8.** Connecting bottom-up and top-down models for single-cell migration. *Top row:* bottom-up models for cell migration. Though there is some ambiguity in this classification, models are ordered by increasing level of coarse-graining from left to right. *Bottom row:* types of cellular datasets and top-down inference approaches at the corresponding levels of coarse-graining. (A) Actin recruitment and branching models describe the establishment of the actin network and how it propels the cell membrane in protrusions [214–220]. (B) Cellular Potts models [25–33] and phase field models [34–39] describe the evolution of the 2D or 3D cell shapes. (C) Mechanical models predict cellular traction forces [229–231]. (D) Hydrodynamic models of actin flow and reaction diffusion models describe the coupling of spatially extended polarity fields to motility [13, 14]. (E) Molecular clutch models describe how actin retrograde flow and polymerisation propel the cell forward by establishing friction with the substrate through adhesions [40, 55, 56]. These are typically minimal 1D description that simplify the actin branching models shown in (A), although more complex 2D versions have also been proposed [226]. (F) Active particle models describe the evolution of cell position based on velocity and polarity [166]. (G) Image fluctuation analysis based on high-resolution live imaging of polarity factors allows causality inference of actin binding factors and how they control protrusion growth [232–234]. (H) Segmentation of the 2D or 3D shape of cells can be integrated in shape space models of morphodynamics, providing a dimensional reduction of complex dynamics [235–238]. (I) Integrating traction force microscopy datasets has been done in machine learning models predicting forces based on protein concentrations [239]. (J) Datasets of spatial fields of polarity cues can be treated with partial differential equation (PDE) inference to obtain reaction–diffusion models of cell polarity. (K) Protrusion tracking or actin flow measurements allow inference of nucleus–protrusion models [228]. (L) Cell trajectory data allows inference of underdamped equations of cell motion [19].

### 6.1. Bottom-up models for cell migration

Bottom-up biophysical modelling of single cell migration initially focused on particular aspects of the motility machinery, such as the ratchet model for force generation by actin polymerisation [211, 212], actin branching [213], and the molecular clutch model for adhesion dynamics [40] (figures 8(A) and (E)). To integrate these underlying mechanisms into cell-scale models, effective descriptions of their coupling to the large-scale behaviour of cells are required. We will describe these models in order of increasing level of coarse-graining, starting with computational models that account explicitly for the cell shape and subcellular features, and then moving to more coarse-grained approaches describing cells as gels, mechanical modules, and particles (figure 8).

To couple mechanisms to cell migration at the cell scale and to describe the typical shapes of migrating cells, a number of studies have developed *moving boundary condition models* (figure 8(A)). These models aim to predict the evolution of the cell boundary, thereby predicting both motion and shape. The motion of the boundary can be described by physical models of molecular processes, including polarity signalling [214–217], hydrostatics and membrane tension [218], and actin network

dynamics [219, 220]. Zooming out from implementations of cell shape dynamics that assume specific biophysical mechanisms, a popular model providing an effective formulation of cell shape dynamics is the *cellular Potts model* [25–33] (figure 8(B)). In this lattice-based model, each cell is described by a set of lattice sites, and the cell shape is evolved by addition and removal of lattice sites based on an energy function which effectively models cell interfacial tension (perimeter elasticity) and a preferred cell area. To implement migration, the energy additionally includes a polarity term. In contrast to this effective energy-based approach, the *phase field model* describes the cell as a field  $\Phi(x, t)$  that is equal to 1 inside the cell and 0 outside [34–39] (figure 8(B)). Cell shape dynamics are then simulated through evolution of this field and coupled to cell polarity and cell velocity through force balance. Other computational cell migration models include the evolution of a cell contour function [221], models that are based on fibre network implementations of the cytoskeleton [222] and particle-based models with stochastic adhesions to a fibrous ECM [209].

To connect these computational models to experiments, a key challenge is that these models often have many parameters that are difficult to constrain based on experimental data.

Furthermore, if a specific mechanism for the cell shape evolution is assumed, it remains unclear how to systematically rule out alternative explanations. Indeed, in a systematic study comparing various computational approaches with different underlying assumptions [223] showed that all these models were able to faithfully capture the migration and typical shape of keratocytes. This apparent degeneracy of possible mechanisms may be a consequence of real redundancy in biological mechanisms responsible for a given behaviour. However, this observation also points to a problem with using complex bottom-up approaches for conceptual insight, as their parameters may remain under-constrained based on phenomenological observations. Therefore, we argue that connecting these computational models to inference approaches can be a promising path to constrain and better understand these models. Specifically, there is no reason why data-driven inference should only be applied to experimental data: we can similarly simulate a computational model, record the trajectories, and learn the effective equation of motion from simulated data. This learned effective model may then provide a much stronger constraint when comparing to the learned model from experimental data. Note that this approach is not limited only to nucleus or centre-of-area trajectories of cells, but can equally be applied to more complex cellular features, such as cell shapes, as discussed in the next section. Besides constraining parameters, this connection may allow both better insight into the emergent behaviour in the mechanistic model, as well as the mechanistic basis of a learned behaviour. For example, simulating a cellular Potts or phase field model in confining geometry would allow us to infer how the parameters of these models determine the response of the cell to the confinement; and conversely which mechanistic ingredients are relevant for setting the observed response in the experiment. Thus, connecting these approaches could yield a much more principled approach for constraining mechanistic models and understanding their emergent behaviours.

An alternative route to computational models has been to coarse-grain further and develop simplified, often 1D descriptions of cell polarity and migration dynamics based on underlying physical principles. First, actin polymerisation and retrograde flow have been described using continuum theories [43–48, 50], which may be coupled to advection-diffusion models of polarity cue concentrations [13, 51–54] (figure 8(D)). Such models have also been extended to account for adhesion-independent cell migration in structured systems where cells actively use friction with the walls or the local topography of the environment to self-propel [44, 140, 224]. Secondly, the molecular clutch model [40] describes the stochastic binding and unbinding of adhesions and their coupling to actin flows (figure 8(E)). Simulation models based on the molecular clutch can predict whole-cell trajectories [225, 226]. More minimal approaches extended the model to account for the mechano-sensitive binding dynamics of focal adhesions [55, 56, 58], predicting cell behaviours featuring periodic extension-retraction cycles, that have been observed as so-called ‘stick-slip processes’ in the biological literature [57, 227]. Coarse-graining further, the most

minimal models describe cells as active particles with a polarity (figure 8(F)). Interestingly, such active particle models can be directly derived from active gel theories, providing a mapping between the two [49].

These 1D cell migration models have given insight into some of the key cellular behaviours observed experimentally. However, it still remains difficult to make predictions for the full stochastic dynamics of cells, and in particular how the intra-cellular mechanisms to structured environments. To connect these models to inference approaches, a promising avenue may be to hierarchically coarse-grain minimal models into a description that is sufficiently simple that it can be inferred directly from experimentally accessible DOFs, such as descriptions of the  $(x, v)$ -dynamics of the cell. Specifically, an active particle model describing the cell position and polarity as DOFs may be coarse-grained into an equation for the underdamped dynamics of cell velocities, as shown in equation (10). Thus, the inferred description (e.g. the function  $F(x, v)$ ) can be matched to the dynamics predicted by more interpretable active particle model. This can be challenging as the inferred functions can contain large amounts of features, some of which may be more relevant than others. Furthermore, the functions may not always have a simple analytical form, making the derivation of an exact description difficult. A challenge for future work is therefore to identify ways to link inferred descriptions to bottom-up models in a principled way. Going further, an active gel or molecular clutch model could be mapped into an active particle model, and thereby indirectly linked to the inferable underdamped equation of motion for the cell. For instance, in [228], we provided a mapping between a model for the coupled dynamics of cell nucleus, protrusion and polarity and the underdamped equation of motion of the nuclear dynamics alone. Such mappings will be very useful in providing conceptual links between different models, and may help to test and constrain existing models, in particular when they are generalised to non-trivial external confinements.

A central challenge for bottom-up cell migration modelling is to link the different types of modelling approaches to each other. Ideally, there should be clear mappings between models, allowing to explicitly contrast assumptions and predictions of different approaches. Furthermore, as different models allow descriptions at different levels of detail, there should be a correspondence between the length- and time-scales of the experimental observations or the behaviour of interest, and the type of model employed. Having consistent mappings between models would then allow switching from one model to the other without contradictions. However, the principles of how to link these models together may still be outstanding. To address this, inference methods could help leverage rapidly increasing experimental data sets to constrain how these models fit together. Indeed, beyond inference from cell migration trajectories, expanding the experimentally tracked DOFs, and performing inference on other cellular features such as protrusions, polarities, traction forces or actin flows directly may provide an important tool, which we will discuss in the next section.

## 6.2. Inference from cellular features

To gain insight into cell migration mechanisms, a promising emerging avenue is to apply inference approaches to cellular features beyond just cell trajectories, such as imaging of the cell or nucleus shapes, the cytoskeleton or concentrations of focal adhesion and polarity molecules. While we can obtain a self-consistent description of the migration dynamics from the underdamped dynamics inferred from the cell trajectories, this effectively treats all intracellular DOFs as hidden variables. The key hidden variable in the simplest bottom-up model, the active particle model, is the cell polarity. However, cell polarity is notoriously hard to define, and there is no unique, generally agreed-upon molecular marker for cell polarity. In this section, we will discuss various approaches to infer dynamical models from cellular features of increasing detail, starting from cell shapes, to protrusions, polarity markers and traction forces, and point out the recurrent challenge of linking these features to the polarity of the cell.

Analysing the dynamics of cell shapes is attractive as cell shapes are easy to observe experimentally, as simple bright-field microscopy paired with modern segmentation pipelines based on machine learning [240, 241] can provide high accuracy, high throughput shape data sets. Clearly, developing a model for the entire cell shape as a function of time is a large jump in complexity from the 1D cell trajectories we have considered so far, and a low-dimensional stochastic equation of motion may not suffice to capture these dynamics. Thus, a first challenge is to determine the dominant contributions to the cell morphology through dimension reduction. By identifying the principle components of the cell shape, recent works have proposed to study cell morphology in a low dimensional space feature space [235–238] (figure 8(H)). From an analysis point of view, these approaches have been very successful by demonstrating that clustering in shape space can be predictive of metastatic potential [242, 243], stem cell lineage decisions [244] and drug response [235], highlighting the rich information content of cell morphologies. Furthermore, morphodynamic feature approaches have allowed comparative mapping of different cell types [245], identification of migration strategies in 3D matrices [238, 246], and revealed adaptive switching between different modes of mesenchymal migration [247].

However, to build dynamical models of cell morphology, there are two key challenges: (i) to establish a self-consistent shape space for morphodynamics, meaning that future morphological features can be predicted based on the current features, meaning the space is constructed such that the dynamics are Markovian; and (ii) to predict whole-cell motion from the evolution in shape space. Interestingly, in the case of *Dictyostelium* cell morphodynamics, just three principle morphological components on the second timescale were found to be predictive of migration behaviours on the minute timescale [236]. In contrast, for neutrophil migration, morphological features obtained from different dimension reduction techniques were shown to be insufficient to predict the migration velocity of the cell, suggesting that additional

information beyond the shape is required to capture cell polarisation in general [237]. Thus, to fully resolve cellular dynamics at the level of cell shapes may require the addition of further cellular DOFs which contain polar information.

To include polar information in a morphodynamic feature space, one could include information about the intra-cellular organisation, such as the relative position of the cell nucleus or the traction forces, or the velocity of the cell shape. Shape velocity is simply the derivative of the cell shape, corresponding to a ‘ribbon’ of alternating protrusion and retraction areas around the cell. At the scale of individual protrusions, morphodynamic profiling of shape velocities has revealed drastic spatiotemporal heterogeneity at the time-scales of minutes to hours, far below the time-scale of migration [248, 249], suggesting that high time-resolution experiments are likely to give most insight. To extract the key information of these shape velocities, an alternative approach is to simplify protrusion and retraction areas into a protrusion and retraction ‘centre of mass’ of the cell, i.e. a 1D readout (figure 8(K)).

In previous work, we extracted such protrusion trajectories from confined migrating cells, and inferred the coupled dynamics of cell nucleus and protrusion motion [228]. Interestingly, considering only nucleus and protrusion motion was not predictive of cell motion, and thus a time-correlated polarity-driven protrusion formation was required to capture the dynamics. Constraining the description of these polarity dynamics based on the observed protrusion trajectories revealed that the cell polarity is sensitive to the local geometry of the confinement. Specifically, under strong confinement, the polarity dynamics switches from a negative to a positive, self-reinforcing feedback loop. This geometry adaptation effect leads to a stereotypical cycle of protrusion extension into the constriction, followed by contraction and transmigration of the cell nucleus. The model then predicted, in agreement with experiments, that the protrusion–nucleus cycling disappears when the constriction is removed. This suggests that the positive polarity feedback loop emerges as a consequence of an adaptation of the cellular dynamics to the presence of the thin constriction. By performing inference on data-sets with cellular features beyond the cell nucleus, this approach resulted in equations of motion with mechanically interpretable terms, including the nucleus adhesiveness, the mechanical nucleus–protrusion coupling, and the coupling of cell polarity to protrusion confinement. Importantly, this model also correctly predicted the inferred underdamped dynamics of the nucleus trajectories alone, providing a link between the more phenomenological approach at the nucleus level to the intracellular polarity dynamics. In further work [250], the more interpretable protrusion–nucleus model was then used as a prediction target for a mechanistic model of confined cell migration, based on a generalised molecular clutch approach [56]. This revealed how membrane tension, actin alignment, and polarity cue diffusion interplay to generate the geometry adaptation effect.

Beyond cell shape and protrusion dynamics, the shape of the cell nucleus can give important insights into the forces acting during cell migration in 3D confining systems. In cell migration through tight 3D channel confinements

(unlike the flat 2D micropatterns discussed in the previous paragraph), the deformation and translocation of the cell nucleus has been shown to be a rate-limiting step in migration [155, 251–254]. To understand how the mechanics of nucleus deformation controls the migration dynamics, measuring the deformation forces acting on the nucleus could yield important insights. However, direct measurement of these forces acting is experimentally challenging. To circumvent this problem, recent work proposed a data-driven approach to infer the deformation force field acting on the nucleus directly from the observed nucleus shapes, relying on a mechanical model of the nucleus as either an elastic solid or an elastic shell [255]. Interestingly, varying the width of the constriction of 3D compliant hydrogel channels in a system mimicking the two-state geometry introduced in section 3.2 revealed that transmigration speeds exhibit a biphasic dependence on width, but with similar non-linear deterministic dynamics as on 2D micropatterns [256]. Specifically, there was an optimal width comparable to the nucleus diameter where transitions are fastest, which suggested a transition from a pulling- to a pushing-dominated mechanism. Another approach proposed an inferred model of coupled cell migration and nucleus geometry dynamics, resulting in a predictive model across geometries [257]. In the future, these inferred forces and dynamics could be used to constrain bottom-up models of how the nucleus affects cell migration [258–263].

A key element that is lacking in inferred models of cellular features is a direct measurement of cell polarity. Potential definitions include the localisation of polarity cues such as Cdc42, Rac or Rho GTPases [264] (such as PBD-YFP, a reporter of Rac1/Cdc42 activation [167]), the localisation of nucleus–actin binding proteins [141], or the relative positioning of cell nucleus and organelles such as the Golgi apparatus and the microtubule organising centre [265, 266]. However, large-scale data sets of tracked molecular markers are rare, likely due to the challenges associated with photo-toxicity of the markers [267, 268]. An interesting route forward could be machine learning approaches that allow prediction of fluorescent labels based on brightfield microscopy images [269, 270]. Alternatively, a more accessible intracellular observable are the traction forces, i.e. the forces exerted by cell adhesions onto the substrate, which can be measured using traction force microscopy [271]. Indeed, relationships between cell velocity and asymmetries in the spatial distribution of traction forces have been established both at the single cell [57, 272] and the collective level [273].

To develop data-driven models of how traction forces are determined by underlying molecular concentration fields, recent approaches based on deep-learning provide a promising avenue. Specifically, recent work showed that such image-based deep-learning can be used to predict traction forces from 2D images of focal adhesion proteins [239] (figure 8(I)). This revealed that a single focal adhesion protein, zyxin, carries the vast majority of predictive power for traction force prediction. Using physics-constrained neural networks, this leads to an interpretable model of how a spatially varying focal adhesion concentration field couples to a mechanical model of the cell as an effective 2D active elastic gel adhered to

a substrate [229–231]. This approach of learning maps from intra-cellular features to cellular behaviours could provide an exciting perspective for building models of the complex relationships between molecular processes and cell behaviour.

At an even more molecular level, a common question is often how different classes of molecular players affect each other causally, in particular in protrusion formation and polarity establishment. This is hard to establish based on pharmacological or genetic perturbations, as this usually perturbs the entire network. An alternative, non-perturbative approach has been developed using a data-driven method that uses relative temporal correlations of signalling molecule recruitment and actin polymerisation within cell protrusions to infer regulatory networks in a perturbation-free manner [232–234] (figure 8(G)).

In conclusion, these examples show how inference from additional cellular features combined with bottom-up mechanistic models can help identify the mechanistic underpinnings of cell migration in complex environments.

## 7. Learning the collective dynamics of multicellular systems

In physiological contexts, cells do not only interact with their confining extracellular environment, but also with one another [274–276]. Cell–cell interactions allow cells to organise collective behaviours and thereby address tasks that they could not solve on their own, such as shaping an embryo or healing a wound. Cellular interactions depend on complex molecular mechanisms, including cadherin-dependent pathways and receptor-mediated cell–cell recognition [204, 205, 277–281]. These mechanisms can lead to well-defined, stereotypical cell behaviours upon collision. A prominent type of collision behaviour was discovered in the 1950s by Abercrombie and coworkers [282], and was termed *contact inhibition of locomotion* (CIL). CIL refers to the tendency of cells to retract their lamellipodia, repolarise, and migrate apart upon contact. While these observations were made in a simple cell culture on 2D substrates, the relevance of CIL for physiological processes was later demonstrated, for example in the development of the neural crest [204, 205, 283].

At larger scales, cell–cell interactions lead to coordinated collective migration, which has been described with a variety of physical modelling approaches. These include active hydrodynamic theories [284], vertex [285–287], mechanical [288], and mechano-chemical [289] models, cellular automata [27, 290], phase-field models [38, 39], as well as active particle models [291–296] (see [297–300] for reviews). These modelling avenues typically make *a priori* assumptions on the types of interactions between individual cells, and therefore classify as bottom-up approaches. Cell–cell interactions are frequently modelled using repulsive potentials as an implementation of excluded volume interactions, alignment terms [292–295], or explicit implementations of CIL-like reorientation events upon collision [291, 296]. However, in these approaches, the structure of these interactions are usually assumed based on physical intuition (i.e. they are

bottom-up models), not derived directly from experimental data. Deriving cell-cell interactions directly from data could have several advantages. Bottom-up models can exhibit model degeneracy, meaning that multiple possible mechanistic interactions can reasonably well capture the qualitative cell behaviours. Furthermore, interacting cells can exhibit complex and unexpected types of interactions, which might be missed in bottom-up models that are limited to physics-inspired interactions such as alignment and attraction/repulsion interactions. Therefore, bottom-up models could be complemented by top-down inference of interactions directly from data, providing stronger constraints on such models.

We want to highlight three key hurdles that make the development of data-driven approaches for cell-cell interactions difficult. First, inference in interacting active many-body systems is technically challenging. To perform inference on such high-dimensional stochastic systems, a number of approaches have been developed that we discuss in section 7.1. Second, the structure of the interactions between cells may be substantially more complex than typically encountered in active matter theory, such as nonreciprocal interactions or interactions that adapt and change over time. Third, the complexity of the biological settings in which cell-cell interactions take place make it difficult to disentangle the distinct contributions of single-cell behaviour, interaction with the local micro-environment, cell proliferation, and cell-cell interactions. To overcome these problems, studying interacting cells in simplified artificial environments is a promising direction, which we discuss in section 7.2.

### 7.1. Inference approaches for interacting active systems

The inference of interactions from experimental tracking data has been a subject of interest in the field of animal behaviour for a long time [301]. The basic problem in inferring collective animal behaviour is very similar to the challenges faced in collective cell migration data, suggesting that these fields could learn from or even help each other. Specifically, in both systems, the basic problem is how to estimate the response of individuals to the presence of another individual as a function of their relative distance and orientation. In the context of the social interactions of fish a number of approaches to address this problem have been developed [302–304]. These inference approaches mainly focused on zonal interaction models which infer how the animal response varies as a function of angle at which another animal is observed, due to the key role of the field of vision in animal interactions.

To learn a predictive model of interacting cellular systems, we require approaches that can infer the interacting equations of motion of the system, which include both single-cell behaviour and interacting terms. A number of such approaches have been developed for deterministic systems [305, 306], as well as for stochastic systems in the context of animal behaviour [88, 307], and more generally for interacting stochastic active particle systems [85, 86]. Performing inference on collective systems is challenging due to the high dimensionality of the problem: a 3D swarm of  $N$  particles has  $6N$  DOFs (counting only positions and velocities, although more

variables may be relevant), and ‘curse of dimensionality’ arguments make this problem seem intractable. Indeed, the very simple approach of grid-based binning of the phase space (section 3.3) is unfeasible as it would require  $\gg 6N$  parameters to accurately represent the dynamics, therefore necessitating prohibitively large data sets to constrain such an approach. To overcome this problem, the trick is to adapt the selection of basis functions in such a way that the inference problem becomes effectively low-dimensional and thereby tractable. In the context of a basis expansion (equation (11)), we can think of binning as using top-hat basis functions at regularly spaced locations in the phase-space. To make better choices for interacting systems, essentially only step 2.2 of the inference procedure in section 4.1 has to be adapted, while the other steps are largely unaffected. The key idea is to simplify the inference by assuming symmetries of the interactions that reduce the number of fitting parameters.

The first important simplifying assumption is to treat particles as identical, such that all particles obey the same equation of motion. With this assumption, systems with more particles actually become effectively easier to infer from, as there is more data per recorded time-step. As proposed in [85, 86], a natural choice of basis functions is then to expand the deterministic contribution to the underdamped dynamics of cell  $i$  as a sum of one-body and two-body (interaction) terms:

$$\mathbf{F}_i \approx \sum_{\alpha} \mathbf{F}_{\alpha}^{(1)} c_{\alpha}^{(1)}(\mathbf{x}_i, \mathbf{v}_i) + \sum_{\beta} \mathbf{F}_{\beta}^{(2)} \sum_{i \neq j} c_{\beta}^{(2)}(\mathbf{x}_i, \mathbf{v}_i, \mathbf{x}_j, \mathbf{v}_j) + \dots \quad (16)$$

where higher orders  $\mathbf{F}^{(3)}, \mathbf{F}^{(4)}, \dots$  can in principle be included to account for multi-body interactions. This provides a generalisation of the basis expansion in equation (11) to interacting systems.

The second important assumption is to choose two-body basis functions  $\{c_{\beta}^{(2)}\}$ , which reflect additional symmetries of the interactions, such as radial symmetry of the interactions. In that case, all position dependence simply becomes a function of the radial distance  $r_{ij}$  of each pair of cells. A further promising approach is the inclusion of small convolutional neural networks as the basis functions of the expansion (equation (16)), which has been applied to the case of interacting active colloids [308]. An advantage of this approach is that it may reduce the risk of overfitting and provide a flexible basis for complex interaction functions. Furthermore, combining symmetry-based constraints with sparse regression [78, 79, 87, 89–91] could provide an avenue to constrain interacting cellular systems, which was recently demonstrated for simulated models of collective cell migration [309]. As an alternative approach, physics-informed graph neural networks have been proposed and applied to epithelial monolayers [310]. In this case, instead of the continuous phase space coordinates  $\{\mathbf{x}_i, \mathbf{v}_i, \mathbf{x}_j, \mathbf{v}_j\}$ , graph-based features were used as the argument of the interaction terms in equation (16). Together, these approaches allow decomposing the inferred dynamics into single-cell and interaction terms and imposing symmetry constraints, and can therefore provide a solution for the curse of dimensionality problem. This could facilitate inference of

cell–cell interactions from experimental collective migration data.

In the context of collective cell migration, a key objective is to identify different classes of interactions that may have biological interpretation. First, it is important to distinguish between two classes of interactions: **positional (isotropic) interactions**, which only depend on the relative position of two cells, such as excluded volume (repulsion) or adhesion (attraction) interactions. In contrast, **orientational (anisotropic) interactions** couple the directional properties of cells to each other. These can be either nematic, such as cell elongation, planar cell polarity, or myosin distributions in epithelia, or polar, such as polarity or velocity vectors. Examples for this are polarity and velocity alignment, stress–polarity coupling, and CIL [299]. Importantly, these different types of interactions imply different couplings of the position, velocity and polarity vectors of the cells. Thus, inferring the structure and parameters of cell–cell interactions with a general ansatz such as equation (11) directly from experimental data could then give insight into which of these interactions are at play.

In the discussion above, cell collectives are approximated as interacting active particles. The validity of this description may depend on the biological context. For low density, non-confluent assemblies of cells with transient interactions (collisions) between cells active particle model have been shown to provide a good approximation to the dynamics [296, 311]. While active particle models have also been successful in describing confluent cell monolayers [292], in this case, the contact geometry between neighbouring cells is important for cell–cell interactions, which is defined by the cell shapes. Thus, accounting for shape is likely more important in the confluent than in the non-confluent regime. Furthermore, cells are consistently connected to each other, implying that interactions may be governed by the topological graph connecting the cells (as a function of neighbour relationships) rather than metric interactions (as a function of distance). Such epithelia are often described using the so-called vertex model, in which cell shapes are represented by a set of vertices that denote the common point of three or more neighbouring cells [286, 287, 312]. The commonly used energy function of this model contains area and perimeter terms, which depend on the vertices connected to the (on average) six neighbouring cells. A direct inference of the parameters and interactions of such models could yield novel insights into how tissues control the collective states of cells. In general, expanding the overdamped equivalent of equation (16) could account for these interactions, but it will not provide an adapted basis to the geometry of the problem. Thus, fitting a set of hypothesised dynamical terms (such as the vertex model) could be a better approach. However, cellular states in epithelial tissues are known to undergo dynamical changes, such as changing myosin distributions, and remodelling of the cell edges [10, 313], with key implications for tissue dynamics [314–316]. This implies that the dynamical parameters are non-stationary in time, making the inference problem significantly more difficult.

One avenue to circumvent this problem is to use a basis of modes that uses the geometry of the epithelial layer more directly [314, 317]. This approach applies in a regime that is dominated by tension in the cell edges, and that exhibits a time-scale separation between the tissue dynamics and the relaxation of individual cell edges, implying that the vertices are in force balance. Under these assumptions, the tissue dynamics decomposes into two independent contributions: the tensions determine the dynamics of the angles (at fixed areas), while isogonal modes predict the evolution of areas (at fixed angles) [314]. Due to force balance at each vertex, the tensions can then be directly inferred from the experimentally observed angles at each vertex [317]. Thus, in this decomposition, the tensions are no longer (hidden) parameters that have to be fitted indirectly, but can be measured directly from the images. Incorporating simplifying assumptions therefore allowed a decomposition of the dynamics that makes it feasible to reliably infer time-dependent cellular adaptation. We will discuss the implications of such tension inference in more detail in the next section.

Beyond these models of cells as discrete entities, active polar or nematic hydrodynamic models provide important conceptual frameworks to describe cellular assemblies. To learn such models from observed data, inference and machine learning approaches for active nematics have been developed and applied in the context of *in vitro* microtubule assays [318, 319], and active polar particle experiments [320]. These approaches use the observed velocity fields or cell tracking data to uncover the hydrodynamic equations governing these active matter systems, which could provide a promising approach for inference from collective cellular systems.

The development of inference approaches that can be applied to stochastic interacting active systems opens up new avenues to learn the dynamics of cell–cell interactions directly from observed data, and we will review such applications in the next section.

## 7.2. From cell pairs to collective migration

Having discussed approaches to make inference from active interacting systems tractable, we turn to the second challenge: reducing the biological complexity of interacting cellular systems, to make the decomposition into various contributions from single-cell dynamics, proliferation, and interactions tractable. To achieve this, a broad variety of *in vitro* approaches to confine groups of cells to defined geometries have been developed. In some cases, this confinement is kept for the entire duration of the experiment, while in others it serves as an initial condition from which cell spreading is observed. To highlight how cell–cell interactions can be inferred from these different approaches, we will systematically go up in the complexity of the experimental approaches, such as increasing cell number and dimensionality of the system, and highlight how interaction inference was or could be applied to these systems (figures 9(A)–(I)). Note that our discussion here is primarily focused on substrate-dependent migration, where

cells are exerting active migration forces on a substrate in the direction of their polarity. We will briefly touch on the case of substrate-independent tissue flow due to active stresses exerted between the cells through their junctional actomyosin (rather than between cells and substrate via protrusions) at the end of this section.

Many systems that are controlled by cell–cell interactions rely on the concerted dynamics of small groups of cells, including the dynamics of pairs of cells [204, 321] and migration of small clusters [322, 323] in developmental systems, as well as migrating tumour clusters of up to eight cells [324, 325]. Studying cell–cell interactions at fixed total number of cells  $N$  in simplified *in vitro* systems provides a major simplification, as the dynamics of cell proliferation can be neglected, and allows building complexity step-by-step.

The simplest possible cell–cell interaction system is to keep  $N = 2$  and study the collisions of pairs of cells. This has been done by studying cell collisions on 1D micropatterned tracks [96, 97, 148, 326, 327], microfluidics [137], and suspended fibres [144]. Furthermore, cell pairs have been confined into closed confinements in which they continuously interact over long periods of times, including circular [328] and two-state micropatterns [98] (figures 9(A) and (B)). An advantage of such closed confinements is that cells interact with each other repeatedly in a standard environment, leading to long interacting trajectories.

In previous work, we applied stochastic inference to such interacting cell pair trajectories to learn the positional and orientational interactions of the cells. To separate these interactions, we used a simplified version of equation (16) and postulated that the dynamics of the system can be described by the following equation of motion:

$$\frac{dv}{dt} = F(x, v) + f(r) \Delta x + \gamma(r) \Delta v + \sigma \eta(t) \quad (17)$$

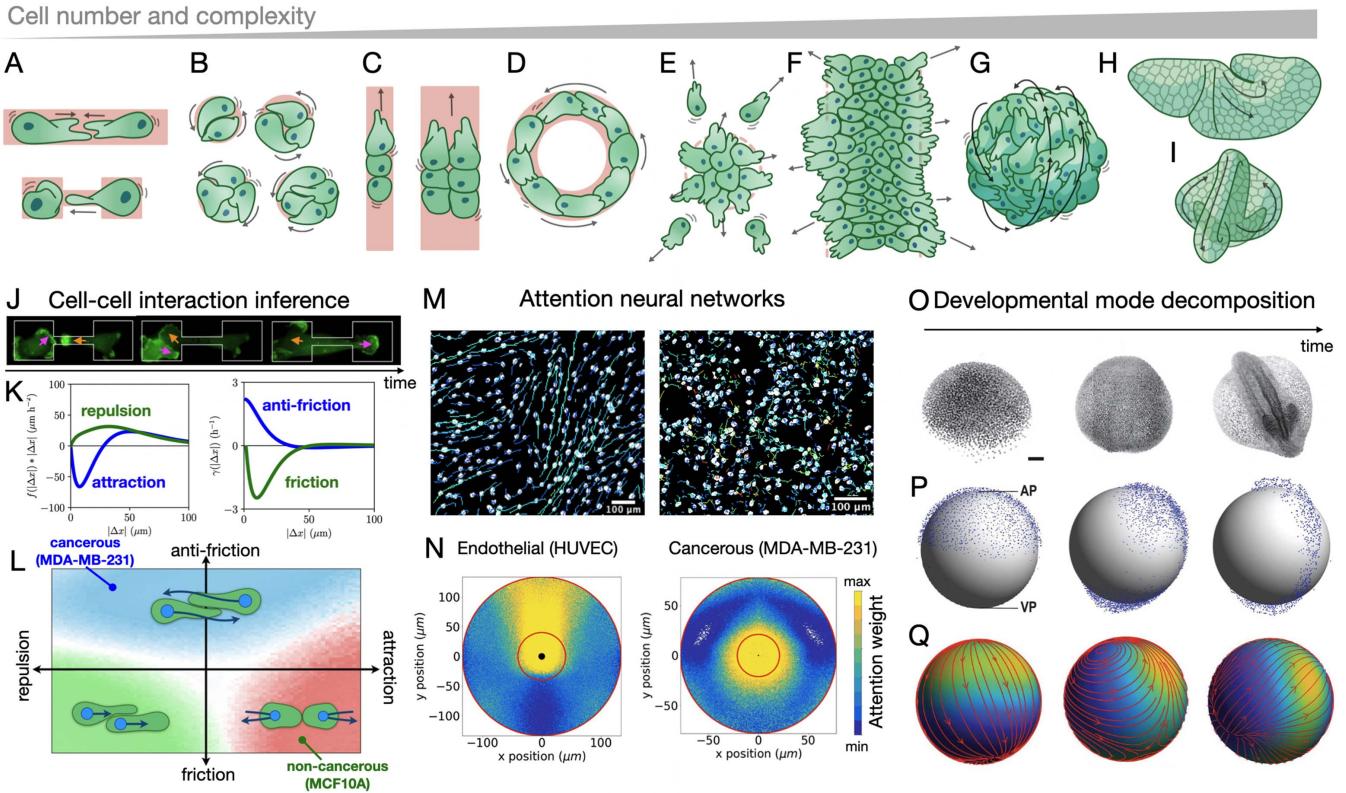
where  $r = |\Delta x|$  is the distance between the two cells and we simplified the dynamics to one dimension since cells move predominantly along the  $x$ -direction of the pattern. This approach assumes that the deterministic dynamics of the system can be decomposed into two separate components: a single-cell term  $F(x, v)$ , similar to that inferred from single-cell experiments, and interactive components, which depend on the relative position  $\Delta x$  and the relative velocity  $\Delta v$  of the cells. The term  $f(r)$  thus represents positional cell–cell interactions such as repulsion and attraction. In contrast,  $\gamma(r) \Delta v$  depends on the relative motion of the cells, and is therefore an orientational interactions with the mathematical form of an effective frictional interaction. For  $\gamma < 0$ , this interaction accounts for alignment between cells, as it seeks to minimise differences in relative velocity [292, 293]. To infer this model (equation (17)), we use a basis expansion of the two-body terms  $\{c_\beta^{(2)}\}$  that assumes radial symmetry of the interactions as the kernels  $f$  and  $\gamma$  only depend on the distance  $r$ . The basic inference procedure for this case is the same as for single cells: we infer the model using a suitable basis expansion, and then make predictions for long time-scale statistics of the interacting dynamics, which match those

observed experimentally. Importantly, we also find that the single-cell term inferred from two-cell experiments matches that inferred from single cell experiments, suggesting that the separation of interactions and single-cell behaviour was successful.

Interestingly, the inference revealed that a non-cancerous (MCF10A) and a cancerous (MDA-MB-231) breast tissue cell line exhibit distinct types of interactions: while the MCF10A cells exhibit repulsive and regular frictional interactions, the MDA-MB-231 attract at short distances and exhibit a positive friction term ( $\gamma > 0$ ) (figure 9(K)). This ‘anti-friction’ interaction ensures that rather than slowing down upon collision, cells deterministically accelerate, leading to the characteristic sliding events observed for this cell line. The model (equation (17)) furthermore suggests an ‘interaction behaviour space’, which relates the physical interaction terms to the cell–cell collision behaviour of the system, suggesting that this framework could potentially describe various cell–cell interaction modes known in the biological literature [97, 205, 326, 331, 332], including reversing, sliding and following interactions (figure 9(L)). These inferred interactions could in the future provide constraints for bottom-up models of cell pair collisions [144, 333].

Going beyond cell pair dynamics, several studies have systematically investigated the effects of increasing the number of interacting cells one by one. First, on circular micropatterns going from  $N = 2$  to  $N = 8$  revealed cell number-dependent rotational behaviours [27] (figure 9(B)). Secondly, a number of studies have investigated the behaviour of small clusters of cells on linear and circular microstripes [334–339] (figures 9(C) and (D)). For instance, in trains of keratocyte cells confined to 1D stripes, the speed of the train was shown to be independent of the number of cells in the axial direction, i.e. parallel to the direction of motion [337]. In contrast, train speed decreased with cell number in the lateral direction, i.e. orthogonal to the direction of motion. This observation constrained a bottom-up active matter simulation to identify the cell–cell interactions in the clusters.

The approach of using experimental data to constrain the interactions in bottom-up active particle simulations has been successful in a broad variety of larger scale systems at the scale of hundreds of cells. One strategy to regularise collective behaviour has been to confine cells to a circular confinement which is then released by mechanical or chemical means to allow spreading of cells (figure 9(E)). In *Dictyostelium* colonies, this revealed that a new type of interaction had to be included in an active particle model for the system, which enhances rather than inhibits motility upon collision and was termed contact enhancement of locomotion [296]. In cancer cell colonies, a similar experimental and theoretical approach revealed how E-cadherin junctions control excluded volume interactions between cells by ‘sharpening’ inter-cellular boundaries [311]. At the scale of confluent monolayers of cells (figure 9(F)), the cell–cell interactions were captured employing active matter models, using the velocity distributions and correlations as a constraint [292]. In these large-scale systems, direct inference of cell–cell interactions is challenging, although recent work proposed a systematic fitting procedure of a Viscek-type



**Figure 9.** Learning cell–cell interactions from collective cell migration experiments. (A)–(I) Systems to study collective cell migration with increasing number of cells and complexity: cell pairs (A); circular micropatterns (B); linear (C) and periodic (D) microstripes; collective escape from a confined area (E); spreading monolayers (F), 3D spheroids (G) and embryos such as fly (H) and zebrafish (I). (J) Time series of two interacting MDA-MB-231 cells transfected with LifeAct-GFP. Arrows highlight regions of pronounced actin activity, and the arrow colour indicates the cell identity. (K) Inferred cohesive interaction term  $f(r)\Delta x$  (left), and effective frictional interaction term  $\gamma(r)$  for non-cancerous MCF10A cells (green) and cancerous MDA-MB-231 cells (blue). (L) Interaction behaviour space predicted by varying the amplitude of the cohesive and friction interactions. Colour and cartoons indicate the dominant collision behaviour at each point, and the inferred locations of both cell types are indicated. Reproduced with permission from [98]. CC BY-NC-ND 4.0. (M), (N) Representative cell trajectories within living tissues, from human umbilical vein endothelial cells (HUVEC, left) and metastatic human breast cancer cells (MDA-MB-231, right). Scale bars: 100  $\mu\text{m}$ . Normalised attention weight contour plots for both cell types. Red circle: radius of the closest neighbour location. Reproduced from [329]. CC BY 4.0. (O)–(P) Mode decomposition of early zebrafish development. Microscopy images (O, scale bar: 100  $\mu\text{m}$ ) and tracked cell positions (P, blue dots) as a function of developmental time. (P) Density (colour) and flow fields (arrows) corresponding to the same images. Reproduced from [330]. CC BY 4.0.

alignment model to collective monolayer migration [340]. A key challenge in this context is twofold: firstly, trajectory data of sufficient quality is required to perform inference. Specifically, cells disappearing and reappearing from the tracking are a problem when inferring particle-based interactions, since all cells that are present should be considered to infer the interactions. Secondly, a formulation of equation (16) has to be found that is on the one hand flexible enough to capture potentially complex types of cell–cell interactions (which for instance do not have to obey radial symmetry), and on the other hand be restricted enough to allow accurate inference. Here, combining stochastic inference with sparsity constraints may be a way forward [78].

To test the potential symmetries of cell–cell interactions, a data-driven approach for cells in 2D monolayers using attention neural networks was recently proposed [329] (figures 9(M) and (N)). This approach detects how predictive the behaviour of neighbouring cells is for the behaviour of a given cell. In the case of radially symmetric interactions,

such as in a Viscek-type alignment model, this attention map should be radially symmetric. However, this approach revealed that the single-cell response of fibroblast and epithelial cells are mainly affected by interactions with the neighbours ahead of them in the direction of motion (figure 9(N)). In contrast, the interactions appeared isotropic in cancer cell collectives, showing how different cell types may exhibit different types of interacting symmetries.

Understanding cell–cell interactions becomes significantly more complicated in 3D systems, where cell migration often occurs on complex, curved surfaces. Therefore, data-driven theoretical approaches to these systems have primarily relied on a tissue-level or continuum description, rather than cell-resolved analysis as considered in the previous examples. Based on this, data-driven approaches for these systems often seek to decompose the dynamics into a low-dimensional set of modes, which we discuss next.

To study 3D collective migration *in vitro*, minimal systems include cylindrical and spherical confinements (figure 9(G)).

In the case of spheroids, migrating cells usually setup global rotations of the tissue [341–345]. Based on experimental cell trajectories, these rotational velocity fields could be decomposed into the basic mode of a rotational velocity field and the fluctuations in the co-rotating frame [344]. In cancer organoids, this revealed travelling velocity waves with vortex flows [344], while in pancreas spheroids, a chiral velocity field was identified [345]. These collective modes could then be recapitulated with active cell migration models confined to the sphere, demonstrating that these modes are a generic response of active polar dynamics of cells to curvature.

The problem of inferring cell–cell interactions becomes more complex in *in vivo* systems such as developing embryos. A popular model organism for cell migration in embryogenesis is zebrafish (figure 9(I)). During zebrafish gastrulation, the tissue performs major rearrangements using a range of biophysical processes, including guidance of cells by self-generated gradients [346], motility-driven unjamming [347], and ECM-independent cell migration [348]. While these various processes have been addressed with bottom-up active particle models, learning models from such embryo data could provide insight into the collective dynamics of this complex system [24]. This was recently done at the scale of the entire embryo by decomposing the motion of the cells into a set of low-dimensional ‘developmental modes’ [330] (figures 9(O)–(Q)). Specifically, density and flow fields were decomposed into a combination of basis functions (analogously to the basis functions for dynamical terms (equation (11))), using the spherical harmonics due to the shape of the system. The dynamics of these modes could then be described with equations of motion identified by sparse regression. This allowed direct inference of a hydrodynamic model, revealing similarities between whole-embryo cell migration and active Brownian particle dynamics on curved surfaces.

In addition to cell migration, tissue flows due to active stresses exerted between cells are a key feature of morphogenesis, such as in *Drosophila* gastrulation (figure 9(H)). A common approach here is to develop continuum tissue mechanics models of embryos that match constitutive relations of active materials with experimental observations [349–352]. In a further step, a machine learning approach was proposed to infer the coupled dynamics of the key molecular components (myosin and E-cadherin) and the morphogenetic flow [353]. To understand how these tissue-level processes are orchestrated at the cellular scale, recent work has proposed an approach to infer the active tensions driving tissue flow from the observed cell geometry based on the assumption of force balance [314, 317]. This allows a mode decomposition of the tissue dynamics into the dynamics of junctional angles (determined by tension) and cell areas (isogonal modes) (also refer to more general discussion in section 7.1). This also allowed disentangling active vs passive T1 transitions in the tissue, which have been shown to enable convergence extension-movement in large-scale tissue deformation processes such as during gastrulation [315–317]. In general, a key challenge to test how well these models are constrained and their predictive power is whether one can for instance

predict mutants, generalise the findings to other organisms, or make predictions for new experiments.

As a perspective for future research using data-driven approaches for interacting cells, we foresee two primary directions. On the one hand, data-driven approaches for cell–cell interactions could provide an avenue to better understand how molecular processes control interacting behaviours. One possible approach is to pharmacologically target molecular components that are known to be important in controlling cell–cell interactions, and then inferring the resulting change in dynamics. This may provide a way to identify the link between individual components with their role in the emergent behaviour. For example, a key question raised by the inference on two-cell collisions is how the separate positional and effective frictional components are controlled by molecular components. In particular, it is unclear what underlying mechanism controls the switch from friction or anti-friction interactions observed in non-cancerous and cancerous cells, respectively. Candidates are E-cadherin mediated cell–cell junctions, which are down-regulated in cancer cells [97], or ephrins, which play a key role in cell–cell recognition [283]. Furthermore, to understand the emergence of the repulsive interaction between cells, which is responsible for CIL, polarity cues, such as Rho GTPases, could be perturbed. These components are likely important in how cells change their direction of motion, an important process in the reversal events associated with CIL [96]. Thus, combining interaction inference with molecular perturbation in cell pair collision experiments could provide an avenue to link mechanisms and behaviour in interacting cellular systems.

On the other hand, learning cellular interactions could help constrain active matter models which can then be used to make predictions for new experiments and biological systems beyond the dataset the model was trained on. For instance, inferring cell–cell interactions from pairs of colliding cells (figure 9(A)) allows making predictions for these types of cells in more complex collective systems (figures 9(B)–(G)). This would allow testing whether collective systems are explainable based on two-body interactions, reciprocal interactions, and whether these interactions exhibit adaptation to their environment. Inference of cell–cell interactions can therefore provide an important route towards understanding the active matter physics of interacting cells.

## 8. Outlook

In this review, we have discussed how data-driven approaches make it possible to learn dynamical models of single and collectively migrating cells directly from experiments. The first part of this review on single-cell dynamics contained three main themes: how to infer and conceptualise models of cell behaviour in unstructured (free 2D) and structured (confining) systems; how to generalise these models to account for temporal and cell-to-cell variability in behaviour; and how to relate these behaviours to bottom-up models and underlying molecular mechanisms. While significant progress has been made in recent years to address these problems, much

remains unknown. The basic molecular mechanisms that drive migration are increasingly well understood, but an integrated understanding of how these mechanisms interplay to set the emergent stochastic behaviours of cells at long time-scales still remains elusive. Furthermore, it remains unclear how the motility and cytoskeletal machinery of cells respond to external confinements at the molecular level, and how these responses determine the emergent behaviour.

So far, data-driven models of cell migration dynamics, from persistent random motion, to confined cell migration and interacting cells, have often been limited to the treatment of low-dimensional sets of cellular DOFs, such as cell nucleus trajectories. These models could in the future provide important constraints for bottom-up biophysical models (see section 6.1). This could yield more interpretable, yet strongly data-constrained descriptions, of the underlying mechanisms. Another exciting avenue is to take data-driven approaches to a more mechanistic level by tracking and analysing subcellular DOFs, such as actin flows, polarity markers, or traction forces (see section 6.2).

In the second part of the review, we discussed how data-driven approaches for cell–cell interactions could provide an avenue to better understand how molecular processes control interacting behaviours by performing model inference on a range of genetic and pharmacological perturbations of underlying migration and interaction mechanisms. Inferring these interactions for a range of cell types and extra-cellular environment could reveal new and unexpected types of interactions, such as non-reciprocal, adaptive, or time-dependent terms. The inferred interactions could then allow simulation and prediction of the collective dynamics at larger scales. This would allow testing whether two-body dynamics are predictive of many-body dynamics. In systems where the inferred two-body interactions are non-standard, this would also allow exploring their consequences in more complex systems. Moreover, applying inference directly to collective migration datasets, allowing to disentangle the separate contributions of single-cell motility, cell–cell interactions, cell proliferation and external confinements.

An attractive perspective is that beyond providing tools for building predictive physical models and constraining underlying mechanisms, data-driven approaches can also help provide new conceptual insights. Cell migration requires the coordination of a multitude of molecular players on the cell level and of the behaviours of a large number of cell on the tissue scale, such as the equilibration to force balance in epithelia. Achieving such coordination has been proposed to occur on low-dimensional ‘slow’ manifolds in high-dimensional systems in a variety of contexts. An example for this are cell fate decisions, where high-dimensional expression profiles can be described by effective 2D dynamical systems [354, 355], as originally introduced through the idea of Waddington’s landscape [356]. Similar principles could potentially play a role in how cells and tissues coordinate their behaviours, and data-driven inference could allow us to identify the low-dimensional dynamical systems and attractor manifolds governing these behaviours.

These perspectives demonstrate how data-driven approaches have the potential to address key open questions in single and collective cell migration. A common thread in these ideas is that by applying data-driven inference to experimental datasets, we can complement models developed from the bottom up, by inferring models directly from data. Based on these inferred dynamics, we can then attempt to constrain underlying mechanisms, and predict emergent behaviours of the system.

## Data availability statement

No new data were created or analysed in this study.

## Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation)—Project-ID 201269156—SFB 1032 (Project B12). D B B was supported by an NOMIS Fellowship and an EMBO Fellowship (ALTF 343-2022). We thank Joachim Rädler, Alexandra Fink, Erwin Frey, Pierre Ronceray, Ricard Alert, Edouard Hannezo, Henrik Flyvbjerg, Ulrich Schwarz, Joshua Shaevitz, Greg Stephens, Andrea Cavagna, Grzegorz Gradziuk, Fridtjof Brauns, Nikolas Claussen, Tom Brandstätter, Johannes Flommersfeld, Christoph Schreiber, Nicolas Arlt, Matthew Schmitt, Joris Messelink, Federico Gnesotto, Federica Mura, Bram Hoogland, Manon Wigbers, Isabella Graf, Jessica Lober, and many others for inspiring discussions. We also thank Claudia Flandoli for the artwork in figures 1, 5, 8 and 9.

## ORCID iDs

David B Brückner  <https://orcid.org/0000-0001-7205-2975>  
Chase P Broedersz  <https://orcid.org/0000-0001-7283-3704>

## References

- [1] Franz C M, Jones G E and Ridley A J 2002 Cell migration in development and disease *Dev. Cell* **2** 153–8
- [2] Scarpa E and Mayor R 2016 Collective cell migration in development *J. Cell Biol.* **212** 143–55
- [3] Luster A D, Alon R and von Andrian U H 2005 Immune cell migration in inflammation: present and future therapeutic targets *Nat. Immunol.* **6** 1182–90
- [4] Li B and Wang J H-C 2011 Fibroblasts and myofibroblasts in wound healing: force generation and measurement *J. Tissue Viab.* **20** 108–20
- [5] Bainbridge P 2013 Wound healing and the role of fibroblasts *J. Wound Care* **22** 407–8, 410–2
- [6] Friedl P and Wolf K 2003 Tumour-cell invasion and migration: diversity and escape mechanisms *Nat. Rev. Cancer* **3** 362–74
- [7] Yamaguchi H, Wyckoff J and Condeelis J 2005 Cell migration in tumors *Curr. Opin. Cell Biol.* **17** 559–64
- [8] Paul C D, Mistriotis P and Konstantopoulos K 2017 Cancer cell motility: lessons from migration in confined spaces *Nat. Rev. Cancer* **17** 131–40

- [9] Bausch A R and Kroy K 2006 A bottom-up approach to cell mechanics *Nat. Phys.* **2** 231–8
- [10] Kasza K E, Rowat A C, Liu J, Angelini T E, Brangwynne C P, Koenderink G H and Weitz D A 2007 The cell as a material *Curr. Opin. Cell Biol.* **19** 101–7
- [11] Fletcher D A and Mullins R D 2010 Cell mechanics and the cytoskeleton *Nature* **463** 485–92
- [12] Broedersz C P and Mackintosh F C 2014 Modeling semiflexible polymer networks *Rev. Mod. Phys.* **86** 995–1036
- [13] Maiuri P *et al* 2015 Actin flows mediate a universal coupling between cell speed and cell persistence *Cell* **161** 374–86
- [14] Callan-Jones A C and Voituriez R 2016 Actin flows in cell migration: from locomotion and polarity to trajectories *Curr. Opin. Cell Biol.* **38** 12–17
- [15] Kruse K, Joanny J F, Jülicher F, Prost J and Sekimoto K 2005 Generic theory of active polar gels: a paradigm for cytoskeletal dynamics *Eur. Phys. J. E* **16** 5–16
- [16] Prost J, Jülicher F and Joanny J-F 2015 Active gel physics *Nat. Phys.* **11** 111–7
- [17] Jülicher F, Grill S W and Salbreux G 2018 Hydrodynamic theory of active matter *Rep. Prog. Phys.* **81** 076601
- [18] Pollard T D 2003 The cytoskeleton, cellular motility and the reductionist agenda *Nature* **422** 741–5
- [19] Brückner D B, Fink A, Schreiber C, Röttgermann P J F, Rädler J O and Broedersz C P 2019 Stochastic nonlinear dynamics of confined cell migration in two-state systems *Nat. Phys.* **15** 595–601
- [20] Fink A, Brückner D B, Schreiber C, Röttgermann P J, Broedersz C P and Rädler J O 2020 Area and geometry dependence of cell migration in asymmetric two-state micropatterns *Biophys. J.* **118** 552–64
- [21] Flamholz A, Phillips R and Milo R 2014 The quantified cell *Mol. Biol. Cell* **25** 3497–500
- [22] Rosen M E and Dallon J 2022 A mathematical analysis of focal adhesion lifetimes and their effect on cell motility *Biophys. J.* **121** 1070–80
- [23] Anderson P W 1972 More is different *Science* **177** 393–6
- [24] Schwayer C and Brückner D B 2023 Connecting theory and experiment in cell and tissue mechanics *J. Cell Sci.* **136** jcs261515
- [25] Graner F and Glazier J A 1992 Simulation of biological cell sorting using a two-dimensional extended Potts model *Phys. Rev. Lett.* **69** 2013–6
- [26] Glazier J A and Graner F 1993 Simulation of the differential adhesion driven rearrangement of biological cells *Phys. Rev. E* **47** 2128–54
- [27] Segner F J, Thüroff F, Piera Alberola A, Frey E and Rädler J O 2015 Emergence and persistence of collective cell migration on small circular micropatterns *Phys. Rev. Lett.* **114** 228102
- [28] Albert P J and Schwarz U S 2014 Dynamics of cell shape and forces on micropatterned substrates predicted by a cellular Potts model *Biophys. J.* **106** 2340–52
- [29] Albert P J and Schwarz U S 2016 Optimizing micropattern geometries for cell shape and migration with genetic algorithms *Integr. Biol.* **8** 741–50
- [30] Albert P J and Schwarz U S 2016 Modeling cell shape and dynamics on micropatterns *Cell Adhes. Migr.* **10** 516–28
- [31] Albert P J and Schwarz U S 2016 Dynamics of cell ensembles on adhesive micropatterns: bridging the gap between single cell spreading and collective cell migration *PLoS Comput. Biol.* **12** 1–34
- [32] Thüroff F, Goychuk A, Reiter M and Frey E 2019 Bridging the gap between single-cell migration and collective dynamics *eLife* **8** e46842
- [33] Goychuk A, Brückner D B, Holle A W, Spatz J P, Broedersz C P and Frey E 2018 Morphology and motility of cells on soft substrates (arXiv:1808.00314)
- [34] Kockelkoren J, Levine H and Rappel W-J 2003 Computational approach for modeling intra- and extracellular dynamics *Phys. Rev. E* **68** 037702
- [35] Zieber F, Swaminathan S and Aranson I S 2011 Model for self-polarization and motility of keratocyte fragments *J. R. Soc. Interface* **9** 1084–92
- [36] Shao D, Levine H and Rappel W-J 2012 Coupling actin flow, adhesion and morphology in a computational cell motility model *Proc. Natl Acad. Sci.* **109** 6851–6
- [37] Marth W and Voigt A 2014 Signaling networks and cell motility: a computational approach using a phase field description *J. Math. Biol.* **69** 91–112
- [38] Camley B A, Zhang Y, Zhao Y, Li B, Ben-Jacob E, Levine H and Rappel W-J 2014 Polarity mechanisms such as contact inhibition of locomotion regulate persistent rotational motion of mammalian cells on micropatterns *Proc. Natl Acad. Sci.* **111** 14770–5
- [39] Bertrand T, d'Alessandro J, Maitra A, Jain S, Mercier B, Mége R-M, Ladoux B and Voituriez R 2020 Clustering and ordering in cell assemblies with generic asymmetric aligning interactions (arXiv:2012.00785)
- [40] Chan C and Odde D 2008 Traction dynamics of filopodia on compliant substrates *Science* **322** 1687–91
- [41] Elosegui-artola A, Trepat X and Roca-Cusachs P 2018 Control of mechanotransduction by molecular clutch dynamics *Trends Cell Biol.* **28** 356–67
- [42] Kruse K, Joanny J F, Jülicher F and Prost J 2006 Contractility and retrograde flow in lamellipodium motion *Phys. Biol.* **3** 130–7
- [43] Callan-Jones A C, Joanny J-F and Prost J 2008 Viscous-fingering-like instability of cell fragments *Phys. Rev. Lett.* **100** 258106
- [44] Hawkins R J, Piel M, Faure-Andre G, Lennon-Dumenil A M, Joanny J F, Prost J and Voituriez R 2009 Pushing off the walls: a mechanism of cell motility in confinement *Phys. Rev. Lett.* **102** 1–4
- [45] Hawkins R J, Poincloux R, Bénichou O, Piel M, Chavrier P and Voituriez R 2011 Spontaneous contractility-mediated cortical flow generates cell migration in three-dimensional environments *Biophys. J.* **101** 1041–5
- [46] Blanch-Mercader C and Casademunt J 2013 Spontaneous motility of actin lamellar fragments *Phys. Rev. Lett.* **110** 078102
- [47] Khoromskaia D and Alexander G P 2015 Motility of active fluid drops on surfaces *Phys. Rev. E* **92** 062311
- [48] Bergert M, Erzberger A, Desai R A, Aspalter I M, Oates A C, Charras G, Salbreux G and Paluch E K 2015 Force transmission during adhesion-independent migration *Nat. Cell Biol.* **17** 524–9
- [49] Recho P, Putelat T and Truskinovsky L 2019 Active gel segment behaving as an active particle *Phys. Rev. E* **100** 1–15
- [50] Lavi I, Meunier N, Voituriez R and Casademunt J 2020 Motility and morphodynamics of confined cells *Phys. Rev. E* **101** 1–20
- [51] Gracheva M 2004 A continuum model of motility in ameboid cells *Bull. Math. Biol.* **66** 167–93
- [52] Doubrovinski K and Kruse K 2011 Cell motility resulting from spontaneous polymerization waves *Phys. Rev. Lett.* **107** 258103
- [53] Vanderlei B, Feng J J and Edelstein-Keshet L 2011 A computational model of cell polarization and motility coupling mechanics and biochemistry *Multiscale Model. Simul.* **9** 1420–43
- [54] Camley B A, Zhao Y, Li B, Levine H and Rappel W-J 2017 Crawling and turning in a minimal reaction-diffusion cell motility model: coupling cell shape and biochemistry *Phys. Rev. E* **95** 012401

- [55] Ron J E, Monzo P, Gauthier N C, Voituriez R and Gov N S 2020 One-dimensional cell motility patterns *Phys. Rev. Res.* **2** 1–27
- [56] Sens P 2020 Stick-Slip model for actin-driven cell protrusions, cell polarisation and crawling *Proc. Natl Acad. Sci. USA* **117** 24670–8
- [57] Hennig K *et al* 2020 Stick-slip dynamics of cell adhesion triggers spontaneous symmetry breaking and directional migration of mesenchymal cells on one-dimensional lines *Sci. Adv.* **6** 1–13
- [58] Schreiber C, Amiri B, Heyn J C, Rädler J O and Falcke M 2021 On the adhesion–velocity relation and length adaptation of motile cells on stepped fibronectin lanes *Proc. Natl Acad. Sci. USA* **118** e2009959118
- [59] Pearce P, Woodhouse F G, Forrow A, Kelly A, Kusumaatmaja H and Dunkel J 2019 Learning dynamical information from static protein and sequencing data *Nat. Commun.* **10** 5368
- [60] Jumper J *et al* 2021 Highly accurate protein structure prediction with AlphaFold *Nature* **596** 583–9
- [61] Imakaev M V, Fudenberg G and Mirny L A 2015 Modeling chromosomes: beyond pretty pictures *FEBS Lett.* **589** 3031–6
- [62] Abbas A, He X, Niu J, Zhou B, Zhu G, Ma T, Song J, Gao J, Zhang M Q and Zeng J 2019 Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes *Nat. Commun.* **10** 2049
- [63] Messelink J J, van Teeseling M C, Janssen J, Thanhichler M and Broedersz C P 2021 Learning the distribution of single-cell chromosome conformations in bacteria reveals emergent order across genomic scales *Nat. Commun.* **12** 1963
- [64] Gabriele M, Brandão H B, Grosse S, Jha A, Dailey G M, Hsieh T-h S, Mirny L, Zechner C and Hansen A S 2022 Dynamics of CTCF and cohesin mediated chromatin looping revealed by live-cell imaging *Science* **376** eabn6583
- [65] Mach P *et al* 2022 Cohesin and CTCF control the dynamics of chromosome folding *Nat. Genet.* **54** 1907–18
- [66] Brückner D B, Chen H, Barinov L, Zoller B and Gregor T 2023 Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome *Science* **380** 1357–62
- [67] Schneidman E, Berry M J, Segev R and Bialek W 2006 Weak pairwise correlations imply strongly correlated network states in a neural population *Nature* **440** 1007–12
- [68] Tkačik G, Marre O, Mora T, Amodei D, Berry M J and Bialek W 2013 The simplest maximum entropy model for collective behavior in a neural network *J. Stat. Mech. P* **03011**
- [69] Genkin M and Engel T A 2020 Moving beyond generalization to accurate interpretation of flexible models *Nat. Mach. Intell.* **2** 674–83
- [70] Mlynarski W, Hledík M, Sokolowski T R and Tkačik G 2021 Statistical analysis and optimality of neural systems *Neuron* **109** 1227–41
- [71] Berman G J, Choi D M, Bialek W and Shaevitz J W 2014 Mapping the stereotyped behaviour of freely moving fruit flies *J. R. Soc. Interface* **11** 20140672
- [72] Brown A E and de Bivort B 2018 Ethology as a physical science *Nat. Phys.* **14** 653–7
- [73] Stephens G J, Johnson-Kerner B, Bialek W and Ryu W S 2008 Dimensionality and dynamics in the behavior of *C. elegans* *PLoS Comput. Biol.* **4** e1000028
- [74] Stephens G J, Bueno M, Mesquita D, Ryu W S and Bialek W 2011 Emergence of long timescales and stereotyped behaviors in *caenorhabditis elegans* *Proc. Natl Acad. Sci. USA* **108** 7286–9
- [75] Costa A C, Ahamed T and Stephens G J 2019 Adaptive, locally linear models of complex dynamics *Proc. Natl Acad. Sci. USA* **116** 1501–10
- [76] Crutchfield J P and McNamara B 1987 Equations of motion from a data series *Complex Syst.* **1** 417–52 (available at: [www.complex-systems.com/abstracts/v01\\_i03\\_a03/](http://www.complex-systems.com/abstracts/v01_i03_a03/))
- [77] Daniels B C and Nemenman I 2015 Automated adaptive inference of phenomenological dynamical models *Nat. Commun.* **6** 1–8
- [78] Brunton S L, Proctor J L and Kutz J N 2016 Discovering governing equations from data: sparse identification of nonlinear dynamical systems *Proc. Natl Acad. Sci. USA* **113** 3932–7
- [79] Champion K, Lusch B, Nathan Kutz J and Brunton S L 2019 Data-driven discovery of coordinates and governing equations *Proc. Natl Acad. Sci. USA* **116** 22445–51
- [80] Chen Z, Liu Y and Sun H 2021 Physics-informed learning of governing equations from scarce data *Nat. Commun.* **12** 1–13
- [81] Siegert S, Friedrich R and Peinke J 1998 Analysis of data sets of stochastic systems *Phys. Lett. A* **243** 275–80
- [82] Ragwitz M and Kantz H 2001 Indispensable finite time corrections for Fokker-Planck equations from time series data *Phys. Rev. Lett.* **87** 254501
- [83] Beheiry M E, Dahan M and Masson J-B 2015 InferenceMAP: mapping of single-molecule dynamics with Bayesian inference *Nat. Methods* **12** 594–5
- [84] Pérez García L, Donlucas Pérez J, Volpe G, Arzola A V and Volpe G 2018 High-performance reconstruction of microscopic force fields from Brownian trajectories *Nat. Commun.* **9** 5166
- [85] Frishman A and Ronceray P 2020 Learning force fields from stochastic trajectories *Phys. Rev. X* **10** 21009
- [86] Brückner D B, Ronceray P and Broedersz C P 2020 Inferring the dynamics of underdamped stochastic systems *Phys. Rev. Lett.* **125** 58103
- [87] Boninsegna L, Nüske F and Clementi C 2018 Sparse learning of stochastic dynamical equations *J. Chem. Phys.* **148** 241723
- [88] Ferretti F, Chardès V, Mora T, Walczak A M and Giardina I 2020 Building general Langevin models from discrete datasets *Phys. Rev. X* **10** 031018
- [89] Dai M, Gao T, Lu Y, Zheng Y and Duan J 2020 Detecting the maximum likelihood transition path from data of stochastic dynamical systems *Chaos* **30** 113124
- [90] Callaham J L, Loiseau J-C, Rigas G and Brunton S L 2021 Nonlinear stochastic modelling with Langevin regression *Proc. R. Soc. A* **477** 20210092
- [91] Huang Y, Mabrouk Y, Gompper G and Sabass B 2022 Sparse inference and active learning of stochastic differential equations from data *Sci. Rep.* **12** 21691
- [92] Rudy S H, Brunton S L, Proctor J L and Kutz J N 2016 Data-driven discovery of partial differential equations *Sci. Adv.* **3** e1602614
- [93] Borzou A, Patteson A E and Schwarz J M 2021 A data-driven statistical field theory for active matter (arXiv:[2103.03461](https://arxiv.org/abs/2103.03461))
- [94] Selmeczi D, Mosler S, Hagedorn P H, Larsen N B and Flyvbjerg H 2005 Cell motility as persistent random motion: theories from experiments *Biophys. J.* **89** 912–31
- [95] Lo Vecchio S, Thiagarajan R, Caballero D, Vignon V, Navoret L, Voituriez R and Riveline D 2020 Collective dynamics of focal adhesions regulate direction of cell motion *Cell Syst.* **10** 1–8
- [96] Scarpa E *et al* 2013 A novel method to study contact inhibition of locomotion using micropatterned substrates *Biol. Open* **2** 901–6
- [97] Milano D F, Ngai N A, Muthuswamy S K and Asthagiri A R 2016 Regulators of metastasis modulate the migratory

- response to cell contact under spatial confinement *Biophys. J.* **110** 1886–95
- [98] Brückner D B, Arlt N, Fink A, Ronceray P, Rädler J O and Broedersz C P 2020 Learning the dynamics of cell-cell interactions in confined cell migration *Proc. Natl Acad. Sci. USA* **118** e2016602118
- [99] Gail M H and Boone C W 1970 The locomotion of mouse fibroblasts in tissue culture *Biophys. J.* **10** 980–93
- [100] Selmecci D, Li L, Pedersen L I I, Nørrelykke S F, Hagedorn P H, Mosler S, Larsen N B, Cox E C and Flyvbjerg H 2008 Cell motility as random motion: a review *Eur. Phys. J. Spec. Top.* **157** 1–15
- [101] Harris T H *et al* 2012 Generalized Lévy walks and the role of chemokines in migration of effector CD8<sup>+</sup> T cells *Nature* **486** 545–8
- [102] Danuser G, Allard J and Mogilner A 2013 Mathematical modeling of eukaryotic cell migration: insights beyond experiments *Ann. Rev. Cell Dev. Biol.* **4** 501–28
- [103] Przibram K 1913 Über die ungeordnete Bewegung niederer Tiere *Pflügers Arch. Physiol.* **153** 401–5
- [104] Fürth R 1917 Einige Untersuchungen über Brownsche bewegung an einem einzelteilchen *Ann. Phys., Lpz.* **53** 177
- [105] Fürth R 1920 Die Brownsche Bewegung bei Berücksichtigung einer Persistenz der Bewegungsrichtung. Mit Anwendungen auf die Bewegung lebender Infusorien *Z. Phys.* **2** 244–56
- [106] Dunn G A 1983 Characterising a kinesis response: time averaged measures of cell speed and directional persistence *Agents Actions Suppl.* **12** 14–33
- [107] Dieterich P, Klages R, Preuss R and Schwab A 2008 Anomalous dynamics of cell migration *Proc. Natl Acad. Sci. USA* **105** 459–63
- [108] Potdar A A, Jeon J, Weaver A M, Quaranta V and Cummings P T 2010 Human mammary epithelial cells exhibit a bimodal correlated random walk pattern *PLoS One* **5** e9636
- [109] Metzner C, Mark C, Steinwachs J, Lautscham L, Stadler F and Fabry B 2015 Superstatistical analysis and modelling of heterogeneous random walks *Nat. Commun.* **6** 7516
- [110] Banigan E J, Harris T H, Christian D A, Hunter C A and Liu A J 2015 Heterogeneous CD8<sup>+</sup> T cell migration in the lymph node in the absence of inflammation revealed by quantitative migration analysis *PLoS Comput. Biol.* **11** e1004058
- [111] Passucci G, Brasch M E, Henderson J H, Zaburdaev V and Manning M L 2019 Identifying the mechanism for superdiffusivity in mouse fibroblast motility *PLoS Comput. Biol.* **15** e1006732
- [112] Pedersen J N, Li L, Grdinaru C, Austin R H, Cox E C and Flyvbjerg H 2016 How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time *Phys. Rev. E* **94** 062401
- [113] Czirók A, Schleitt K, Madarász E and Vicsek T 1998 Exponential distribution of locomotion activity in cell cultures *Phys. Rev. Lett.* **81** 3038–41
- [114] Cherstvy A G, Nagel O, Beta C and Metzler R 2018 Non-Gaussianity, population heterogeneity and transient superdiffusion in the spreading dynamics of amoeboid cells *Phys. Chem. Chem. Phys.* **20** 23034–54
- [115] Gardiner C 1985 *Handbook of Stochastic Methods* (Springer)
- [116] Takagi H, Sato M J, Yanagida T and Ueda M 2008 Functional analysis of spontaneous cell movement under different physiological conditions *PLoS One* **3** e2648
- [117] Bödeker H U, Beta C, Frank T D and Bodenschatz E 2010 Quantitative analysis of random amoeboid motion *Europhys. Lett.* **90** 28005
- [118] Li L, Cox E C and Flyvbjerg H 2011 Dicty dynamics': *Dictyostelium* motility as persistent random motion *Phys. Biol.* **8** 046006
- [119] Zaburdaev V, Schmiedeberg M and Stark H 2008 Random walks with random velocities *Phys. Rev. E* **78** 011119
- [120] Wu P-H, Giri A and Wirtz D 2015 Statistical analysis of cell migration in 3D using the anisotropic persistent random walk model *Nat. Protocols* **10** 517–27
- [121] Taktikos J, Stark H and Zaburdaev V 2013 How the motility pattern of bacteria affects their dispersal and chemotaxis *PLoS One* **8** e81936
- [122] Tranquillo R T and Lauffenburger D A 1987 Stochastic model of leukocyte chemosensory movement *J. Math. Biol.* **25** 229–62
- [123] Tranquillo R, Lauffenburger D and Zigmond S 1988 A stochastic model for leukocyte random motility and chemotaxis based on receptor binding fluctuations *J. Cell Biol.* **106** 303–9
- [124] Amselem G, Theves M, Bae A, Bodenschatz E and Beta C 2012 A stochastic description of *Dictyostelium* chemotaxis *PLoS One* **7** e37213
- [125] Even-Ram S and Yamada K M 2005 Cell migration in 3D matrix *Curr. Opin. Cell Biol.* **17** 524–32
- [126] Driscoll M K and Danuser G 2015 Quantifying modes of 3D cell migration *Trends Cell Biol.* **25** 749–59
- [127] Vargas P, Barbier L, Sáez P J and Piel M 2017 Mechanisms for fast cell migration in complex environments *Curr. Opin. Cell Biol.* **48** 72–78
- [128] Wu P-H, Giri A, Sun S X and Wirtz D 2014 Three-dimensional cell migration does not follow a random walk *Proc. Natl Acad. Sci. USA* **111** 3949–54
- [129] Bhattacharjee T and Datta S S 2019 Bacterial hopping and trapping in porous media *Nat. Commun.* **10** 2–10
- [130] Bhattacharjee T and Datta S S 2019 Confinement and activity regulate bacterial motion in porous media *Soft Matter* **15** 9920–30
- [131] Alom Ruiz S and Chen C S 2007 Microcontact printing: a tool to pattern *Soft Matter* **3** 168–77
- [132] Théry M 2010 Micropatterning as a tool to decipher cell morphogenesis and functions *J. Cell Sci.* **123** 4201–13
- [133] Paul C D, Hung W-C, Wirtz D and Konstantopoulos K 2016 Engineered models of confined cell migration *Annu. Rev. Biomed. Eng.* **18** 159–80
- [134] Ermis M, Antmen E and Hasirci V 2018 Micro and Nanofabrication methods to control cell-substrate interactions and cell behavior: a review from the tissue engineering perspective *Bioact. Mater.* **3** 355–69
- [135] Singhvi R, Kumar A, Lopez G P, Stephanopoulos G N, Wang D C, Whitesides G M and Ingber D E 1994 Engineering cell shape and function *Science* **264** 696–8
- [136] Chen C S, Mrksich M, Huang S, Whitesides G M and Ingber D E 1998 Micropatterned surfaces for control of cell shape, position and function *Biotechnol. Prog.* **14** 356–63
- [137] Lin B, Yin T, Wu Y I, Inoue T and Levchenko A 2015 Interplay between chemotaxis and contact inhibition of locomotion determines exploratory cell migration *Nat. Commun.* **6** 6619
- [138] Renkawitz J *et al* 2019 Nuclear positioning facilitates amoeboid migration along the path of least resistance *Nature* **568** 546–50
- [139] Kopf A *et al* 2020 Microtubules control cellular shape and coherence in amoeboid migrating cells *J. Cell Biol.* **219** e201907154
- [140] Reversat A *et al* 2020 Cellular locomotion using environmental topography *Nature* **582** 582–5
- [141] Davidson P M, Battistella A, Dejardin T, Betz T, Plastino J, Borghi N, Borghi N and Sykes C 2020 Actin accumulates nesprin-2 at the front of the nucleus during confined cell migration *EMBO Rep.* **21** e49910
- [142] Zhang Y Z, Su B, Venugopal J, Ramakrishna S and Lim C T 2007 Biomimetic and bioactive nanofibrous scaffolds from

- electrospun composite nanofibers *Int. J. Nanomed.* **2** 623–38
- [143] Guetta-Terrier C *et al* 2015 Protrusive waves guide 3D cell migration along nanofibers *J. Cell Biol.* **211** 683–701
- [144] Singh J, Pagulayan A, Camley B A and Nain A S 2021 Rules of contact inhibition of locomotion for cells on suspended nanofibers *Proc. Natl Acad. Sci.* **118** e2011815118
- [145] Maiuri P *et al* 2012 The first world cell race *Curr. Biol.* **22** R673–5
- [146] Fraley S I, Feng Y, Giri A, Longmore G D and Wirtz D 2012 Dimensional and temporal controls of three-dimensional cell migration by zyxin and binding partners *Nat. Commun.* **3** 713–9
- [147] Doyle A D, Wang F W, Matsumoto K and Yamada K M 2009 One-dimensional topography underlies three-dimensional fibrillar cell migration *J. Cell Biol.* **184** 481–90
- [148] Mohammed D *et al* 2019 Substrate area confinement is a key determinant of cell velocity in collective migration *Nat. Phys.* **15** 858–66
- [149] Caballero D, Voituriez R and Riveline D 2014 Protrusion fluctuations direct cell motion *Biophys. J.* **107** 34–42
- [150] Comelles J, Caballero D, Voituriez R, Hortigüela V, Wollrab V, Godeau A L, Samitier J, Martínez E and Riveline D 2014 Cells as active particles in asymmetric potentials: motility under external gradients *Biophys. J.* **107** 1513–22
- [151] Ron J E, Crestani M, Kux J M, Liu J, Al-Dam N, Monzo P, Gauthier N C, Sáez P J and Gov N S 2024 Emergent seesaw oscillations during cellular directional decision-making *Nat. Phys.* **20** 501–11
- [152] Hadjitheodorou A, Bell G R R, Ellett F, Irimia D, Tibshirani R, Collins S R and Theriot J A 2023 Leading edge competition promotes context-dependent responses to receptor inputs to resolve directional dilemmas in neutrophil migration *Cell Syst.* **14** 196–209.E6
- [153] Chang S S, Guo W-H, Kim Y and Wang Y-L 2013 Guidance of cell migration by substrate dimension *Biophys. J.* **104** 313–21
- [154] Park J-A *et al* 2015 Unjamming and cell shape in the asthmatic airway epithelium *Nat. Mater.* **14** 1040–8
- [155] Green B J, Panagiotakopoulou M, Pramotton F M, Stefopoulos G, Kelley S O, Poulikakos D and Ferrari A 2018 Pole shape defines paths of metastatic cell migration *Nano Lett.* **18** 2140–7
- [156] Wolf K, te Lindert M, Krause M, Alexander S, te Riet J, Willis A L, Hoffman R M, Figidor C G, Weiss S J and Friedl P 2013 Physical limits of cell migration: control by ECM space and nuclear deformation and tuning by proteolysis and traction force *J. Cell Biol.* **201** 1069–84
- [157] Mahmud G, Campbell C J, Bishop K J, Komarova Y A, Chaga O, Soh S, Huda S, Kandere-Grzybowska K and Grzybowski B A 2009 Directing cell motions on micropatterned ratchets *Nat. Phys.* **5** 606–12
- [158] Caballero D, Comelles J, Piel M, Voituriez R and Riveline D 2015 Ratchetaxis: long-range directed cell migration by local cues *Trends Cell Biol.* **25** 815–27
- [159] Schreiber C, Segerer F J, Wagner E, Roidl A and Rädler J O 2016 Ring-shaped microlanes and chemical barriers as a platform for probing single-cell migration *Sci. Rep.* **6** 26858
- [160] Köhler B *et al* 2023 Unravelling the metastasis-preventing effect of miR-200c *in vitro* and *in vivo* *bioRxiv Preprint* (<https://doi.org/10.1101/2023.11.14.566527v1>)
- [161] Frey E and Kroy K 2005 Brownian motion: a paradigm of soft matter and biological physics *Ann. Phys., Lpz.* **14** 20–50
- [162] Brückner D B 2021 Stochastic dynamics of migrating cells: a data-driven approach *PhD Thesis* LMU München: Faculty of Physics
- [163] Oliver T, Dembo M and Jacobson K 1999 Separation of propulsive and adhesive traction stresses in locomoting keratocytes *J. Cell Biol.* **145** 589–604
- [164] Lee J, Leonard M, Oliver T, Ishihara A and Jacobson K 1994 Traction forces generated by locomoting keratocytes *J. Cell Biol.* **127** 1957–64
- [165] Lee J, Ishihara A, Theriot J A and Jacobson K 1993 Principles of locomotion for simple-shaped cells *Nature* **362** 167–71
- [166] Romanczuk P, Bär M, Ebeling W, Lindner B and Schimansky-Geier L 2012 Active Brownian particles: from individual to collective stochastic dynamics: from individual to collective stochastic dynamics *Eur. Phys. J. Spec. Top.* **202** 1–162
- [167] d'Alessandro J, Barbier-Chebbah A, Cellerin V, Benichou O, Mége R M, Voituriez R and Ladoux B 2021 Cell migration guided by long-lived spatial memory *Nat. Commun.* **12** 4118
- [168] Yu W, Cho J and Burton J C 2022 Extracting forces from noisy dynamics in dusty plasmas *Phys. Rev. E* **106** 035303
- [169] Boudet J F *et al* 2021 From collections of independent, mindless robots to flexible, mobile and directional superstructures *Sci. Robot.* **6** eabd0272
- [170] Ahamed T, Costa A C and Stephens G J 2021 Capturing the continuous complexity of behaviour in *Caenorhabditis elegans* *Nat. Phys.* **17** 275–83
- [171] Friedrich R, Zeller J and Peinke J 1998 A note on three-point statistics of velocity increments in turbulence *Europhys. Lett.* **41** 153–8
- [172] Vestergaard C L, Pedersen J N, Mortensen K I and Flyvbjerg H 2015 Estimation of motility parameters from trajectory data: a condensate of our recent results *Eur. Phys. J. Spec. Top.* **224** 1151–68
- [173] Guimerà R, Reichardt I, Aguilar-Mogas A, Massucci F A, Miranda M, Pallarès J and Sales-Pardo M 2020 A Bayesian machine scientist to aid in the solution of challenging scientific problems *Sci. Adv.* **6** eaav6971
- [174] Kloeden P and Platen E 1992 *Numerical Solution of Stochastic Differential Equations* (Springer)
- [175] Gnesotto F S, Gradziuk G, Ronceray P and Broedersz C P 2020 Learning the non-equilibrium dynamics of Brownian movies *Nat. Commun.* **11** 1–9
- [176] Stepanians G, Hastewell A D, Skinner D J, Totz J F and Dunkel J 2023 Discovering dynamics and parameters of nonlinear oscillatory and chaotic systems from partial observations (arXiv:[2304.04818](https://arxiv.org/abs/2304.04818))
- [177] Niepel M, Spencer S L and Sorger P K 2009 Non-genetic cell-to-cell variability and the consequences for pharmacology *Curr. Opin. Chem. Biol.* **13** 556–61
- [178] Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, Rosenfeld N, Danon T, Perzov N and Alon U 2006 Variability and memory of protein levels in human cells *Nature* **444** 28–31
- [179] Raj A and van Oudenaarden A 2008 Nature, nurture, or chance: stochastic gene expression and its consequences *Cell* **135** 216–26
- [180] Wagner A, Regev A and Yosef N 2016 Revealing the vectors of cellular identity with single-cell genomics *Nat. Biotechnol.* **34** 1145–60
- [181] Altschuler S J and Wu L F 2010 Cellular heterogeneity: do differences make a difference? *Cell* **141** 559–63
- [182] Cohen A A *et al* 2008 Dynamic proteomics of individual cancer cells in response to a drug *Science* **322** 1511–6
- [183] Feinerman O, Veiga J, Dorfman J R, Germain R N and Altan-Bonnet G 2008 Variability and robustness in T cell

- activation from regulated heterogeneity in protein levels *Science* **321** 1081–4
- [184] Gascoigne K E and Taylor S S 2008 Cancer cells display profound intra- and interline variation following prolonged exposure to antimitotic drugs *Cancer Cell* **14** 111–22
- [185] Wieser S, Weghuber J, Sams M, Stockinger H and Schütz G J 2009 Cell-to-cell variability in the diffusion constants of the plasma membrane proteins CD59 and CD147 *Soft Matter* **5** 3287–94
- [186] Camley B A and Rappel W-J 2017 Cell-to-cell variation sets a tissue-rheology-dependent bound on collective gradient sensing *Proc. Natl Acad. Sci.* **114** E10074–82
- [187] Li X, Das A and Bi D 2019 Mechanical heterogeneity in tissues promotes rigidity and controls cellular invasion *Phys. Rev. Lett.* **123** 58101
- [188] Boehm M and Nabel E G 2001 Cell cycle and cell migration *Circulation* **103** 2879–81
- [189] Poddar A A, Lu J, Jeon J, Weaver A M and Cummings P T 2009 Bimodal analysis of mammary epithelial cell migration in two dimensions *Ann. Biomed. Eng.* **37** 230–45
- [190] Jordan D, Kuehn S, Katifori E and Leibler S 2013 Behavioral diversity in microbes and low-dimensional phenotypic spaces *Proc. Natl Acad. Sci.* **110** 14018–23
- [191] Bentley S A, Laeverenz-Schlogelhofer H, Anagnostidis V, Cammann J, Mazza M G, Gielen F and Wan K Y 2022 Phenotyping single-cell motility in microfluidic confinement *eLife* **11** e76519
- [192] Wadhwa N and Berg H C 2022 Bacterial motility: machinery and mechanisms *Nat. Rev. Microbiol.* **20** 161–73
- [193] Korabel N, Clemente G D, Han D, Feldman F, Millard T H and Waigh T A 2022 Hemocytes in Drosophila melanogaster embryos move via heterogeneous anomalous diffusion *Commun. Phys.* **5** 269
- [194] Sasso S, Stibor H, Mittag M and Grossman A R 2018 From molecular manipulation of domesticated Chlamydomonas reinhardtii to survival in nature *eLife* **7** e39233
- [195] Wan K Y and Goldstein R E 2018 Time irreversibility and criticality in the motility of a flagellate microorganism *Phys. Rev. Lett.* **121** 058103
- [196] Schumacher L J, Maini P K and Baker R E 2017 Semblance of heterogeneity in collective cell migration *Cell Syst.* **5** 119–27
- [197] Brückner D B, Fink A, Rädler J O and Broedersz C P 2020 Disentangling the behavioural variability of confined cell migration *J. R. Soc. Interface* **17** 20190689
- [198] Chan T E, Stumpf M P and Babtie A C 2017 Gene regulatory network inference from single-cell data using multivariate information measures *Cell Syst.* **5** 251–67
- [199] Todorov H, Cannoodt R, Saelens W and Saeys Y 2019 Network inference from single-cell transcriptomic data BT - gene regulatory networks: methods and protocols *Gene Regulatory Networks* ed G Sanguinetti and V A Huynh-Thu (Springer) pp 235–49
- [200] Li B and You L 2013 Predictive power of cell-to-cell variability *Quant. Biol.* **1** 131–9
- [201] Sachs K, Perez O, Pe'er D, Lauffenburger D A and Nolan G P 2005 Causal protein-signaling networks derived from multiparameter single-cell data *Science* **308** 523–30
- [202] Beroz F, Jawerth L M, Münster S, Weitz D A, Broedersz C P and Wingreen N S 2017 Physical limits to biomechanical sensing in disordered fibre networks *Nat. Commun.* **8** 16096
- [203] Berthier E, Yang H, Guo M, Ronceray P and Broedersz C P 2022 Nonlinear mechanosensation in fiber networks *Phys. Rev. Research* **6** 013327 (available at: <https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.6.013327>)
- [204] Carmona-Fontaine C, Matthews H K, Kuriyama S, Moreno M, Dunn G A, Parsons M, Stern C D and Mayor R 2008 Contact inhibition of locomotion *in vivo* controls neural crest directional migration *Nature* **456** 957–61
- [205] Stramer B and Mayor R 2017 Mechanisms and *in vivo* functions of contact inhibition of locomotion *Nat. Rev. Mol. Cell Biol.* **18** 43–55
- [206] Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Gustavo Pesce C and Brent R 2005 Regulated cell-to-cell variation in a cell-fate decision system *Nature* **437** 699–706
- [207] Snijder B, Sacher R, Rämö P, Damm E-M, Liberali P and Pelkmans L 2009 Population context determines cell-to-cell variability in endocytosis and virus infection *Nature* **461** 520
- [208] Snijder B and Pelkmans L 2011 Origins of regulated cell-to-cell variability *Nat. Rev. Mol. Cell Biol.* **12** 119–25
- [209] Dietrich M, Le Roy H, Brückner D B, Engelke H, Zantl R, Rädler J O and Broedersz C P 2018 Guiding 3D cell migration in deformed synthetic hydrogel microstructures *Soft Matter* **14** 2816–26
- [210] Ipíñia E P, D'Alessandro J, Ladoux B and Camley B A 2023 Secreted footprints let cells switch between confined, oscillatory, and exploratory migration *bioRxiv Preprint* (<https://doi.org/10.1101/2023.09.14.557437>)
- [211] Peskin C S, Odell G M and Oster G F 1993 Cellular motions and thermal fluctuations: the Brownian ratchet *Biophys. J.* **65** 316–24
- [212] Mogilner A and Oster G 1996 Cell motility driven by actin polymerization *Biophys. J.* **71** 3030–45
- [213] Mogilner A and Edelstein-Keshet L 2002 Regulation of actin dynamics in rapidly moving cells: a quantitative analysis *Biophys. J.* **83** 1237–58
- [214] Marée A F M, Jilkine A, Dawes A, Grieneisen V A and Edelstein-Keshet L 2006 Polarization and movement of keratocytes: a multiscale modelling approach *Bull. Math. Biol.* **68** 1169–211
- [215] Satulovsky J, Lui R and Wang Y-I 2008 Exploring the control circuit of cell migration by mathematical modeling *Biophys. J.* **94** 3671–83
- [216] Nishimura S I, Ueda M and Sasai M 2009 Cortical factor feedback model for cellular locomotion and cytoskeleton fission *PLoS Comput. Biol.* **5** e1000310
- [217] Raynaud F, Ambühl M E, Gabella C, Bornert A, Sbalzarini I F, Meister J-J and Verkhovsky A B 2016 Minimal model for spontaneous cell polarization and edge activity in oscillating, rotating and migrating cells *Nat. Phys.* **12** 367–73
- [218] Stéphanou A, Mylona E, Chaplain M and Tracqui P 2008 A computational model of cell migration coupling the growth of focal adhesions with oscillatory cell protrusions *J. Theor. Biol.* **253** 701–16
- [219] Shao D, Rappel W-J and Levine H 2010 Computational model for cell morphodynamics *Phys. Rev. Lett.* **105** 108104
- [220] Herant M and Dembo M 2010 Form and function in cell motility: from fibroblasts to keratocytes *Biophys. J.* **98** 1408–17
- [221] Coburn L, Cerone L, Torney C, Couzin I D and Neufeld Z 2013 Tactile interactions lead to coherent motion and enhanced chemotaxis of migrating cells *Phys. Biol.* **10** 046002
- [222] Zhu J and Mogilner A 2016 Comparison of cell migration mechanical strategies in three-dimensional matrices: a computational study *Interface Focus* **6** 20160040
- [223] Wolgemuth C W, Stajic J and Mogilner A 2011 Redundant mechanisms for stable cell locomotion revealed by minimal models *Biophys. J.* **101** 545–53

- [224] Hawkins R J and Voituriez R 2010 Mechanisms of cell motion in confined geometries *Math. Model. Nat. Phenom.* **5** 84–105
- [225] Bangasser B L *et al* 2017 Shifting the optimal stiffness for cell migration *Nat. Commun.* **8** 15313
- [226] Prahler S, Stanslaski M R, Vargas P, Piel M and Odde D J 2020 Predicting confined 1D cell migration from parameters calibrated to a 2D motor-clutch model *Biophys. J.* **118** 1709–20
- [227] Monzo P *et al* 2016 Mechanical confinement triggers glioma linear migration dependent on formin FHOD3 *Mol. Biol. Cell* **27** 1246–61
- [228] Brückner D B, Schmitt M, Fink A, Ladurner G, Flommersfeld J, Arlt N, Hannezo E, Rädler J O and Broedersz C P 2022 Geometry adaptation of protrusion and polarity dynamics in confined cell migration *Phys. Rev. X* **12** 031041
- [229] Edwards C M and Schwarz U S 2011 Force localization in contracting cell layers *Phys. Rev. Lett.* **107** 128101
- [230] Mertz A F, Banerjee S, Che Y, German G K, Xu Y, Hyland C, Marchetti M C, Horsley V and Dufresne E R 2012 Scaling of traction forces with the size of cohesive cell colonies *Phys. Rev. Lett.* **108** 198101
- [231] Oakes P W, Banerjee S, Marchetti M C and Gardel M L 2014 Geometry regulates traction stresses in adherent cells *Biophys. J.* **107** 825–33
- [232] Welf E S and Danuser G 2014 Using fluctuation analysis to establish causal relations between cellular events without experimental perturbation *Biophys. J.* **107** 2492–8
- [233] Lee K, Elliott H L, Oak Y, Zee C-T, Groisman A, Tytell J D and Danuser G 2015 Functional hierarchy of redundant actin assembly factors revealed by fine-grained registration of intrinsic image fluctuations *Cell Syst.* **1** 37–50
- [234] Isogai T and Danuser G 2018 Discovery of functional interactions among actin regulators by analysis of image fluctuations in an unperturbed motile cell system *Phil. Trans. R. Soc. B* **373** 20170110
- [235] Gordonov S, Hwang M K, Wells A, Gertler F B, Lauffenburger D A and Bathe M 2016 Time series modeling of live-cell shape dynamics for image-based phenotypic profiling *Integr. Biol.* **8** 73–90
- [236] Tweedy L, Witzel P, Heinrich D, Insall R H and Endres R G 2019 Screening by changes in stereotypical behavior during cell motility *Sci. Rep.* **9** 8784
- [237] Chan C K, Hadjitheodorou A, Tsai T Y and Theriot J A 2020 Quantitative comparison of principal component analysis and unsupervised deep learning using variational autoencoders for shape analysis of motile cells *bioRxiv Preprint* (<https://doi.org/10.1101/2020.06.26.174474>)
- [238] Cavanagh H, Kempe D, Mazalo J K, Biro M and Endres R G 2022 T cell morphodynamics reveal periodic shape oscillations in three-dimensional migration *J. R. Soc. Interface* **19** 20220081
- [239] Schmitt M S, Colen J, Sala S, Devany J, Seetharaman S, Caillier A, Gardel M L, Oakes P W and Vitelli V 2024 Machine learning interpretable models of cell mechanics from protein images *Cell* **187** S0092867423013314
- [240] Ronneberger O, Fischer P and Brox T 2015 *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Springer) pp 234–41 (available at: [https://link.springer.com/chapter/10.1007/978-3-319-24574-4\\_28](https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28))
- [241] Falk T *et al* 2019 U-Net: deep learning for cell counting, detection and morphometry *Nat. Methods* **16** 67–70
- [242] Hermans T M, Pilans D, Huda S, Fuller P, Kandere-Grzybowska K and Grzybowski B A 2013 Motility efficiency and spatiotemporal synchronization in non-metastatic vs. metastatic breast cancer cells *Integr. Biol.* **5** 1464–73
- [243] Elbez R, Folz J, McLean A, Roca H, Labuz J M, Pienta K J, Takayama S and Kopelman R 2021 Cell-morphodynamic phenotype classification with application to cancer metastasis using cell magnetorotation and machine-learning *PLoS One* **16** e0259462
- [244] Buggenthin F *et al* 2017 Prospective identification of hematopoietic lineage choice by deep learning *Nat. Methods* **14** 403–6
- [245] Imoto D *et al* 2021 Comparative mapping of crawling-cell morphodynamics in deep learning-based feature space *PLOS Comput. Biol.* **17** e1009237
- [246] Eddy C Z, Raposo H, Manchanda A, Wong R, Li F and Sun B 2021 Morphodynamics facilitate cancer cells to navigate 3D extracellular matrix *Sci. Rep.* **11** 20434
- [247] Shafqat-Abbas H, Kowalewski J M, Kiss A, Gong X, Hernandez-Varas P, Berge U, Jafari-Mamaghani M, Lock J G and Strömbäck S 2016 An analysis toolbox to explore mesenchymal migration heterogeneity reveals adaptive switching between distinct modes *eLife* **5** e11384
- [248] Machacek M and Danuser G 2006 Morphodynamic profiling of protrusion phenotypes *Biophys. J.* **90** 1439–52
- [249] Ma X, Dagliyan O, Hahn K M and Danuser G 2018 Profiling cellular morphodynamics by spatiotemporal spectrum decomposition *PLoS Comput. Biol.* **14** 1–29
- [250] Flommersfeld J, Stöberl S, Shah O, Rädler J O and Broedersz C P 2023 Geometry-sensitive protrusion growth directs confined cell migration (arXiv:2308.08372)
- [251] Davidson P M, Denais C, Bakshi M C and Lammerding J 2014 Nuclear deformability constitutes a rate-limiting step during cell migration in 3-D environments *Cell. Mol. Bioeng.* **7** 293–306
- [252] Denais C M, Gilbert R M, Isermann P, McGregor A L, Te Lindert M, Weigelin B, Davidson P M, Friedl P, Wolf K and Lammerding J 2016 Nuclear envelope rupture and repair during cancer cell migration *Science* **352** 353–8
- [253] Fruleux A and Hawkins R J 2016 Physical role for the nucleus in cell migration *J. Phys.: Condens. Matter* **28** 363002
- [254] Lomakin A J *et al* 2020 The nucleus acts as a ruler tailoring cell responses to spatial constraints *Science* **370** eaba2894
- [255] Estabrook I D, Thiam H R, Piel M and Hawkins R J 2021 Calculation of the force field required for nucleus deformation during cell migration through constrictions *PLOS Comput. Biol.* **17** e1008592
- [256] Stoeberl S, Flommersfeld J, Kreft M, Benoit M, Broedersz C and Raedler J O 2023 Nuclear deformation and dynamics of migrating cells in 3D confinement reveal adaptation of pulling and pushing forces *bioRxiv Preprint* (<https://doi.org/10.1101/2023.10.30.564765>)
- [257] Amiri S, Zhang Y, Gerardo A, Sykes C and Ronceray P 2023 Inferring geometrical dynamics of cell nucleus translocation (arXiv:2312.12402)
- [258] Leong F Y, Li Q, Lim C T and Chiam K-H 2011 Modeling cell entry into a micro-channel *Biomech. Model. Mechanobiol.* **10** 755–66
- [259] Scianna M and Preziosi L 2013 Modeling the influence of nucleus elasticity on cell invasion in fiber networks and microchannels *J. Theor. Biol.* **317** 394–406
- [260] Le Berre M, Liu Y-J, Hu J, Maiuri P, Bénichou O, Voituriez R, Chen Y and Piel M 2013 Geometric friction directs cell migration *Phys. Rev. Lett.* **111** 198101
- [261] Tozluoglu M, Tournier A L, Jenkins R P, Hooper S, Bates P A and Sahai E 2013 Matrix geometry determines optimal cancer cell migration strategy and modulates response to interventions *Nat. Cell Biol.* **15** 751–62
- [262] Aubry D, Thiam H, Piel M and Allena R 2015 A computational mechanics approach to assess the link between cell morphology and forces during confined migration *Biomech. Model. Mechanobiol.* **14** 143–57

- [263] Cao X, Moendarbary E, Isermann P, Davidson P M, Wang X, Chen M B, Burkart A K, Lammerding J, Kamm R D and Shenoy V B 2016 A chemomechanical model for nuclear morphology and stresses during cell transendothelial migration *Biophys. J.* **111** 1541–52
- [264] Tapon N and Hall A 1997 Rho, Rac and Cdc42 GTPases regulate the organization of the actin cytoskeleton *Curr. Opin. Cell Biol.* **9** 86–92
- [265] Nabi I 1999 The polarization of the motile cell *J. Cell Sci.* **112** 1803–11
- [266] Pouthas F, Girard P, Lecaudey V, Ly T B N, Gilmour D, Boulin C, Pepperkok R and Reynaud E G 2008 In migrating cells, the Golgi complex and the position of the centrosome depend on geometrical constraints of the substratum *J. Cell Sci.* **121** 2406–14
- [267] Icha J, Weber M, Waters J C and Norden C 2017 Phototoxicity in live fluorescence microscopy and how to avoid it *BioEssays* **39** 1700003
- [268] Kiepas A, Voorand E, Mubaid F, Siegel P M and Brown C M 2020 Optimizing live-cell fluorescence imaging conditions to minimize phototoxicity *J. Cell Sci.* **133** jcs.242834
- [269] Ounkomol C, Seshamani S, Maleckar M M, Collman F and Johnson G R 2018 Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy *Nat. Methods* **15** 917–20
- [270] Christiansen E M *et al* 2018 *In silico* labeling: predicting fluorescent labels in unlabeled images *Cell* **173** 792–803.e19
- [271] Gómez-González M, Latorre E, Arroyo M and Trepat X 2020 Measuring mechanical stress in living tissues *Nat. Rev. Phys.* **2** 300–17
- [272] Tanimoto H and Sano M 2014 A simple force-motion relation for migrating cells revealed by multipole analysis of traction stress *Biophys. J.* **106** 16–25
- [273] Rossetti L, Grosser S, Abenza J F, Valon L, Roca-Cusachs P, Alert R and Trepat X 2024 Optogenetic generation of leader cells reveals a force-velocity relation for collective cell migration *bioRxiv preprint*, (<https://doi.org/10.1101/2024.01.23.576733>)
- [274] Poujade M, Grasland-Mongrain E, Hertzog A, Jouanneau J, Chavrier P, Ladoux B, Buguin A and Silberzan P 2007 Collective migration of an epithelial monolayer in response to a model wound *Proc. Natl Acad. Sci. USA* **104** 15988–93
- [275] Stramer B, Wood W, Galko M J, Redd M J, Jacinto A, Parkhurst S M and Martin P 2005 Live imaging of wound inflammation in *Drosophila* embryos reveals key roles for small GTPases during *in vivo* cell migration *J. Cell Biol.* **168** 567–73
- [276] Weavers H, Liepe J, Sim A, Wood W, Martin P and Stumpf M P 2016 Systems analysis of the dynamic inflammatory response to tissue damage reveals spatiotemporal properties of the wound attractant gradient *Curr. Biol.* **26** 1975–89
- [277] Astin J W, Batson J, Kadir S, Charlet J, Persad R A, Gillatt D, Oxley J D and Nobes C D 2010 Competition amongst Eph receptors regulates contact inhibition of locomotion and invasiveness in prostate cancer cells *Nat. Cell Biol.* **12** 1194–204
- [278] Davis J R, Luchici A, Mosis F, Thackery J, Salazar J A, Mao Y, Dunn G A, Betz T, Miodownik M and Stramer B M 2015 Inter-cellular forces orchestrate contact inhibition of locomotion *Cell* **161** 361–73
- [279] Moore R, Theveneau E, Pozzi S, Alexandre P, Richardson J, Merks A, Parsons M, Kashef J, Linker C and Mayor R 2013 Par3 controls neural crest migration by promoting microtubule catastrophe during contact inhibition of locomotion *Development* **140** 4763–75
- [280] Matthews H K, Marchant L, Carmona-Fontaine C, Kuriyama S, Larraín J, Holt M R, Parsons M and Mayor R 2008 Directional migration of neural crest cells *in vivo* is regulated by Syndecan-4/Rac1 and non-canonical Wnt signaling/RhoA *Development* **135** 1771–80
- [281] Kadir S, Astin J W, Tahtamouni L, Martin P and Nobes C D 2011 Microtubule remodelling is required for the front-rear polarity switch during contact inhibition of locomotion *J. Cell Sci.* **124** 2642–53
- [282] Abercrombie M and Heaysman J E 1954 Observations on the social behaviour of cells in tissue culture. II. “Monolayering” of fibroblasts *Exp. Cell Res.* **6** 293–306
- [283] Mayor R and Carmona-Fontaine C 2010 Keeping in touch with contact inhibition of locomotion *Trends Cell Biol.* **20** 319–28
- [284] Marchetti M C, Joanny J F, Ramaswamy S, Liverpool T B, Prost J, Rao M and Simha R A 2013 Hydrodynamics of soft active matter *Rev. Mod. Phys.* **85** 1143–89
- [285] Honda H 1983 Geometrical models for cells in tissues *International Review of Cytology* vol 81, ed G Bourne, J Danielli and K Jeon (Academic) pp 191–248 (available at: [www.sciencedirect.com/science/article/abs/pii/S0074769608623396](http://www.sciencedirect.com/science/article/abs/pii/S0074769608623396))
- [286] Fletcher A G, Osterfield M, Baker R E and Shvartsman S Y 2014 Vertex models of epithelial morphogenesis *Biophys. J.* **106** 2291–304
- [287] Alt S, Ganguly P and Salbreux G 2017 Vertex models: from cell mechanics to tissue morphogenesis *Phil. Trans. R. Soc. B* **372** 20150520
- [288] Serra-Picamal X, Conte V, Vincent R, Anon E, Tambe D T, Bazellieres E, Butler J P, Fredberg J J and Trepat X 2012 Mechanical waves during tissue expansion *Nat. Phys.* **8** 628–34
- [289] Boocock D, Hino N, Ruzickova N, Hirashima T and Hannezo E 2020 Theory of mechanochemical patterning and optimal migration in cell monolayers *Nat. Phys.* **17** 267–74
- [290] Ilina O *et al* 2020 Cell–cell adhesion and 3D matrix confinement determine jamming transitions in breast cancer invasion *Nat. Cell Biol.* **22** 1103–15
- [291] Smeets B, Alert R, Pešek J, Pagonabarraga I, Ramon H and Vincent R 2016 Emergent structures and dynamics of cell colonies by contact inhibition of locomotion *Proc. Natl Acad. Sci. USA* **113** 14621–6
- [292] Sepúlveda N, Petitjean L, Cochet O, Grasland-Mongrain E, Silberzan P and Hakim V 2013 Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model *PLoS Comput. Biol.* **9** e1002944
- [293] Basan M, Elgeti J, Hannezo E, Rappel W-J and Levine H 2013 Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing *Proc. Natl Acad. Sci. USA* **110** 2452–9
- [294] Copenhagen K, Malet-Engra G, Yu W, Scita G, Gov N and Gopinathan A 2018 Frustration-induced phases in migrating cell clusters *Sci. Adv.* **4** eaar8483
- [295] Garcia S, Hannezo E, Elgeti J, Joanny J-F, Silberzan P and Gov N S 2015 Physics of active jamming during collective cellular motion in a monolayer *Proc. Natl Acad. Sci.* **112** 15314–9
- [296] d’Alessandro J, Solon A P, Hayakawa Y, Anjard C, Detcheverry F, Rieu J-P and Rivière C 2017 Contact enhancement of locomotion in spreading cell colonies *Nat. Phys.* **13** 999–1005
- [297] Hakim V and Silberzan P 2017 Collective cell migration: a physics perspective *Rep. Prog. Phys.* **80** 076601
- [298] Camley B A and Rappel W-J 2017 Physical models of collective

- cell motility: from cell to tissue *J. Phys. D: Appl. Phys.* **50** 113002
- [299] Alert R and Trepaut X 2020 Physical models of collective cell migration *Annu. Rev. Condens. Matter Phys.* **11** 77–101
- [300] Buttenschön A and Edelstein-Keshet L 2020 Bridging from single to collective cell migration: a review of models and links to experiments *PLoS Comput. Biol.* **16** e1008411
- [301] Partridge B L 1981 Internal dynamics and the interrelations of fish in schools *J. Comp. Physiol. A* **144** 313–25
- [302] Lukeman R, Li Y-X and Edelstein-Keshet L 2010 Inferring individual rules from collective behavior *Proc. Natl Acad. Sci. USA* **107** 12576–80
- [303] Katz Y, Tunstrøm K, Ioannou C C, Huepe C and Couzin I D 2011 Inferring the structure and dynamics of interactions in schooling fish *Proc. Natl Acad. Sci. USA* **108** 18720–5
- [304] Gautrais J, Ginelli F, Fournier R, Blanco S, Soria M, Chaté H and Theraulaz G 2012 Deciphering interactions in moving animal groups *PLoS Comput. Biol.* **8** e1002678
- [305] Lu F, Zhong M, Tang S and Maggioni M 2019 Nonparametric inference of interaction laws in systems of agents from trajectory data *Proc. Natl Acad. Sci. USA* **116** 14424–33
- [306] Miller J, Tang S, Zhong M and Maggioni M 2020 Learning theory for inferring interaction kernels in second-order interacting agent systems (arXiv:2010.03729)
- [307] Escobedo R, Lecheval V, Papaspyros V, Bonnet F, Mondada F, Sire C and Theraulaz G 2020 A data-driven method for reconstructing and modelling social interactions in moving animal groups *Phil. Trans. R. Soc. B* **375** 20190380
- [308] Ruiz-Garcia M, Gutierrez C M B, Alexander L C, Aarts D G A L, Ghiringhelli L and Valeriani C 2022 Discovering dynamic laws from observations: the case of self-propelled, interacting colloids (arXiv:2203.14846)
- [309] Messenger D A, Wheeler G E, Liu X and Bortz D M 2022 Learning anisotropic interaction rules from individual trajectories in a heterogeneous cellular population *J. R. Soc. Interface* **19** 20220412
- [310] Yang H, Meyer F, Huang S, Yang L, Lungu C, Olayioye M A, Buehler M J and Guo M 2024 Learning dynamics from multicellular graphs with deep neural networks (arXiv:2401.12196)
- [311] Zisis T, Brückner D B, Brandstätter T, Siow W X, D’Alessandro J, Vollmar A M, Broedersz C P and Zahler S 2022 Disentangling cadherin-mediated cell-cell interactions in collective cancer cell migration *Biophys. J.* **121** 44–60
- [312] Farhadifar R, Röper J-C, Aigouy B, Eaton S and Jülicher F 2007 The influence of cell mechanics, cell-cell interactions and proliferation on epithelial packing *Curr. Biol.* **17** 2095–104
- [313] Wozniak M A and Chen C S 2009 Mechanotransduction in development: a growing role for contractility *Nat. Rev. Mol. Cell Biol.* **10** 34–43
- [314] Noll N, Mani M, Heemskerk I, Streichan S J and Shraiman B I 2017 Active tension network model suggests an exotic mechanical state realized in epithelial tissues *Nat. Phys.* **13** 1221–6
- [315] Sknepnek R, Djafer-Cherif I, Chuai M, Weijer C and Henkes S 2023 Generating active T1 transitions through mechanochemical feedback *eLife* **12** e79862
- [316] Ioratim-Uba A, Liverpool T B and Henkes S 2023 Mechano-chemical active feedback generates convergence extension in epithelial tissue (arXiv:2303.02109)
- [317] Brauns F, Claussen N H, Wieschaus E F and Shraiman B I 2023 Epithelial flow by controlled transformation of internal force-balance geometry *bioRxiv Preprint* (<https://doi.org/10.1101/2023.05.30.542935>)
- [318] Colen J *et al* 2021 Machine learning active-nematic hydrodynamics *Proc. Natl Acad. Sci.* **118** e2016708118
- [319] Golden M, Grigoriev R O, Nambisan J and Fernandez-Nieves A 2023 Physically informed data-driven modeling of active nematics *Sci. Adv.* **9** eabq6120
- [320] Supek R, Song B, Hastewell A, Choi G P T, Mietke A and Dunkel J 2023 Learning hydrodynamic equations for active matter from particle simulations and experiments *Proc. Natl Acad. Sci.* **120** e2206994120
- [321] Kozak E L, Miranda-Rodríguez J R, Borges A, Dierkes K, Mineo A, Pinto-Teixeira F, Viader-Llargués O, Solon J, Chara O and López-Schier H 2023 Quantitative videomicroscopy reveals latent control of cell-pair rotations *in vivo* *Development* **150** dev200975
- [322] Bianco A, Poukkula M, Cliffe A, Mathieu J, Luque C M, Fulga T A and Rørt P 2007 Two distinct modes of guidance signalling during collective migration of border cells *Nature* **448** 362–5
- [323] Dai W *et al* 2020 Tissue topography steers migrating *Drosophila* border cells *Science* **370** 987–90
- [324] Friedl P, Noble P B, Walton P A, Laird D W, Chauvin P J, Tabah R J, Black M and Zänker K S 1995 Migration of coordinated cell clusters in mesenchymal and epithelial cancer explants *in vitro* *Cancer Res.* **55** 4557–60 (available at: <https://pubmed.ncbi.nlm.nih.gov/7553628/>)
- [325] Hou J-M, Krebs M, Ward T, Sloane R, Priest L, Hughes A, Clack G, Ranson M, Blackhall F and Dive C 2011 Circulating tumor cells as a window on metastasis biology in lung cancer *Am. J. Pathol.* **178** 989–96
- [326] Li D and Wang Y-L 2018 Coordination of cell migration mediated by sitedependent cell-cell contact *Proc. Natl Acad. Sci. USA* **115** 10678–83
- [327] Desai R A, Gopal S B, Chen S and Chen C S 2013 Contact inhibition of locomotion probabilities drive solitary versus collective cell migration *J. R. Soc. Interface* **10** 20130717
- [328] Huang S, Brangwynne C P, Parker K K and Ingber D E 2005 Symmetry-breaking in mammalian cell cohort migration during tissue pattern formation: role of random-walk persistence *Cell Motil. Cytoskel.* **61** 201–13
- [329] LaChance J, Suh K, Clausen J and Cohen D J 2022 Learning the rules of collective cell migration using deep attention networks *PLoS Comput. Biol.* **18** e1009293
- [330] Romeo N, Hastewell A D, Mietke A and Dunkel J 2021 Learning developmental mode dynamics from single-cell trajectories *eLife* **10** 1–36
- [331] Abercrombie M 1979 Contact inhibition and malignancy *Nature* **281** 259–62
- [332] Hayakawa M, Hiraiwa T, Wada Y, Kuwayama H and Shibata T 2020 Polar pattern formation induced by contact following locomotion in a multicellular system *eLife* **9** e53609
- [333] Zadeh P and Camley B A 2022 Picking winners in cell-cell collisions: Wetting, speed and contact *Phys. Rev. E* **106** 054413
- [334] Vedula S R K, Leong M C, Lai T L, Hersen P, Kabla A J, Lim C T and Ladoux B 2012 Emerging modes of collective cell migration induced by geometrical constraints *Proc. Natl Acad. Sci.* **109** 12974–9
- [335] Tarle V, Gauquelin E, Vedula S R K, D’Alessandro J, Lim C T, Ladoux B and Gov N S 2017 Modeling collective cell migration in geometric confinement *Phys. Biol.* **14** 035001
- [336] Jain S *et al* 2020 The role of single-cell mechanical behaviour and polarity in driving collective cell migration *Nat. Phys.* **16** 802–9
- [337] Vercrusse E, Brückner D B, Gómez-González M, Luciano M, Kalukula Y, Rossetti L, Trepaut X, Hannezo E and Gabriele S 2022 Geometry-driven migration

- efficiency of minimal cell clusters *bioRxiv Preprint* (<https://doi.org/10.1101/2022.07.17.500364>)
- [338] Pagès D-L *et al* 2022 Cell clusters adopt a collective amoeboid mode of migration in confined nonadhesive environments *Sci. Adv.* **8** eab8416
- [339] Ron J E, d'Alessandro J, Cellerin V, Voituriez R, Ladoux B and Gov N S 2023 Polarization and motility of one-dimensional multi-cellular trains *Biophys. J.* **122** 4598–613
- [340] Gu M, Fang X and Luo Y 2023 Data-driven model construction for anisotropic dynamics of active matter *PRX Life* **1** 013009
- [341] Wang H, Lacoche S, Huang L, Xue B and Muthuswamy S K 2013 Rotational motion during three-dimensional morphogenesis of mammary epithelial acini relates to laminin matrix assembly *Proc. Natl Acad. Sci.* **110** 163–8
- [342] Chin A S, Worley K E, Ray P, Kaur G, Fan J and Wan L Q 2018 Epithelial cell chirality revealed by three-dimensional spontaneous rotation *Proc. Natl Acad. Sci. USA* **115** 12188–93
- [343] Palamidessi A *et al* 2019 Unjamming overcomes kinetic and proliferation arrest in terminally differentiated cells and promotes collective motility of carcinoma *Nat. Mater.* **18** 1252–63
- [344] Brandstätter T, Brückner D B, Han Y L, Alert R, Guo M and Broedersz C P 2023 Curvature induces active velocity waves in rotating spherical tissues *Nat. Commun.* **14** 1643
- [345] Tan T H, Amiri A, Seijo-Barandiarán I, Staddon M F, Materne A, Tomas S, Duclut C, Popović M, Grapin-Botton A and Jülicher F 2022 Emergent chirality in active solid rotation of pancreas spheres *bioRxiv Preprint* (<https://doi.org/10.1101/2022.09.29.510101>)
- [346] Stock J, Kazmar T, Schlumm F, Hannezo E and Pauli A 2022 A self-generated Toddler gradient guides mesodermal cell migration *Sci. Adv.* **8** eadd2488
- [347] Pinheiro D, Kardos R, Hannezo E and Heisenberg C-P 2022 Morphogen gradient orchestrates pattern-preserving tissue morphogenesis via motility-driven unjamming *Nat. Phys.* **18** 1482–93
- [348] Tavano S, Brückner D B, Tasciyan S, Tong X, Kardos R, Schauer A, Hauschild R and Heisenberg C-P 2023 BMP-dependent patterning of ectoderm tissue material properties modulates lateral mesendoderm cell migration during early zebrafish gastrulation *bioRxiv Preprint* (<https://doi.org/10.1101/2023.07.21.550024>)
- [349] Etournay R *et al* 2015 Interplay of cell dynamics and epithelial tension during morphogenesis of the *Drosophila* pupal wing *eLife* **4** e07090
- [350] Morita H, Grigolon S, Bock M, Krens S G, Salbreux G and Heisenberg C-P 2017 The physical basis of coordinated tissue spreading in zebrafish gastrulation *Dev. Cell* **40** 354–66.e4
- [351] Streichan S J, Lefebvre M F, Noll N, Wieschaus E F and Shraiman B I 2018 Global morphogenetic flow is accurately predicted by the spatial distribution of myosin motors *eLife* **7** e27454
- [352] Münster S, Jain A, Mietke A, Pavlopoulos A, Grill S W and Tomancak P 2019 Attachment of the blastoderm to the vitelline envelope affects gastrulation of insects *Nature* **568** 395–9
- [353] Lefebvre M, Colen J, Claussen N, Brauns F, Raich M, Mitchell N, Fruchart M, Vitelli V and Streichan S J 2023 Learning a conserved mechanism for early neuroectoderm morphogenesis *bioRxiv Preprint* (<https://doi.org/10.1101/2023.12.22.573058>)
- [354] Rand D A, Raju A, Sáez M, Corson F and Siggia E D 2021 Geometry of gene regulatory dynamics *Proc. Natl Acad. Sci.* **118** e2109729118
- [355] Sáez M, Blassberg R, Camacho-Aguilar E, Siggia E D, Rand D A and Briscoe J 2021 Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions *Cell Syst.* **12** 1–17
- [356] Waddington C 1957 *The Strategy of the Genes* (Routledge)