1 How to Track Cells and Molecules

Seeing is believing and the development of high-resolution microscopes originally provided the most conclusive evidence for the existence of bacteria or *animalcules* (tiny animals) as they were first described.[1] Microscopy continues to be a central tool in modern bacterial biophysics and, when combined with quantitative image analysis tools, microscopes can provide unambiguous quantitative data to answer many of the questions related to bacterial behaviour.

A simple inverted optical microscope is shown in Figure 1.1. It follows a simple *4f* geometry, where *f* is the focal length of the condenser and objective lenses, and in practice, additional improvements are standard on laboratory microscopes e.g. Köhler illumination (to illuminate specimens in a uniform manner), phase contrast (to provide additional contrast for thin or transparent specimens), fluorescence optics (to allow imaging of fluorescent samples, Figure 13.8) and confocal pin holes (to improve background rejection, Figure 13.7).[2]

Once objects are identified in a microscopy image (the process of *segmentation*), analysing their dynamics by linking objects in consecutive images provides a rich source of additional information i.e. *tracks* are created that can be statistically analysed. Tracks can describe the motion of whole cells on the microscale, single molecules on the nanoscale or organelles on intermediate length scales e.g. the swimming behaviour of *Escherichia coli* (of micrometre length scales), the motion of proteins attached to a membrane (of nanometre length scales) or the transverse fluctuations of the endoplasmic reticulum in eukaryotic cells (of 10–1 000 nm length scales).[3]

1.1 How to Track Cells

Experimentally, tracking single cells is less demanding than single molecules due to their larger size, so it is a good place to start.[4] For strains of readily culturable bacteria, cells can be imaged with standard microscopy techniques using absorption contrast i.e. no complicated sample preparation is required, such as staining techniques. To track cells, first a sequence of well-resolved microscopy images need to be acquired. Standard types of imaging modality that can be used to create movies of dispersed bacteria or early stages of biofilms include *bright-field microscopy* (very high speeds are possible i.e. ~10^5 frames per second), *fluorescence microscopy* (specific labelling is possible, but the technique is relatively slow due to the low photon yield of fluorescent processes) and *confocal microscopy* (allows three-dimensional [3D]

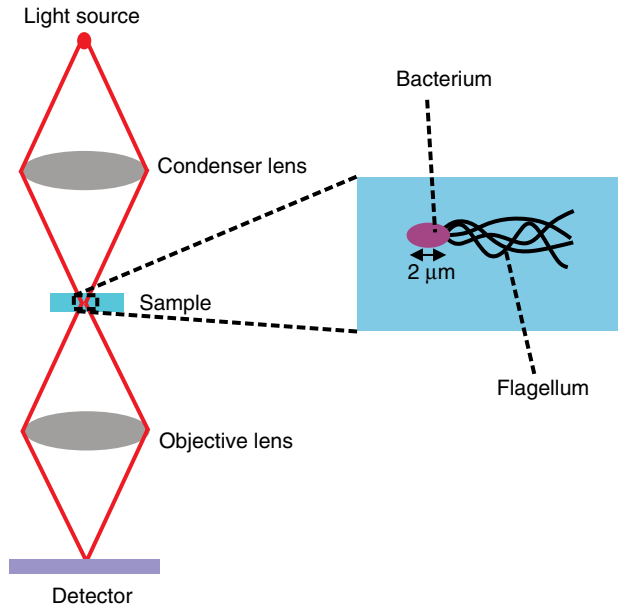

Figure 1.1 A simple *inverted optical microscope* in a *4f* configuration (*f* is the focal length of both the condenser and the objective lenses).[2] An *E. coli* cell is shown, and the flagella would be invisible without a dedicated contrast mechanism (fluorescence or phase contrast is needed).

imaging, but the scanning process of image acquisition often makes it slow). State-of-the-art super-resolution fluorescence microscopy techniques include *lattice sheet microscopy* that can achieve ~50 nm resolution at video rates (~50 frames per second)[5] and MINIFLUX (a variety of *stimulated emission depletion microscopy*, STED, Figure 13.10) that can achieve ~1 nm resolution at 1 000 frames per second.[6] The super-resolution techniques tend to be technically challenging (Section 13.2.4), and bright-field microscopy is much easier for beginner microscopists.

Ideally, movies of cells should be as long as possible, in terms of the number of frames, to maximise the amount of information available in the resultant tracks. Track length can be limited by the depth of focus of the microscope (z sectioning), field of view of the microscope (sampling in x and y), excessive particle speeds, the available memory on the camera (particularly an issue with ultrafast cameras), photobleaching of fluorescent labels and phototoxicity that damages the cells.

Once a movie of the cells has been made, the next challenge is to segment the images of cells using image analysis software. Gaussian trackers can be used to locate the positions of compact symmetrical bacterial cells that may be reasonably approximated by Gaussian functions, but more complicated cellular shapes need more sophisticated forms of segmentation, such as neural networks (NNs) or snakes algorithms[9] (Figures 1.2 and 1.4). Tracks are then made by connecting the centres of the segmented cells together in consecutive frames to form a linked list. Software searches for the closest positions of cells in consecutive images to link the cell centres together.

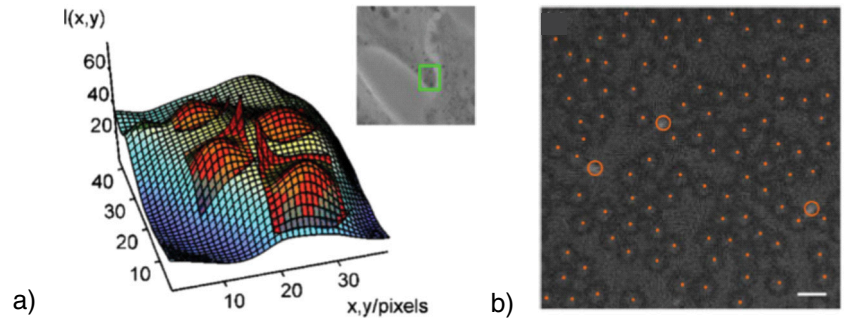

Figure 1.2 (a) A *Gaussian tracker* segments objects[7] in a human epithelial cell.[7] Endosomes (red, ~100 nm in size) are identified within a frame from a bright-field microscope (inset). (b) A *convolutional neural network* (CNN) segments *Bacillus subtilis* cells (red circles) immersed in a suspension of Brownian particles (red dots).[8]

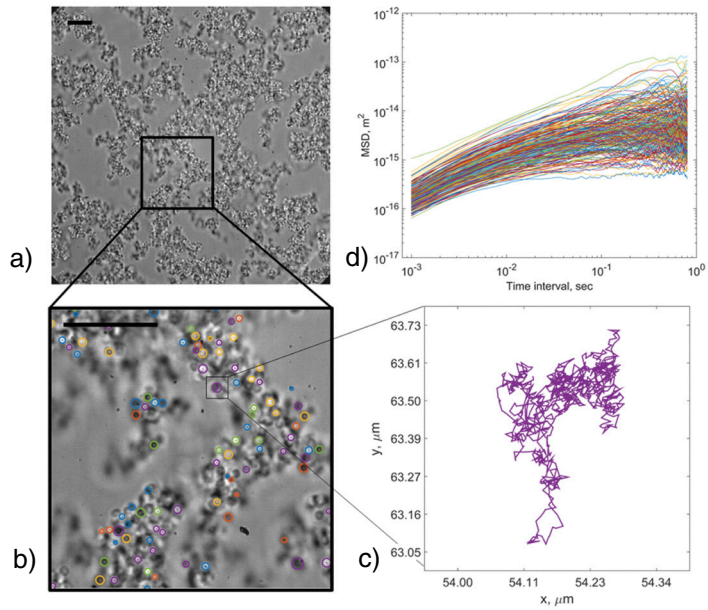

Figure 1.3 (a) An image of an early-stage *Staphylococcus aureus* biofilm from bright-field optical microscopy.[13,14] (b) Zoomed in region where single bacterial cells (1–2 μm in size) can be observed. (c) A track of a single *S. aureus* cell made using Gaussian tracker software (Figure 1.2a). (d) Mean square displacements (MSD) as a function of time interval of hundreds of single *S. aureus* cells in the biofilm calculated from the tracks. *S. aureus* is immotile, so the cells' motions are due to thermal forces modulated by the viscoelasticity of the biofilm.

If particles move substantial distances between consecutive images or the particle concentrations are too high, it can be an impossible task to unambiguously identify which particle contributes to which track. Particle positions can typically be measured with subdiffraction-limit resolution (often an order of magnitude improvement is possible on the diffraction limit) and sub-camera pixel resolution, because the weighted mean

of measurements of the optical centre of mass of a particle is used i.e. averages over many pixels are calculated. Thus cell positions can be routinely tracked with ~10 nm resolution at ~10^4 frames per second using standard optical microscopes combined with fast complementary metal oxide semiconductor cameras (Figure 1.3).[10] Higher resolutions have been achieved with quantum metrology using squeezed light.[11,12]

1.2 How to Track Single Molecules

Single molecules can now be routinely tracked both *in vitro* and *in vivo* inside single cells, although it was seen as a dramatic advance when single molecules were first imaged in condensed phases (Nobel Prize 2014).[15,16] Many people were surprised that it was possible to discriminate single molecules in condensed phases against backgrounds containing vast numbers of molecules of the order of Avogadro's number (6×10^{23}) with a sufficient signal-to-noise ratio (SNR). Single-molecule imaging techniques are primarily based on fluorescence microscopy, and this requires specific labelling of the molecules of interest with fluorophores. An emission filter based on the wavelength shift of emitted photons from a fluorophore compared with the wavelength of the excitation light source (the Stokes shift) can be used to discriminate single fluorescent molecules against the background of a huge number of non-fluorescent molecules.

Large catalogues are available for commercial fluorophores that can label biomolecules with varying degrees of specificity, such as proteins, nucleic acids, carbohydrates and lipids. The specificity of the labels needs to be determined in a biological experiment to be certain of what is labelled using careful control experiments due to the large number of factors that affect fluorophore binding. An elegant solution for labelling proteins is to genetically modify them to add an extra fluorescent protein domain to their structure. This can be very effective for *in vivo* studies, but green fluorescent proteins (GFPs) can suffer from fast photobleaching (synthetic fluorophores often are much more photostable), bulky GFPs can perturb protein functionality (control experiments are needed), and there are time lags introduced by the GFP transcription that can limit studies of fast intracellular dynamics.

For molecular imaging, the choice of segmentation algorithm is determined in part by the geometry of the molecule. Extended molecules with extensive labelling (e.g. a large DNA molecule in which all the base pairs are fluorescently labelled) require snakes algorithms (Figure 1.4), whereas molecules with point-like labelling often use Gaussian trackers (Figure 1.2a).[9] AI techniques (e.g. convolutional neural networks [CNNs]) can be more flexible in the types of molecular geometry they can analyse[19] but will suffer from poor SNRs if they are not properly constrained (Figure 1.2b). Often it is best to constrain artificial intelligence (AI) algorithms using simple physical models e.g. the probabilities of particle displacements can be constrained on the basis that particles will not teleport between different locations, which is a Bayesian approach. Current AI techniques often require extensive data sets to perform the training procedure i.e. they involve *supervised learning* (Chapter 14).

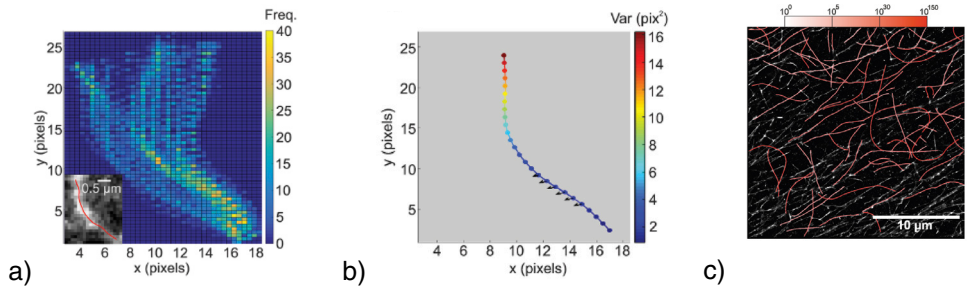

Figure 1.4 *Snakes algorithms* allow tracking of extended objects, such as the endoplasmic reticulum (ER) in human cells[3] or a peptide fibre in a gel.[17] (a) Time dependence of the tracked contours of the ER tubules from fluorescence microscopy. (b) Mean position of the ER tubule, where the variances of the transverse motions are highlighted. The tracked contours of ER tubules indicate active motion due to motor proteins. (c) Peptide fibre positions in a gel from fluorescence microscopy.[17,18]

The choice of algorithm to link the positions of segmented particles together into a track also has a variety of options[20] e.g. *nearest neighbour linking* or *multi-track optimisation* are possible. Particular care is required when particles closely approach one another (they can easily switch labels), and tracks can become fragmented due to low SNRs (they can be stitched together, but often with limited success). Our experience is that the segmentation algorithm plays a more important role than the linking algorithm in the quality of the final tracks, but both are important.

Bayesian tracking techniques (Chapter 14) can be very useful to remove the noise on tracks e.g. Kalman filtering,[21,22] and the methodology has been extensively developed with satellite imaging. However, care must be taken that this noise is random and Markovian (independent noise fluctuations occur with no memory), since it is an assumption used in Kalman filtering. It is particularly an issue when considering non-Markovian processes e.g. the motility of microorganisms or the intracellular motion of molecules are frequently non-Markovian.[23]

1.3 The Statistics of Structures

The static images of molecules and cells from microscopy experiments can provide a range of useful information e.g. calculating their sizes, conformations and relative organisation. Standard freeware software allows the segmentation of bacteria in microscopy images and can provide quantitative descriptors of cell shape.[24] Sophisticated software has also been developed to segment bacterial biofilms and quantify their structures in three dimensions.[25]

Different statistical tools are needed to quantify the relative positions of bacteria or the molecules associated with them. The *Ripley K function* ($K(r)$) quantifies the intuitive notion of whether particles have been placed at random across a surface or they

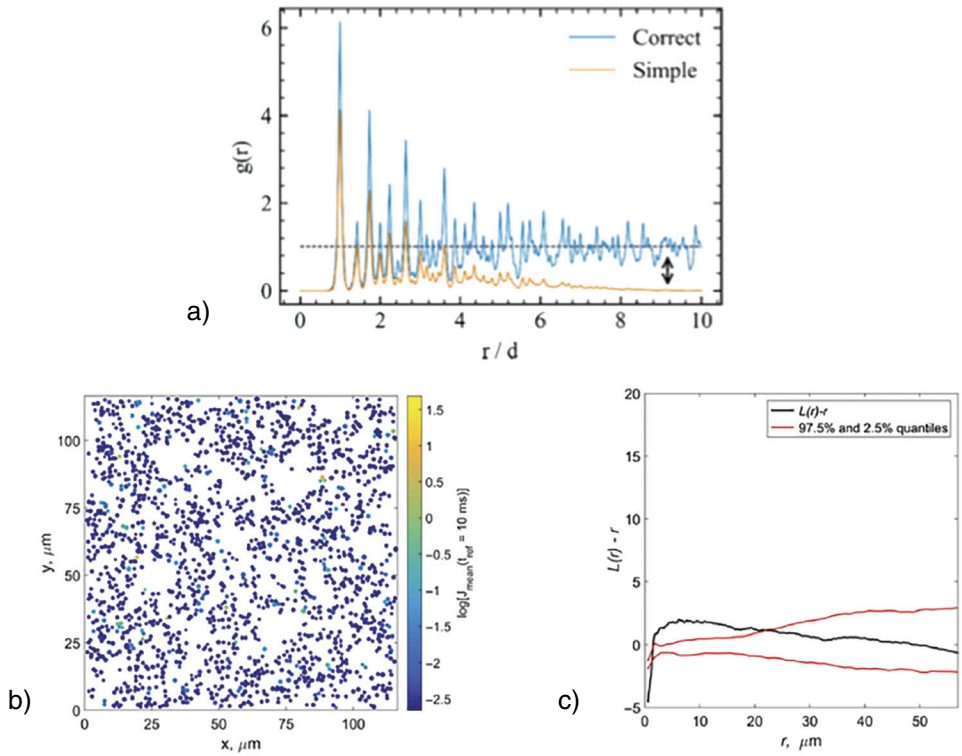

Figure 1.5 (a) The *radial distribution function* $g(r)$ for a random lattice of points[26] (blue). The red curve is from a naïve numerical calculation. (b) *Segmented positions* of *S. aureus* bacteria in a biofilm (coloured with a measure of the linear viscoelasticity via the creep compliance, $J(t)$, Chapter 10) and (c) the *Ripley K function* of the *S. aureus* bacteria in the biofilm[13] shown in (b) (r is the distance from a test point, $K(r)$ is the Ripley K function and $L(r) = \sqrt{\frac{K(r)}{\pi}}$ is the black line). It shows considerable clumping of the *S. aureus*.

are clustered together or dispersed.[27] It is defined so that $K(r)$ is the expected number of additional points within a distance r of a given point. This is a useful tool to understand the distributions of bacteria as they adsorb to surfaces. An alternative measure is given by the pair correlation function ($g(r)$), which is widely used in condensed matter physics, particularly liquid-state theory and models of colloidal matter.[28] $g(r)$ is defined as

$$g(r) = \frac{1}{2\pi r}\frac{\mathrm{d}K(r)}{\mathrm{d}r}, \tag{1.1}$$

where r is again the distance from a test point (Figure 1.5). With a stationary Poisson distribution of points, $g(r) = 1$ i.e. a complete random arrangement with no correlations. $g(r) < 1$ indicates an anti-correlation between points (dispersion), whereas $g(r) > 1$ indicates clustering.[27] $g(r)$ can be related to the interparticle potential if Boltzmann statistics are assumed for systems in thermal equilibrium[28,29] and has been

extensively developed in liquid-state theory. In anisotropic systems, $g(r)$ needs to be generalised[26] e.g. correlations along separate lattice directions should be averaged separately to maintain the additional information needed to quantify the degree of anisotropy, such as with liquid crystalline materials.

1.4 How to Analyse Particle Tracking Data

Statistical tools for handling tracking data can be very powerful. They play a central role in modern biological physics, since microscopy methods can provide high-resolution time series of images of living cells, biofilms and single molecules. Robust statistics are needed to test hypotheses on the behaviour of particles e.g. how they move, react, sense and oscillate. Furthermore, the analysis of particle tracking data can be conveniently extended to include tools from machine learning, since they have a common statistical basis, greatly increasing the possibilities for pattern recognition and large-scale automation[22] (Chapter 14).

Tracks of individual bacteria provide a rich source of information on their behaviour e.g. their motility, chemosensing and interactions. Statistical tools need to be applied to the tracks to make sense of them. A wide variety of *ad hoc* bespoke statistical parameters could be defined e.g. a bacterium is motile if its average velocity over 1 s is 1 μm s^{-1}, but they are often unsatisfactory. To choose between alternative possible statistical parameters, standard mathematically elegant methods are preferable, since they provide better prospects for quantitative comparison with both analytical models and simulations. They can also be more robust to varying experimental conditions and thus generalise more easily.

The transport of bacterial cells and molecules in the cells is often *anomalous* e.g. the central limit theorem breaks down and the probability density functions of their displacements are non-Gaussian. Mathematical models have been developed to describe anomalous transport (Chapter 2), although the relative merits of competing models are still debated.[30] A recent innovation is to train NNs on anomalous transport models since NNs can then provide the dynamic segmentation of tracks (Chapter 14).[31] Particle tracks represent a special case of *time series analysis*, which find wide-ranging applications inside and outside biology e.g. forecasting the stock market or diagnosing heart disease based on electrocardiograms.[32] There is thus a huge literature, and a wide range of mathematical tools have been developed.

A central tool for quantifying stochastic motion of particles is the *mean square displacement* (MSD, Chapter 2). For a random walk, the MSD $\left(\left\langle \Delta r^2 \right\rangle\right)$ has a simple scaling dependence on time, $\left\langle \Delta r^2 \right\rangle \sim t^1$ e.g. during Brownian motion (Figure 1.6). Furthermore, scaling of the MSDs is used to define anomalous transport, $\left\langle \Delta r^2 \right\rangle \sim t^\alpha$, where $\alpha \neq 0,1,2$, and this is the type of stochastic motion most commonly observed for cellular motility and the motion of larger molecules inside cells.[23,33] Average displacements $\left(\left\langle \Delta r \right\rangle\right)$ of particles are often not a useful measure of stochastic transport, since for symmetric stochastic processes, they average to zero, $\left\langle \Delta r \right\rangle = 0$. Higher moments

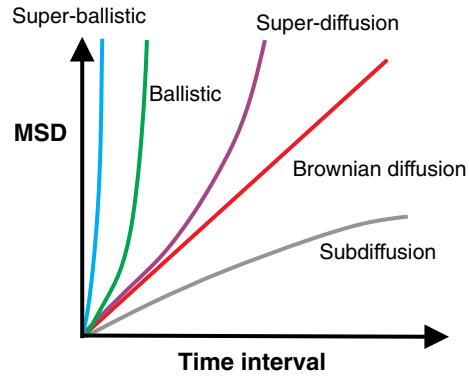

Figure 1.6 MSD as a function of time interval showing *sub-diffusive*, *diffusive* (Brownian), *super-diffusive*, *ballistic* and *super-ballistic* scaling behaviours. Note that super-ballistic scaling is rarely observed in low Reynolds number systems (they are overdamped), although it is possible in turbulent flows.

of the displacement probability distribution based on the third and fourth moments $\left(\left\langle \Delta r^3 \right\rangle \text{and} \left\langle \Delta r^4 \right\rangle\right)$ are useful for quantifying the skew and the degree of peakedness (the kurtosis), respectively. Moments of probability distributions of the displacements can provide average quantities to describe stochastic motility, which are reasonably robust to noise, but probability distribution functions (pdfs) contain additional information. Mathematically, the moment distribution is insufficient to unambiguously determine a pdf.[34]

For stationary statistical processes,[35] often MSDs are averaged over time, and the MSD is then considered as a function of time interval (TAMSD(τ)) i.e.

$$TAMSD(\tau)=\frac{1}{T-t}\int_0^{T-t}\left(\left(x_i\left(\tau+t\right)-x_i\left(\tau\right)\right)^2+\left(y_i\left(\tau+t\right)-y_i\left(\tau\right)\right)^2+\left(z_i\left(\tau+t\right)-z_i\left(\tau\right)\right)^2\right)dt \tag{1.2}$$

where T is the duration of the track and t is the time. x_i, y_i and z_i are the Cartesian coordinates of the particle i. The calculation of time-averaged MSDs (TAMSDs) can provide a major improvement in the SNR at short time intervals in experiments. If there are n steps in a track, the error bars scale as $\left(n-1\right)^{-1/2}$ for the shortest time interval of the TAMSD, $\left(n-2\right)^{-1/2}$ for the next shortest time interval and so on. Ensemble averaging of MSDs (EMSDs) over different particles is also possible to improve the SNR i.e. the MSDs are averaged over i in Equation (1.2).

There is a general theorem by Birkhoff from dynamical systems theory[36,37] that states TAMSD = EMSD for an *ergodic process*, and it can be used as a diagnostic for ergodicity breaking e.g. whether glassy behaviour occurs in the tracks. MSDs can be calculated in one, two and three dimensions, and their analysis conveniently generalises to different dimensionalities e.g. to describe the motion of a motor along a DNA chain (one-dimensional [1D]), a particle in the plane of focus of a conventional optical microscope (two-dimensional [2D]) or a particle in a confocal microscope (3D).

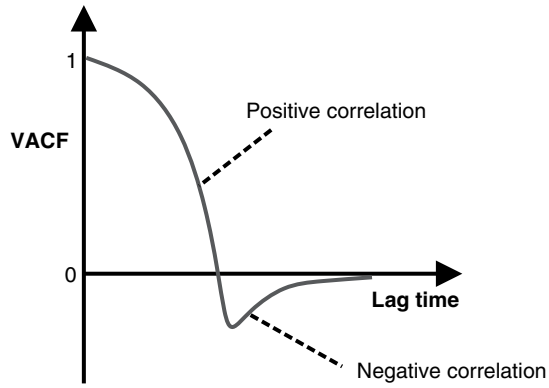

Figure 1.7 *Velocity autocorrelation function* (*VACF*) as a function of lag time for a particle moving inside a cell. The negative values of the *VACF* are due to anti-correlation.

Stochastic aging (SA) is a separate issue to ergodicity and is often observed in biology e.g. stochastic processes are not stationary and evolve with time during the growth of a cell.[37] SA can be diagnosed by delaying tracks by different aging times (i.e. chop off the start of the data corresponding to the aging time) and then comparing the resultant *MSD*s. Aging and glassy phenomena have direct implications in medicine e.g. scarring during wound healing that involves non-ergodic glassy fibrous composites of collagen.

Velocities need to be handled carefully with stochastic processes, since with random walks they depend on the time scale at which they are measured.[38] Often instantaneous velocities are defined in experiments as $\frac{\Delta r}{\Delta t}$ (Δr is the displacement of a particle over a time interval, Δt), but this quantity is sensitive to the choice of Δt e.g. a smaller choice of Δt can correspond to higher values of velocity for sub-ballistic processes. Lots of values of motor protein velocities in the literature are mishandled due to such issues and when faster cameras are manufactured, the quoted motor protein velocities often also increase. More robust methods to quantify velocities are to consider velocity autocorrelation functions (*VACF*) or velocities calculated via first passage probabilities[39] (see later). *VACF*s (Figure 1.7) can be defined as

$$VACF(t) = \int_0^{T-t-\delta} \frac{v(t'+t)v(t')}{T-t-\delta} dt', \tag{1.3}$$

$$v = \frac{r(t+\delta) - r(t)}{\delta}, \tag{1.4}$$

where v is the velocity at time t, δ is the time interval, T is the duration of the experiment and r is the displacement.

The use of probability distributions of *survival times* has its origins in medicine (Figure 1.8a).[40] Histograms of the number of patients in a medical trial can be plotted

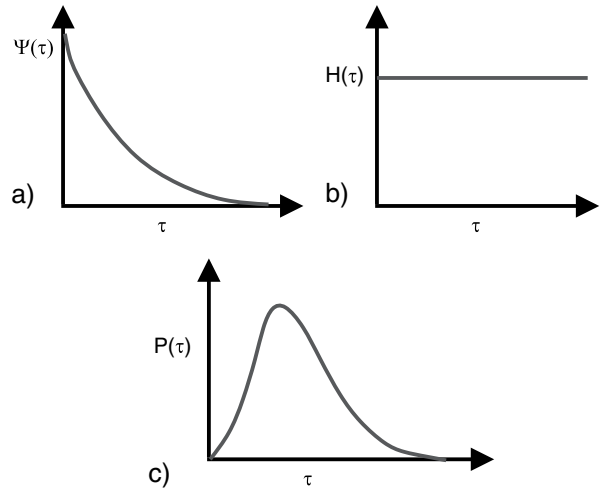

Figure 1.8 Plots of (a) the *survival time distribution* $\left(\Psi(\tau)\right)$, (b) the *hazard function* $\left(H(\tau)\right)$ and the *probability density function* $\left(P(\tau)\right)$ as a function of time (τ) for the run times of a bacterium.

as a function of time. If the death rate occurs at a constant value per unit time, the survival distribution has an exponential decay ($\Psi(\tau)$, a Poisson process). Decays due to more complex processes can be non-exponential, and hazard rates $\left(H(\tau)\right)$ can be introduced to make it easier to visualise them i.e. a constant hazard rate as a function of time is equivalent to a single exponential decay for the survival time (Figure 1.8b). The hazard rate is the rate of death of a subject of age t.

A practical problem for calculating survival times is when patients leave trials before they die, which biases the data due to a form of censuring. Kaplan–Meier estimators can be used to correct for these biases,

$$\Psi(t_i) = \left(1 - \frac{d_i}{n_i}\right)\Psi(t_{i-1}), \tag{1.5}$$

where $\Psi(t_i)$ is the survival distribution at time t_i, d_i is the number of events that happen at t_i and n_i is the number of events that survive up to t_i. Survival times can be used in the more general context of biological physics using such Kaplan–Meier corrections e.g. the run times of bacteria can be considered as a distribution of survival times. Biases introduced by the finite length of tracks in tracking experiments can be corrected using Equation (1.5). Survival times can also be used to understand the residence times of bacteria on surfaces.[41] Survival times (e.g. for runs, Figure 1.8a) can be simpler than just considering histograms (or pdfs), since they are monotonic decreasing functions (in contrast, the run time pdfs, $P(\tau)$, will be peaked, Figure 1.8c).

The *first passage probability* (FPP) for particle tracks is defined as the probability distribution for the times a particle takes to travel a specific distance for the first time[43] (Figure 1.9a). The *mean* FPP $\left(\text{MFPP}, v_{FPP}\right)$ is the mean of the FPP distribution.

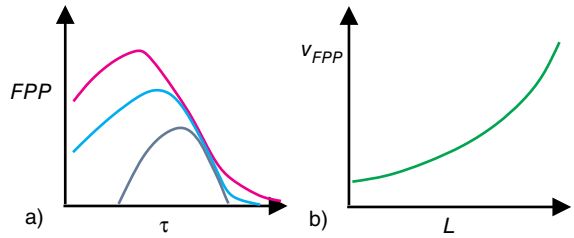

Figure 1.9 (a) *First passage probability* (FPP) of particles moving inside a cell as a function of time (τ), red < blue < navy blue correspond to longer transit lengths. (b) *Mean first passage velocity* (v_{FPP}) of the particles as a function of transit length (L).[39,42]

In numerous biological situations, the FPP is the crucial statistical quantity of interest e.g. times for a chemical reaction to occur or for a particle to leave a maze. MFPPs for particles with multiple scaling regimes as a function of time imply that the reaction kinetics of the particle will also exhibit multiple regimes. Furthermore, FPPs can also provide more robust alternatives to instantaneous velocities to quantify motility and can help separate up the motility of particles using the average FPP velocity (v_{FPP}) as a function of transit length (L, Figure 1.9b) e.g. the question can be asked as to whether long-range transits happen at larger velocities, which is useful with endosomal transport.

MSDs are insensitive to direction (so too are the survival times and the FPPs), and they just provide a measure of the amplitude of motion as a function of time. *Angular correlations* of particle displacements thus provide crucial information to understand particle motility[44] with respect to direction. Analogous to an MSD, the average direction cosine $\left(\left\langle \cos\theta(\tau)\right\rangle\right)$ for segments along a track can be quantified as a function of time interval (averaged in an analogous manner to a TAMSD, Equation (1.1), although three points are required to define the consecutive displacements Δr_i and Δr_{i+1}), and a scalar product is used,

$$\left\langle \cos\theta(\tau)\right\rangle = \frac{\left\langle \Delta r_i . \Delta r_{i+1}\right\rangle}{\left|\Delta r_i\right|\left|\Delta r_{i+1}\right|}. \tag{1.6}$$

Cosines are bounded functions; $-1 \le \cos\theta \le 1$. Negative values of the angular correlation function correspond to anti-persistent motion i.e. the particle is constantly changing direction and tends to move back on itself. $\left\langle \cos\theta\right\rangle = 0$ corresponds to no average directional bias and is expected for an unbiased random walk. Positive values of $\left\langle \cos\theta(\tau)\right\rangle$ correspond to *directional persistence*. Such measures of directionality are useful for the development of models for bacteria, since bacteria act as stochastic swimmers, and for the motility of intracellular cargoes within bacterial cells. Some similar information is encoded in velocity correlation functions (Equation 1.2), but it is useful to have both measures.

More sophisticated statistical measures are needed to describe the correlated motion of particles. Two-point correlation functions are one possibility, $\left\langle \Delta r_1 \Delta r_2 \right\rangle$, where Δr_1 and Δr_2 are the displacements for two different particles, which has been studied from

the perspective of two-particle microrheology.[45] Velocity cross-correlation functions $\left(\langle v_1 v_2 \rangle\right)$ are also useful to study the mutual motion of cells e.g. in chemotactic fields, and provide similar information to $\langle \Delta r_1 \Delta r_2 \rangle$.

Flocking order parameters (Φ) have also been introduced to describe phase transitions during coherent motion in motile particles[46] (such as bacteria, starlings and ants) e.g.

$$\Phi = \frac{1}{Nv}\left|\sum_i v_i\right|, \tag{1.7}$$

where N is the number of particles, v is the average velocity and v_i is the velocity of each particle. Care must be taken in calculating Φ for particles that experience anomalous transport, since the values of v_i will depend on time for non-ballistic particle motion, and other order parameters have been suggested to make them more robust.[47]

A sophisticated modern approach to the motility of both particles inside bacteria and cellular motility follows a framework of *heterogeneous anomalous transport* (HAT).[48] This attempts to quantify the heterogeneity of the anomalous transport of particles in both space and time by considering generalised diffusion coefficients $\left(D_\alpha(r,t)\right)$ and scaling exponents $(\alpha(r,t))$ that vary in space and time, defined via the MSDs of the distributions using

$$\left\langle \Delta r^2 \right\rangle = 2nD_\alpha\left(r,t\right)\tau^{\alpha(r,t)}, \tag{1.8}$$

where n is the number of dimensions. Note that D_α has fractional units, which provides some challenges e.g. it is not possible to plot D_α on a single axis. Rescaling D_α by characteristic length and time scales solves many of these problems. Thus, values of both D_α and α are allowed to vary with time and space during the analysis, which corresponds to a multi-fractal model.[23] There is good evidence that HAT occurs for the majority of cellular motility and intracellular motility of large molecules and aggregates.

Experiments with extended linear objects, such as single molecules, aggregates of cells, organelles or individual cells, lend themselves to Fourier analysis of data segmented using snakes algorithms[49] (Figure 1.4). The equipartition theorem can be used to calculate the energy of each Fourier mode assuming the fibres are in thermal equilibrium and simple continuum models are used for the energy of the snakes e.g. all the energy is stored in Hookean bending modes.[49] Similar analysis is also possible with cell membranes in two dimensions.[50] Challenges occur to describe systems in which quenched disorder or active transport affect the conformations of the extended objects.[3,17]

Other less direct methods of calculating stochastic processes occur in the literature. For example, the square of the Fourier transform of the particle displacement as a function of the correlation time is called the power spectral density ($P(\omega)$) and is often measured in optical tweezer experiments. The information content is similar to a MSD as a function of time interval, but MSDs are often simpler to work with.

1.5 Scattering Alternatives

Instead of working directly with images, *scattering experiments* function in reciprocal space. Historically, scattering techniques were used in situations where imaging was not possible e.g. in experiments with hard X-rays or thermal neutrons where it is hard to construct an imaging lens, such as the Braggs' initial work to study the structure of simple crystals, such as sodium chloride. Images tend to be preferred in modern-day biological physics experiments, since they are easier to interpret and tend to be less ambiguous, but scattering data can also be useful. Inelastic scattering experiments (e.g. dynamic light scattering, X-ray photon correlation spectroscopy or quasi-elastic neutron scattering) detect small energy changes in scattered radiation and often provide much faster dynamic information than is currently possible with imaging experiments e.g. point detectors can stream data much faster than pixel arrays. In bacterial biophysics, *fluorescence correlation spectroscopy* (FCS) and *dynamic differential microscopy* (DDM) are commonly used scattering techniques and can be microscope based to improve their spatial sensitivity.

In DDM, a movie of a biological system is made with a microscope and versions are possible using both coherent (e.g. bright-field contrast) and incoherent (e.g. fluorescence) scattering. Software correlators can then be used with stacks of the images to calculate correlation functions that describe the image dynamics (Figure 1.10). The correlation functions allow access to identical information as inelastic scattering experiments, although they are at relatively slow time scales due to the update times of pixel arrays used on standard digital cameras in optical microscopes.

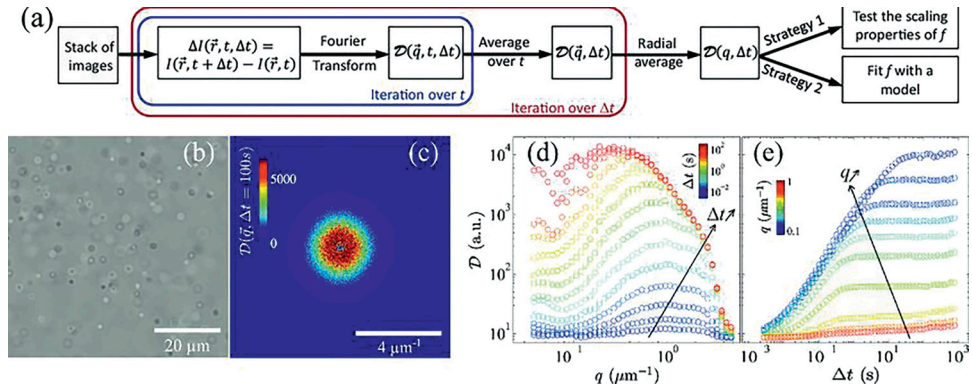

Figure 1.10 (a) A schematic diagram of the algorithm used to analyse DDM[51] experiments. (b) An example of the difference of two images $\left(d\left(r,t,\tau\right)\right)$. (c) The square of the Fourier transform of the difference of two images. (d) $D\left(q,\tau\right)$ (the square of the Fourier transform of $d\left(r,t,\tau\right)$ averaged over t) as a function of momentum transfer, q. (e) $D\left(q,\tau\right)$ as a function of time interval $\left(\Delta t = \tau\right)$. Reprinted from [Germain D., Leocmach M., Gibaud T., Differential dynamics microscopy to characterise Brownian motion and bacteria motility. *American Journal of Physics* 2016, 84, 202], with the permission of AIP Publishing.

Intermediate scattering functions $\left(\text{ISF}, f\left(q,\tau\right)\right)$ are a key statistical tool used in inelastic scattering experiments to quantify the dynamics e.g. with light, neutrons and X-ray.[52] DDM can be used to extract the ISF from stacks of images. There needs to be a source of speckle on the images (which is non-necessarily due to coherent scattering) and the images do not need to be particularly well resolved. DDM works well on images from both bright-field (with both laser and light-emitting diode [LED] illumination) and fluorescence microscopy (the calculations are slightly different in each case). Taking differences between images suppresses the noise due to stationary particles and detector heterogeneities[53] (hence the name DDM), which gives

$$d\left(r,t,\tau\right)=I\left(r,t_0+\tau\right)-I\left(r,t_0\right). \tag{1.9}$$

Next, these image differences are Fourier transformed in space (q is the momentum transfer) and squared,

$$D(q,\tau)=\left\langle \left|d(q,t_0,\tau\right|^2\right\rangle_{t_0}. \tag{1.10}$$

The ISF $\left(f\left(q,\tau\right)\right)$ is constructed by fitting $A(q)$ and $B(q)$ using

$$D(q,\tau)=A(q)\left[1-f\left(q,\tau\right)\right]+B(q), \tag{1.11}$$

where $A(q)$ and $B(q)$ are assumed arbitrary smooth functions.

In DDM, a major advantage compared with tracks from direct imaging is that the dynamics can be averaged and quantified without segmentation of the images. The neglect of segmentation can be an advantage for complex hierarchical structures e.g. the endoplasmic reticulum in eukaryotic cells, where it can be hard to unambiguously locate the objects' boundaries.[3] Challenges with DDM are that some spatial information is lost in the averaging procedures (e.g. during the calculation of the Fourier transforms), analytic calculations are slightly harder in reciprocal space and it can be more challenging to determine which specific structures are being analysed.

Fluorescence correlation spectroscopy considers the fluctuations in fluorescent emission from a small volume that is illuminated in a sample[54] (Figure 1.11). The fluctuations in fluorescent emission can be related to the motion of the fluorophores through the calculation of correlation functions ($G(\tau)$, where τ is the correlation time) and can be performed relatively quickly (using fast point photodetectors) and thus experiments can be performed with quickly photobleaching fluorophores. If only a single detection volume is used in the sample, it is harder to explore the length dependence of the dynamic processes using FCS (via q, the momentum transfer). This can make the exploration of anomalous transport more challenging (the spatial dependence of the motility is ambiguous) and a partial solution is to use a range of pin hole sizes, so different volumes are illuminated in the sample.[55] In this case, the fluorophores need to be long-lived and the statistical processes must be stationary (they should not evolve with time).

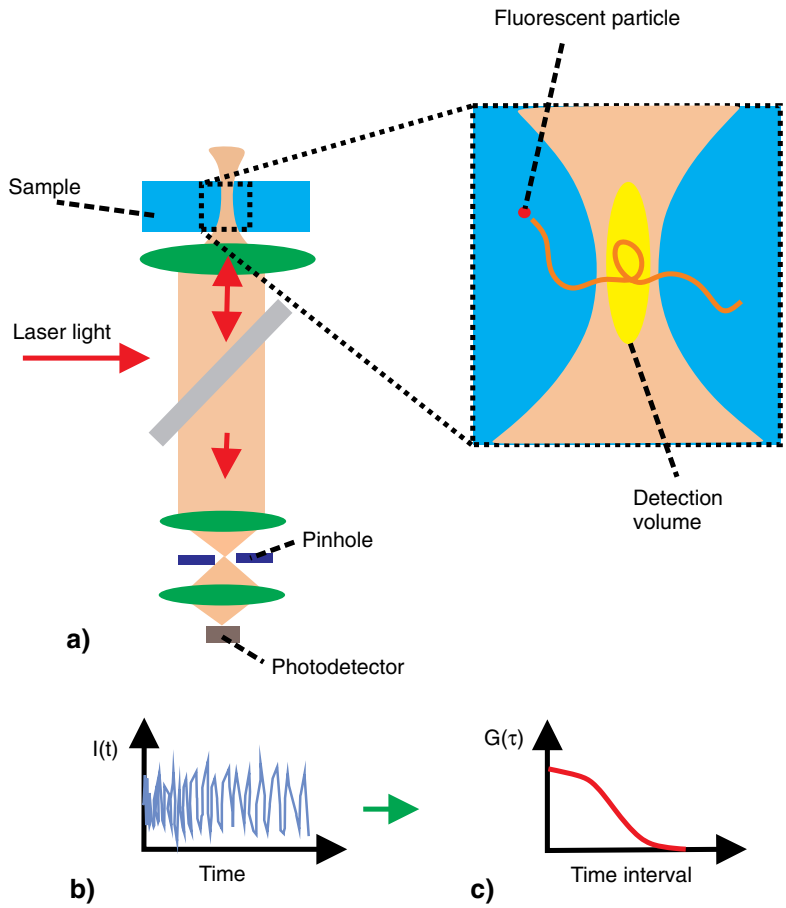

Figure 1.11 (a) Apparatus for FCS is based on a fluorescence microscope, which uses a pinhole to define the detection volume. (b) The intensity of light ($I(t)$) emitted by fluorophores in the detection volume as a function of time. (c) A correlation function ($G(\tau)$) of the intensity fluctuations as a function of time interval (τ) from (b). The correlation function can be used to quantify the motion of the fluorescence particles within the detection volume.

Suggested Reading

Höfling, F.; Franosch, T., Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics* **2013**, *76*, 046602. Good overview of the experimental evidence for anomalous transport inside cells.

Ibe, O. C. *Elements of Random Walks and Diffusion Processes*. Wiley: 2013. Introduces anomalous transport (e.g. fractional Brownian motion) in an intelligible manner for non-mathematicians. Also acts as a good primer for stochastic processes.

Klafter, J.; Sokolov, I. M., *First Steps in Random Walks: From Tools to Applications*. Oxford University Press: 2011. Short and fairly mathematical introduction to some modern models for anomalous transport.

Waigh, T. A.; Korabel, N., Heterogeneous anomalous transport in molecular and cellular biology. *Reports on Progress in Physics* **2023**, *86*, 126601. Considers some challenges in the modelling of anomalous transport in cellular biology e.g. multi-fractal effects.

References

1. Lane, N., The unseen world: Reflections on Leeuwenhoek (1677) 'Concerning little animals'. *Philosophical Transactions of the Royal Society B* **2015**, *370* (1666), 20140344.
2. Mertz, J., *Introduction to Optical Microscopy*. Cambridge University Press: 2019.
3. Perkins, H. T.; Allan, V. J.; Waigh, T. A., Network organisation and the dynamics of tubules in the endoplasmic reticulum. *Scientific Reports* **2021**, *11* (1), 16230.
4. Dubay, M. M.; Acres, J.; Riebeles, M.; Nadeau, J. L., Recent advances in experimental design and data analysis to characterise prokaryotes motility. *Journal of Microbiological Methods* **2023**, *204*, 106658.
5. Chen, B. C.; et al., Lattice light-sheet microscopy-imaging molecules to embryos at high spatiotemporal resolution. *Science* **2014**, *346* (6208), 1257998.
6. Wolff, J. O.; Scheiderer, L.; Engelhardt, T.; Maththias, J.; Hell, S. W., MINIFLUX dissects the unimpeded walking of kinesin-1. *Science* **2023**, *379* (6636), 1004–1010.
7. Rogers, S. S.; Waigh, T. A.; Zhao, X.; Lu, J. R., Precise particle tracking against a complicated background: Polynomial fitting with Gaussian weight. *Physical Biology* **2007**, *4* (3), 220–227.
8. Helgadottir, S.; Argua, A.; Volpe, G., Digital video microscopy enhanced by deep learning. *Optica* **2019**, *6* (4), 506.
9. Szeliski, R., *Computer Vision: Algorithms and Applications*, 2nd ed. Springer: 2022.
10. Waigh, T. A., Advances in the microrheology of complex fluids. *Reports on Progress in Physics* **2016**, *79* (7), 074601.
11. Xu, C.; Zhang, L.; Huang, S.; Ma, T.; Liu, F.; Yonezawa, H.; Zhang, Y.; Xiao, M., Sensing and tracking enhanced by quantum squeezing. *Photonics Research* **2019**, *7* (6), 14.
12. Taylor, M. A.; Janousek, J.; Daria, V.; Knittel, J.; Hage, B.; Bachor, H. A.; Bowen, W. P., Biological measurement beyond the quantum limit. *Nature Photonics* **2013**, *7* (3), 229–233.
13. Hart, J. W.; Waigh, T. A.; Lu, J. R.; Roberts, I. S., Microrheology and spatial heterogeneity of *Staphylococcus aureus* biofilms modulated by hydrodynamic shear and biofilm-degrading enzymes. *Langmuir* **2019**, *35* (9), 3553–3561.

14. Rogers, S. S.; van der Walle, C.; Waigh, T. A., Microrheology of bacterial biofilms *in vitro*: *Staphylococcus aureus* and *Pseudomonas aeruginosa*. *Langmuir* **2008**, *24* (23), 13549–13555.
15. Moerner, W. E.; Kardor, L., Optical detection and spectroscopy of single molecules in a solid. *PRL* **1989**, *62* (21), 2535–2538.
16. Leake, M. C., *Single-Molecular Cellular Biophysics*. Cambridge University Press: 2013.
17. Cox, H.; Xu, H.; Waigh, T. A.; Lu, J. R., Single-molecule study of peptide gel dynamics reveals states of prestress. *Langmuir* **2018**, *34* (48), 14678–14689.
18. Cox, H.; Cao, M.; Xu, H.; Waigh, T. A.; Lu, J. R., Active modulation of states of prestress in self-assembled short peptide gels. *Biomacromolecules* **2019**, *20* (4), 1719–1730.
19. Newby, J. M.; Schaefer, A. M.; Lee, P. T.; Forest, M. G.; Lai, S. K., Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D. *Proceedings of the National Academy of Sciences* **2018**, *115* (36), 9026–9031.
20. Chenouard, N.; et al., Objective comparison of particle tracking methods. *Nature Methods* **2014**, *11* (3), 281–289.
21. Wu, P. H.; Agarwal, A.; Hess, H.; Khargonekar, P. P.; Tseng, Y., Analysis of video-based microscopic particle trajectories using Kalman filtering. *Biophysical Journal* **2010**, *98* (12), 2822–2830.
22. Murphy, K. P., *Probabilistic Machine Learning: An Introduction*. MIT: 2022.
23. Waigh, T. A.; Korabel, N., Heterogeneous anomalous transport in cellular and molecular biology. *Reports on Progress in Physics* **2023**, *86* (12), 126601.
24. Dacret, A.; Quardokus, E. M.; Brun, Y. V., MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nature Microbiology* **2016**, *1* (7), 1.
25. Hartmann, R.; et al., Quantitative image analysis of microbial communities with BiofilmQ. *Nature Microbiology* **2021**, *6* (2), 151.
26. Kopera, B. A. F.; Retsch, M., Computing the 3D radial distribution function from particle positions: An advanced analytic approach. *Analytic Chemistry* **2018**, *90* (23), 13909–13914.
27. Holmes, S.; Huber, W., *Modern Statistics for Modern Biology*. Cambridge University Press: 2019.
28. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Oxford University Press: 2017.
29. Hansen, J. P.; McDonald, I. R., *Theory of Simple Liquids: With Applications to Soft Matter*. Academic Press: 2013.
30. Metzler, R.; Klafter, J., The restaurant at the end of the random walk. *Journal of Physics A: General Physics* **2004**, *37* (31), R161–R208.
31. Han, D.; Korabel, N.; Chen, R.; Johnston, M.; Gavrilova, A.; Allan, V. J.; Fedotov, S.; Waigh, T. A., Deciphering anomalous heterogeneous intracellular transport with neural networks. *eLife* **2020**, *9*, e52224.
32. Nielsen, A., *Practical Time Series Analysis: Prediction with Statistics and Machine Learning*. O'Reilly: 2020.

33. Hofling, F.; Franosch, T., Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics* **2013**, *76*, 046602.
34. Sornette, D., *Critical Phenomena in Natural Sciences*. Springer: 2003.
35. Ibe, O. C., *Elements of Random Walks and Diffusion Processes*. Wiley: 2013.
36. Birkhoff, G. D., Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences* **1931**, *17* (12), 656–660.
37. Korabel, N.; Taloni, A.; Pagnini, G.; Allan, V. J.; Fedotov, S.; Waigh, T. A., Ensemble heterogeneity mimics ageing for endosomal dynamics within eukaryotic cells. *Scientific Reports* **2023**, *13* (1), 8789.
38. Berg, H. C., *Random Walks in Biology*. Princeton University Press: 1993.
39. Rogers, S. S.; Flores-Rodriguez, N.; Allan, V. J.; Woodman, P. G.; Waigh, T. A., The first passage probability of intracellular particle trafficking. *PCCP* **2010**, *12* (15), 3753–3761.
40. Aalen, O.; Borgan, O.; Gjessing, H., *Survival and Event History Analysis: A Process Point of View*. Springer: 2008.
41. Blee, J. A.; Roberts, I. S.; Waigh, T. A., Membrane potentials, oxidative stress and the dispersal response of bacterial biofilms to 405 nm light. *Physical Biology* **2020**, *17* (4), 036001.
42. Flores-Rodriguez, N.; Rogers, S. S.; Kenwright, D. A.; Waigh, T. A.; Woodman, P. G.; Allan, V. J., Roles of dynein and dynactin in early endosome dynamics revealed using automated tracking and global analysis. *PLOS One* **2011**, *6* (9), e24479.
43. Redner, S., *A Guide to First Passage Processes*. Cambridge University Press: 2001.
44. Harrison, A. W.; Kenwright, D. A.; Waigh, T. A.; Woodman, P. G.; Allan, V. J., Modes of correlated angular motion in live cells across three distinct time scales. *Physical Biology* **2013**, *10* (3), 036002.
45. Levine, A. J.; Lubensky, T. C., One- and two-particle microrheology. *Physical Review Letters* **2000**, *85*, 1774.
46. Vicsek, T.; Czirok, A.; Ben-Jacob, E.; Cohen, I.; Shochet, O., Novel type of phase transition in a system of self-driven particles. *Physical Review Letters* **1995**, *75* (6), 1226–1229.
47. Cavagna, A.; Giardina, I.; Grigera, T. S., The physics of flocking: Correlation as a compass from experiments to theory. *Physics Reports* **2018**, *728* (3), 1–62.
48. Itto, Y.; Beck, C., Superstatistical modelling of protein diffusion dynamics in bacteria. *Journal of the Royal Society – Interface* **2021**, *18* (176), 20200927.
49. Gittes, F.; Mickey, B.; Nettleton, J.; Howard, J., Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape. *Journal of Cellular Biology* **1993**, *120* (4), 923–934.
50. Monzel, C.; Sengupta, K., Measuring shape fluctuations in biological membranes. *Journal of Physics D: Applied Physics* **2016**, *49* (24), 243002.
51. Germain, D.; Leocmach, M.; Gibaud, T., Differential dynamic microscopy to characterize Brownian motion and bacteria motility. *American Journal of Physics* **2016**, *84* (3), 202.

52. Berne, B. J.; Percora, R., *Dynamic Light Scattering: With Applications to Chemistry, Biology and Physics*. Dover: 2003.
53. Cerbino, R.; Cicuta, P., Perspective: Differential dynamic microscopy extracts multiscale activity in complex fluids and biological systems. *Journal of Chemical Physics* **2017**, *147* (11), 110901.
54. Rigler, R.; Elson, E. S., *Fluorescence Correlation Spectroscopy: Theory and Applications*. Springer: 2001.
55. Stolle, M. D.; Fradin, C., Anomalous diffusion in inverted variable-lengthscale fluorescence correlation spectroscopy. *Biophysical Journal* **2019**, *116* (5), 791–806.