POLITECNICO DI TORINO

MASTER THESIS
in Physics of Complex Systems

Analysis & modelling of biophysical processes from single-particle tracking data

October 2018

Mentors

Andrea Pagnani
Politecnico di Torino

David Holcman
École Normale Supérieure, Paris

Candidate

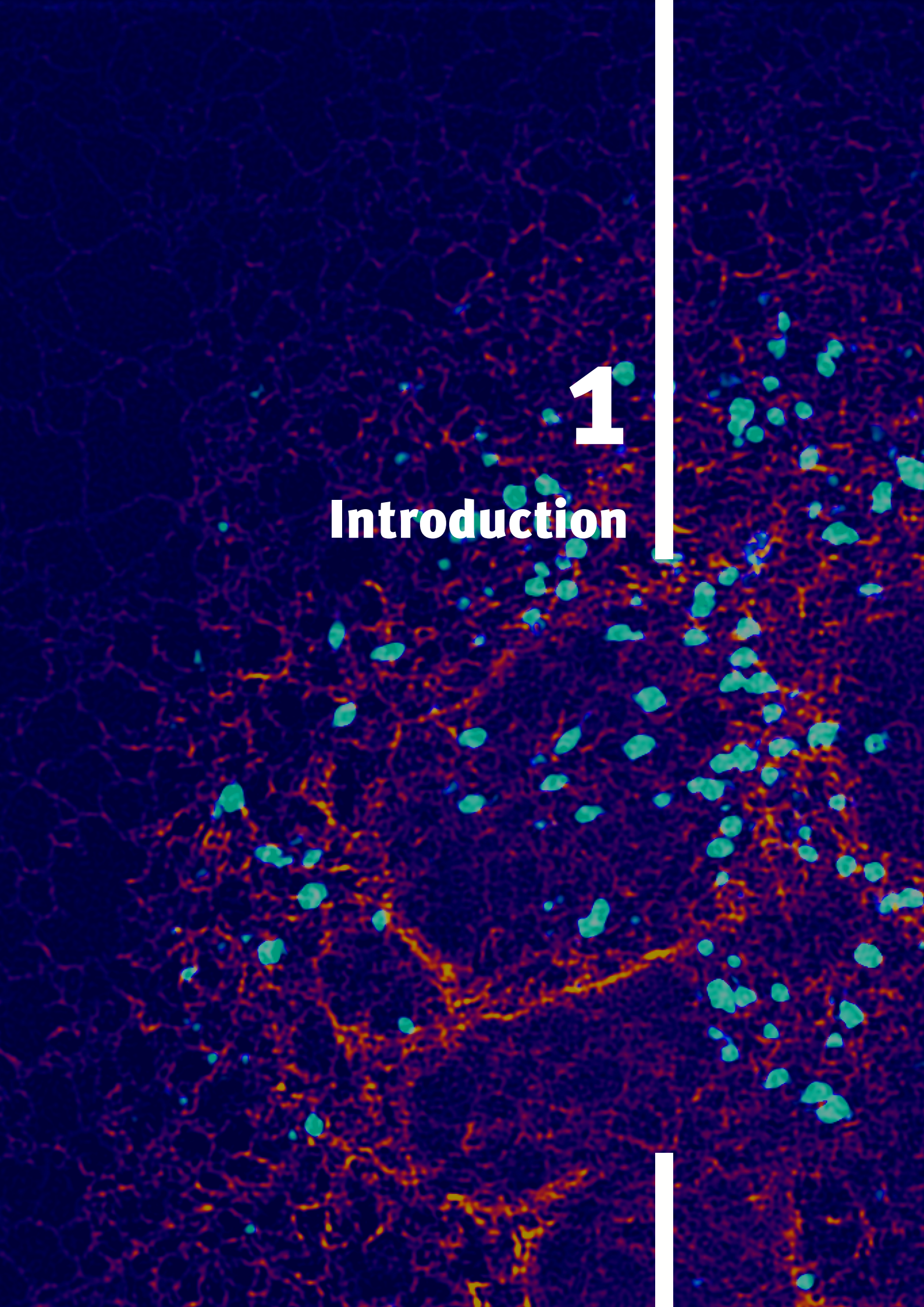
Matteo Dora

Typeset by the author in Meta & Meta Serif.

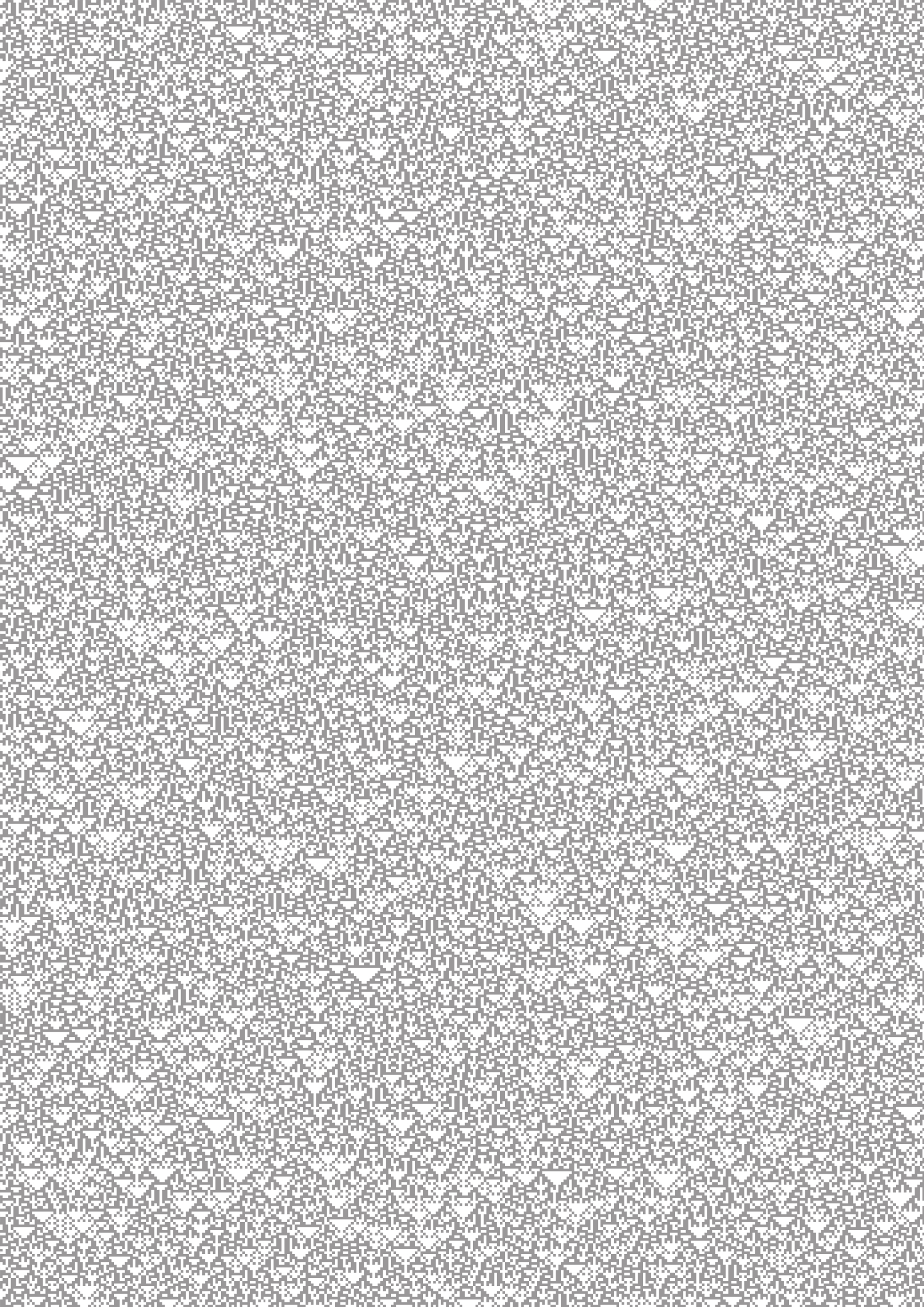
Contents

Les théories servent à irriter les philistins, à séduire les esthètes et à faire rire les autres.

— Amélie Nothomb, *Le sabotage amoureux*



1 Introduction



1.1 Overview

Super resolution microscopy has become, with recent developments, a fundamental source of data in the field of biology. On top of this, imaging techniques like PALM have been combined with single-particle tracking,[17] allowing to reconstruct thousands of simultaneous particle trajectories at increasingly high spatiotemporal resolution. With the availability of this large quantity of information, new data analysis techniques are needed to extract relevant biophysical features that can give new insights on the biological processes. Existing knowledge from physics and mathematics can provide a solid base to build a toolset of statistical and analytical methods to accomplish this goal. Langevin's equation and stochastic processes, for example, have been extensively—and successfully—used to model the dynamics of biological systems[24]. Likewise, regression and unsupervised learning algorithms represent a powerful tool to classify the large amount of data produced by single-particle tracking. Valuable insights can be obtained by carefully joining data driven methods and theoretical modelling. Either approaches are valid but may not be sufficient, per se, to reveal the inner significance of certain processes; their combination can instead provide hints about the deeper mechanisms of life. Redundancy, for example, seems to be a fundamental principle in many biological processes, as the joint application of narrow escape model and numerical simulations has suggested[21,25,2]. Yet to be understood problems in biology can benefit hugely from this twofold approach. Gathering of high resolution datasets paves the way to new and promising discoveries. On the other hand, data is useless without a solid framework to analyse and explain it. Overwhelming availability of data can even be counterproductive, as it makes the extraction of meaningful patterns more complicated. Dealing with these challenges is going to be, in the very near future, one of the major efforts in all fields of scientific research.

In this thesis I will present some results regarding data analysis, modelling and numerical simulations based on single-particle tracking data. The *fil rouge* connecting the whole work is the search for explanation of the dynamics of amyloid beta (Aβ) aggregates in neurons, which are known to be involved in Alzheimer's disease[4]. In this introductive chapter I will briefly review the current knowledge about Alzheimer's and the main techniques in super resolution microscopy. In chapter 2 I will present the analysis of SPT data. While the dataset used specifically regards Aβ aggregates, the final aim is to build a general methodological framework to allow the extraction of biophysical features from SPT recordings in different contexts. Following the results of the data analysis, in chapter 3 I will reformulate the initial hypothesis and focus on a model of motion[9] involving the endoplasmic reticulum (ER), describing the results of numerical simulations that I developed to better understand the biological implications of the model.

[17] Manley et al. *High-density mapping of single-molecule trajectories with photoactivated localization microscopy* (2008)

[24] Schuss. *Theory and applications of stochastic processes: an analytical approach* (2010)

[21] Reynaud et al. *Why so many sperm cells?* (2015)

[25] Schuss, Basnayake and Holcman. *Redundancy principle for optimal random search in biology* (2017)

[2] Basnayake et al. *Extreme Narrow Escape: shortest paths for the first particles to escape through a small window* (2018)

[4] Eisenberg and Jucker. *The Amyloid State of Proteins in Human Diseases* (2012)

[9] Holcman et al. *Single particle trajectories reveal active endoplasmic reticulum luminal flow* (2018)

1.2 Alzheimer's disease

Alzheimer's disease is the most common form of dementia in elder adults. As population ages, the death rate due to the disease is steadily increasing. Statistics published by the Centers for Disease Control and Prevention show that Alzheimer's-related deaths in USA increased 55% between 1999 and 2014,[27] making it the sixth

[27] Taylor et al. *Deaths from Alzheimer's Disease—United States, 1999–2014* (2017)

leading cause of death. The disease causes a progressive loss of brain functions (neuron deaths) to a point where it impaires basic cognitive skills. Most common symptoms are memory loss and language impairment.

As of today, the causes of Alzheimer's are not fully understood by the scientific community. Despite several clinical trials, no treatment has been proved to be—even partially—effective in treating the disease. Most scientists agree that probably there is no single cause for Alzheimer's, but a combination of several genetic and environmental factors.

The amyloid hypothesis

A long standing attempt to explain Alzheimer's disease is the *amyloid hypothesis*. One of the most distinctive characteristics of people affected by the disease is the formation of amyloid-β (Aβ) aggregates in the brain extracellular space (fig. 1.1). These clumps, known as *plaques*, are thought to start a cascade process leading to brain inflammation, cells deaths and synapses dysfunction. Despite solid evidence proving the association between Alzheimer's disease and Aβ plaques, the amyloid hypothesis didn't lead to any beneficial result in terms of medical therapy.[14,16] Various treatments targeting Aβ aggregates have been trialled but, even if successful in dissolving plaques, no effect on patients' cognitive functions has been observed.[10] The general explanation of these disappointing results is that treatments are administered too late in the disease progression, when plaques have already activated an uncontrollable cascade of damaging processes. Supporting this idea, it has been shown that formation of Aβ plaques can begin decades before the manifestation of symptoms. Despite being often criticised for its failures, the amyloid theory remains one of the main rational approach in Alzheimer's disease research.[6]

[14] Lancet. *Alzheimer's disease: expedition into the unknown.* (2016)

[16] Makin. *The amyloid hypothesis on trial.* (2018)

[10] Honig et al. *EXPEDITION 3: a phase 3 trial of solanezumab in mild dementia due to Alzheimer's disease* (2016)

[6] Hardy and Selkoe. *The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics* (2002)

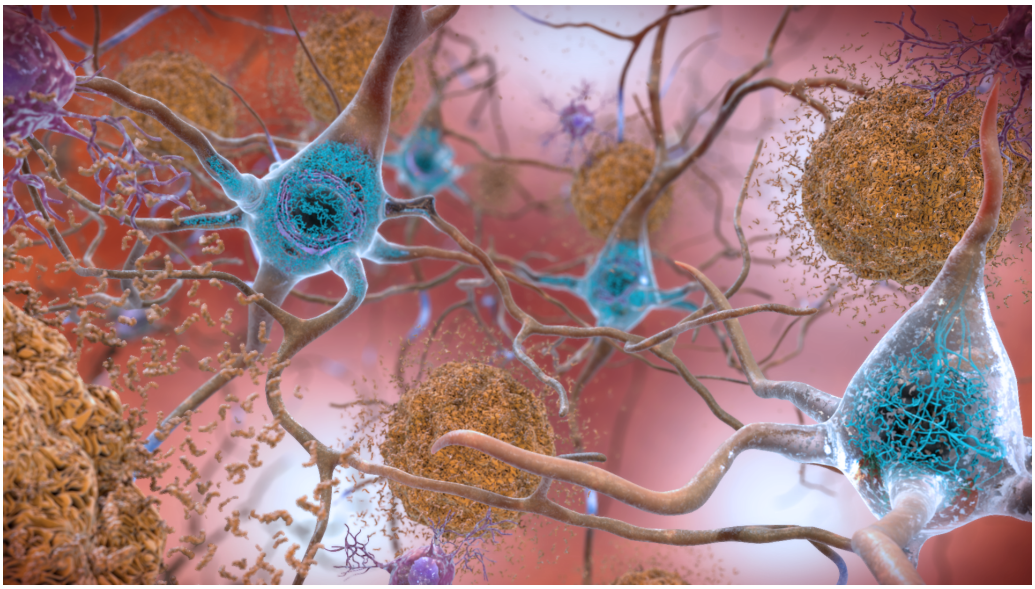


Figure 1.1 Illustration representing Aβ plaques inside the brain (shown as brown spherical aggregates between neurons). *Image courtesy of the National Institute on Aging/NIH.*

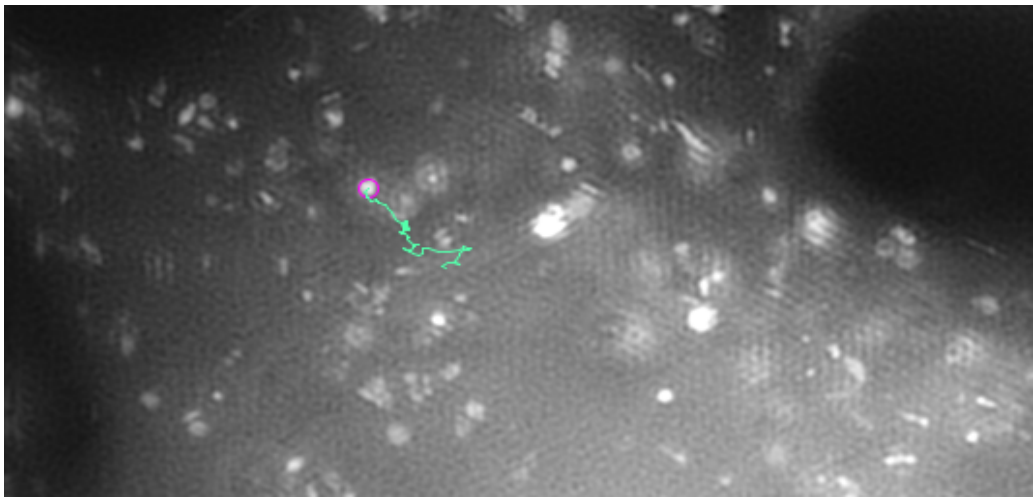


Figure 1.2 A frame taken from a fast SIM recording. Bright spots are amyloid beta aggregates moving inside a cell. The trajectory of one of them (circled in red) obtained by SPT is shown in green.

1.3 Super resolution microscopy and SPT

We will look at Alzheimer's disease from a very small scale, studying the motion of amyloid-β inside neural cells. Observing these kind of biological samples at molecular scale requires a very high resolution. In particular, one has to capture details beyond the light diffraction limit. The ensemble of techniques that allow to overcome this limit is known as *super resolution microscopy*. In the context of wide-field fluorescence imaging, two main strategies exist to achieve super resolution. The first and more traditional one is using patterned light illumination techniques such as *structured imaging microscopy* (SIM).[5] SIM relies on Moiré patterns created by illuminating the sample through different gratings. The patterns can then be analysed in reciprocal space to reconstruct the high resolution image. The second strategy is more recent and consists in localisation based methods such as *photo-activated localisation microscopy* (PALM)[3,17] and *stochastic optical reconstruction microscopy* (STORM).[22] Both PALM and STORM work by selectively activating subsets of fluorophores to avoid diffraction effects between closely spaced samples. This is obtained through *photobleaching* (the spontaneous fading of the fluorophore) in PALM and by *photoswitching* (the controlled switching between on/off states) in STORM.

All these techniques have undergone a remarkable development in recent years, making it possible to capture super resolution images at small time intervals (in the order of tens of milliseconds). These recordings allow to observe the motion of single molecules in cells at nanometer scale. In practice, one can reconstruct the molecules trajectories by linking their position observed in consecutive frames. Several tools and techniques have been designed to perform this operation,[23] that goes under the name of *single-particle tracking* (SPT). The advantage of SPT data with respect to other methods is that it measures individual molecule dynamics (and not just ensemble averages). Although the method is limited by the imaging time resolution, it allows to perform a much fine-grained analysis than other ensemble-based techniques.

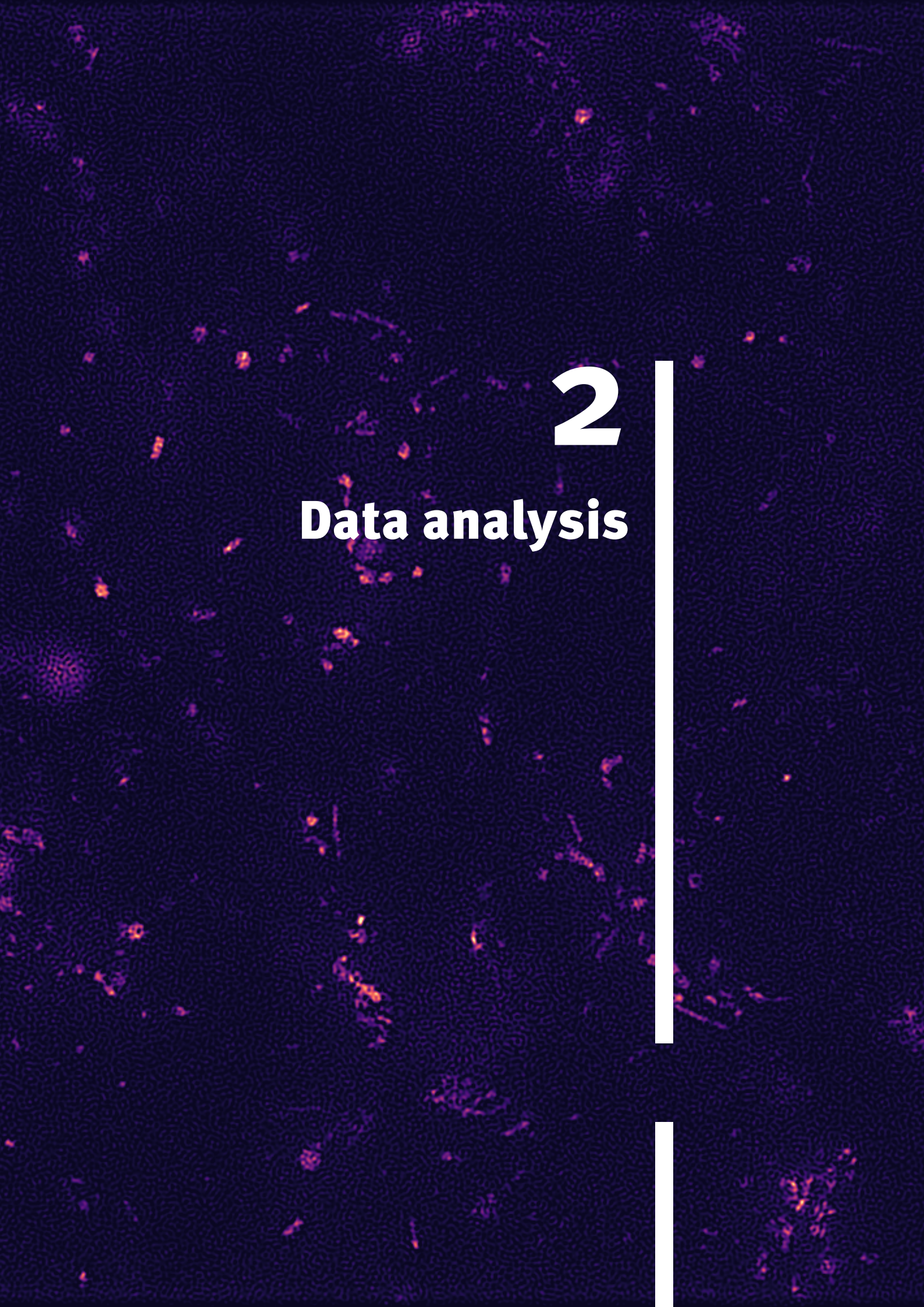
[5] Guerra. *Super resolution through illumination by diffraction-born evanescent waves* (1995)

[3] Betzig et al. *Imaging intracellular fluorescent proteins at nanometer resolution* (2006)

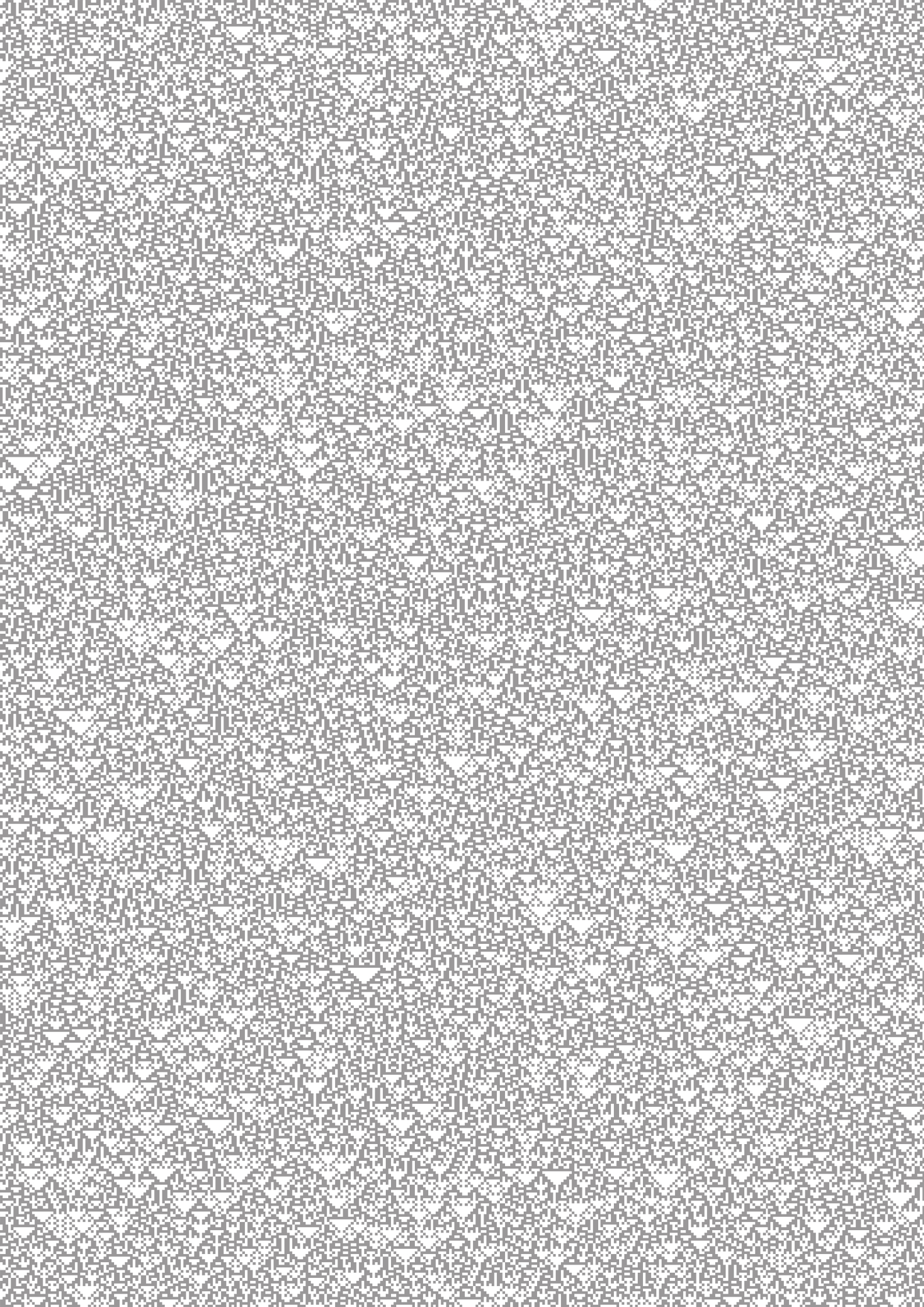
[17] Manley et al. *High-density mapping of single-molecule trajectories with photoactivated localization microscopy* (2008)

[22] Rust, Bates and Zhuang. *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)* (2006)

[23] Saxton. *Single-particle tracking: connecting the dots* (2008)



2
Data analysis



2.1 A statistical framework for tracking data

The modern imaging methods described in chapter 1 have made possible to track simultaneously multiple particles at high spatiotemporal resolution. Single-particle tracking can then be used to reconstruct thousands of particle trajectories. To extract meaningful insights from this huge amount of data, we need a modelling and statistical framework.

In this chapter we will present data analysis techniques to reconstruct the dynamics of amyloid-β aggregates. As explained in section 1.2, the uncontrolled formation of Aβ plaques in the brain tissue is thought to be a cause of Alzheimer's disease. What we would like to understand is how cells react to an excessive concentration of Aβ. How do they try to dispose of this dangerous products? We hypothesise that the cell transports and actively treats Aβ molecules for disposal in a well defined system. We tried to reconstruct the motion of Aβ molecules inside the cell to realise which route they follow and how they are treated. That said, the statistical methods described in the following aim to provide a general framework to extract dynamical parameters of particles at cell scale.

The data presented here was provided by the Molecular Neuroscience Group at the University of Cambridge, UK. It consists into trajectory fragments of Aβ-42 aggregates artificially injected in HEK 293T cells and captured via fast SIM (in 2D) at a frequency of 8 Hz. The trajectories were reconstructed from imaging data using the TrackMate software.[29]

[29] Tinevez et al. *TrackMate: An open and extensible platform for single-particle tracking* (2017)

2.2 Model and methods

At the microscopic level, the motion of the molecules can be described by Langevin's equation. In the case of biological processes we are interested in, the Langevin dynamics can be considered in its large friction limit (Smoluchowski's equation)[13,11]

$$\dot{x} = \frac{F(x)}{y} + \sqrt{2D}\dot{w} \tag{2.1}$$

where $F(x)$ is the drift force exerted on the particle at position x, y is the friction coefficient, D is the diffusion coefficient and $w(t)$ is a two-dimensional Wiener process. At this scale, it makes sense to consider the diffusion to be mainly due to thermal agitation so that it can be considered isotropic.

However, it is not possible to directly recover the microscopic model from the SPT data, since we miss information about the local behaviour both in space (such as the presence of microscopic obstacles undetected by the imaging device) and time (such as thermal fluctuations much faster than the acquisition timescale).[12,7] We can still build a coarse-grained model,[13,11] transforming eq. 2.1 into the effective stochastic equation

$$\dot{x} = a(x) + \sqrt{2}B\dot{w} \tag{2.2}$$

where $a(x)$ is the effective velocity field and $D \equiv B^T B$ is the effective diffusion tensor. It must be noted that, in principle, the effective diffusion coefficient may not be isotropic since it takes into account the local microscopic features (e.g. obstacles). On the other hand, actual analysis of the data show that anisotropic components

[13] Hozé et al. *Heterogeneity of receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging* (2012)

[11] Hozé and Holcman. *Residence times of receptors in dendritic spines analyzed by stochastic simulations in empirical domains* (2014)

[12] Hozé and Holcman. *Statistical methods for large ensembles of super resolution stochastic single particle trajectories in cell biology* (2017)

[7] Holcman, Hozé and Schuss. *Analysis and interpretation of superresolution single-particle trajectories* (2015)

[13] Hozé et al. *Heterogeneity of receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging* (2012)

[11] Hozé and Holcman. *Residence times of receptors in dendritic spines analyzed by stochastic simulations in empirical domains* (2014)

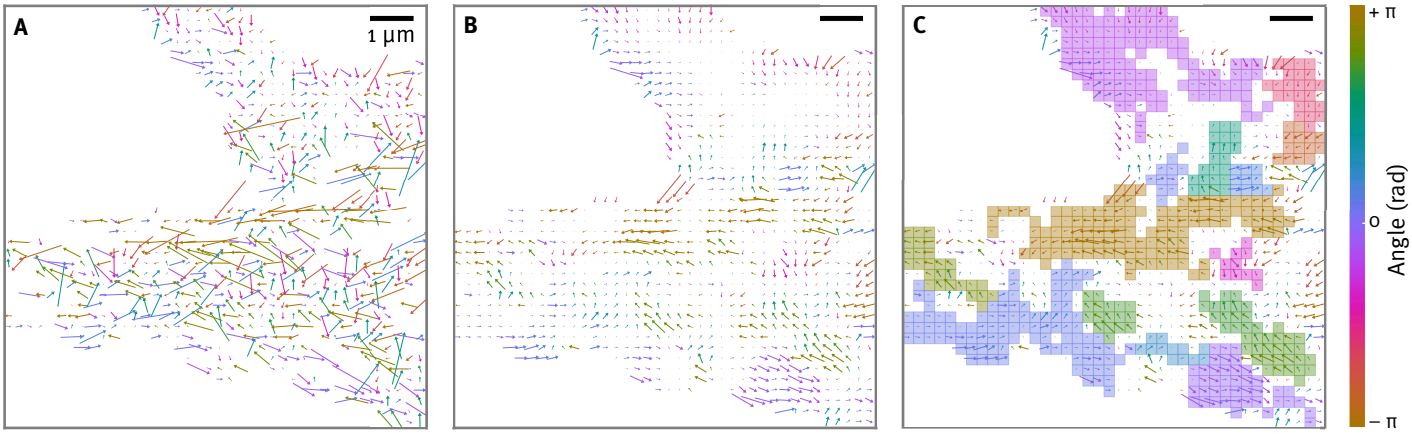


Figure 2.1 Velocity field obtained from the SPT data using bins of size 250 nm. **A** original velocity field, **B** velocity field after smoothening. **C** regions of coherent motion obtained by clustering the smoothened velocity field. The angular distance threshold was set to 0.08 and clusters smaller than 10 bins were discarded. The area color indicates the mean angle of the velocity among the cluster.

are negligible: in the following, the diffusion tensor is reduced for simplicity to a scalar coefficient by averaging on the diagonal entries. Moreover, the velocity field is assumed time invariant in the relatively short time window spanned by the SPT data (30–60 s).

Diffusion and velocity field estimation

To estimate the dynamical parameters of eq. 2.2 a statistical analysis is needed. We follow the approach demonstrated in [13, 11, 12, 7] by partitioning the data in a square grid with fixed bin size. The velocity field and diffusion coefficients are considered constant in each bin. If the acquisition time interval is sufficiently small, the process described by eq. 2.2 can be discretised following a forward Euler scheme:

$$x_{t+1} - x_t = a_{\Delta t}(x_t)\Delta t + \sqrt{2\Delta t D_{\Delta t}(x_t)}n_t \tag{2.3}$$

where Δt is the acquisition time interval.

Then, the trajectory step $\Delta x_t \equiv x_{t+1} - x_t$ starting in bin B is a normally distributed random variable with mean $a(x_B)$ and variance $2D(x_B)\Delta t$ where x_B is the center of the bin. The velocity field and diffusion tensor can thus be recovered by computing the empirical estimate of the moments:

$$a_{\Delta t}(x_B) = \frac{1}{\Delta t}\mathbb{E}_B\left[\Delta x\right] \tag{2.4}$$

$$D_{\Delta t}(x_B) = \frac{1}{2\Delta t}\left(\mathbb{E}_B\left[\Delta x^2\right] - \mathbb{E}_B[\Delta x]^2\right) \tag{2.5}$$

where the expected value is taken on the bin. Notice that the $a_{\Delta t}$ and $D_{\Delta t}$ are not the same of those of eq. 2.2, as they depend on the acquisition interval Δt. In fact, we may recover the coefficients of the continuous process only in the limit $\Delta t \to 0$. To avoid a too cluttered notation, we will drop the Δt index in the following and assume it implicitly.

Field smoothening

To perform analysis and simulations we often need a smooth representation of the velocity or diffusion field, providing estimates also for bins with few—or missing—data. To solve this problem, a convolution with variable kernel is applied on the field. Denoting by n_B the number of datapoints inside bin B, the smoothened field $\tilde{a}$ is obtained as

$$\tilde{a}(x_B) = (k * a)(x_B) \equiv \frac{\sum_{S \in \Gamma(B)} n_S a(x_S)}{\sum_{S' \in \Gamma(B)} n_{S'}} \tag{2.6}$$

where $\Gamma(B)$ is the Moore neighbourhood of B.

Explicitly, representing bins with their grid site indices, the kernel matrix $K(i,j)$ centered in the bin i,j is:

$$K(i,j) = \frac{1}{\sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} n_{k,l}} \begin{bmatrix} n_{i-1,j-1} & n_{i-1,j} & n_{i-1,j+1} \\ n_{i,j-1} & n_{i,j} & n_{i,j+1} \\ n_{i+1,j-1} & n_{i+1,j} & n_{i+1,j+1} \end{bmatrix} \tag{2.7}$$

Finally, bins such that the total number of samples in their Moore neighbourhood is lower than a threshold are excluded from the analysis. A comparison between the velocity field before and after the smoothening is visible in figs. 2.1a and 2.1b.

This approach can be interpreted in terms of Bayesian inference, where the prior (Gaussian) distribution of the parameter of interest is based on the distributions in the neighbouring bins. The mean estimator on the posterior is then obtained by averaging the means of the neighbours with a weight proportional to the number of samples.

2.3 Attractors

The first biophysical features that we want to identify are attracting regions. We would like to infer from the tracking data if there exist regions in the intracellular space where Aβ molecules are collected and, if so, we would like to characterise their dimension and attraction force. Note that in microscopy images we are only able to see the fluorescent tags on the Aβ molecules and not the cell structure. Attracting regions may represent organelles where Aβ is collected for disposal. Being able to identify these regions may allow us to relate them to specific organelles (such as lysosomes) by looking at their distribution and time evolution.

To identify local attractors we use the methodology and formalism developed in [13, 20]. Considering the velocity field to be locally conservative, it can be described by the gradient of a scalar potential:

$$a(x) = -\nabla U(x) \tag{2.8}$$

[13] Hozé et al. *Heterogeneity of receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging* (2012)

[20] Parutto and Holcman. *Detection and interpretation of high density nanometer regions in super resolution single particle trajectories* (2018)

Local attractors can then be seen as wells in the potential. At first non-zero order around a local minumum x_0 the potential is a paraboloid:

$$U(x,y) = U_0 + A\left[\frac{(x-x_0)^2}{r_x^2} + \frac{(y-y_0)^2}{r_y^2}\right] + O(x,y)^2 \tag{2.9}$$

in a coordinate system (x,y) where the axes are rotated by an angle ϕ to match those of the paraboloid. A potential well is thus by the set of parameters A (well depth), x_0 (centre), r_1, r_2 (axes of the ellipse obtained by cutting the potential well at height A), and ϕ (the angle of the ellipse major axis).

If the diffusion coefficient is assumed to be locally constant, the particle density is described by the Boltzmann distribution

$$\rho(x,y) \propto \exp\left(-\frac{U(x,y)}{D}\right) \tag{2.10}$$

Substituting eq. 2.9, the density around a local minimum is approximately Gaussian. Thus, after selecting a small high density region, principal component analysis (PCA) can be used to approximate the location of the attractor (x_0) and the ellipse parameters (ϕ, r_1, r_2) corresponding to the 95% confidence ellipse.

To find the remaining parameter A an iterative procedure is used. A grid centered in x_0 is built, and for each iteration k an ellipse $\mathcal{E}_k$ with increasingly longer axis is obtained by rescaling the original confidence ellipse. At each iteration, we calculate the A_k that minimize the mean squared error (MSE) with respect to the velocity field of all the bins inside $\mathcal{E}$:

$$\mathrm{MSE}_k = \sum_{x_i \in \mathcal{E}_k} \| -\nabla U(x_i) - a(x_i)\|^2 \tag{2.11}$$

The best fit for A is then the A_k corresponding to the iteration with minimal MSE. A parabolic error score S indicating how much the potential well resembles a paraboloid is defined as

$$S \equiv \frac{\mathrm{MSE}_k}{\sum_{x_i \in \mathcal{E}_k} \|a(x_i)\|^2} \tag{2.12}$$

where $S \in [0,1]$ and $S = 0$ indicates a perfect fit. The initial high density regions are localised by using the DBSCAN clustering algorithm.

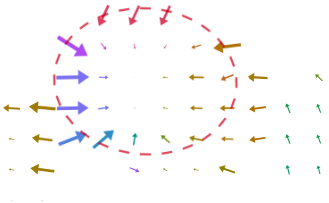


Figure 2.2 Example of a potential well in the velocity field obtained from the SPT data.

2.4 Regions of coherent motion

We now study regions in the cell space where amyloid beta particles follow a coherently directed motion. The study of this kind of regions is not new in the literature and can find relevant applications in many biological contexts. As an example, coherent motion regions have been observed in interphase chromatin[30], yielding new hypothesis about chromatin biological functions. In our case, we would like to find out if Aβ is transported along well defined pathways.

We identified regions of coherent motion by clustering adjacent bins based on the angular distance of their velocity vectors. The angular distance between two vectors u and v is defined as:

$$\mathrm{distance}(u,v) \equiv \frac{1}{\pi}\arccos\left(\frac{u\cdot v}{\|u\|\|v\|}\right) \tag{2.13}$$

[30] Zidovska, Weitz and Mitchison. *Micron-scale coherence in interphase chromatin dynamics* (2013)

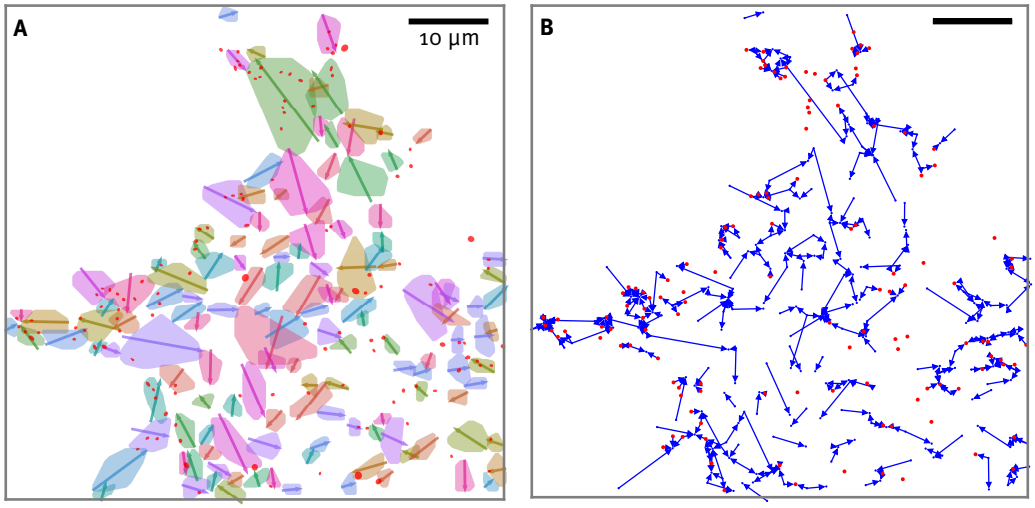


Figure 2.3 Reconstruction of the dynamical structure. **A** coherent motion regions with main axis highlighted and attractors (in red), **B** motion graph reconstructed by linking the regions of coherent motion and attractors.

We adopted a very simple algorithm to perform the clustering:

1. Create an empty cluster (i.e. a set of adjacent bins) for each bin in the grid and assign the bin to it.

2. For every bin B, consider each neighbouring bin $S \in \Gamma(B)$. If the angular distance between $\tilde{a}(x_B)$ and $\tilde{a}(x_S)$ is lower than a given threshold, merge the clusters of B and S.

This extremely simple method is quite effective in localising coherently directed motion. An example of the result is shown in fig. 2.1c.

2.5 The full picture

In the previous sections we localised regions of attraction and pathways. One can hypothesise that the dynamics of Aβ is composite. Molecules may escape (by Brownian motion) from an attractor, then travel along a pathway and eventually drop into another attracting region. We considered this approach by connecting the regions of coherent motion and the potential wells to form a directed graph representing a rough dynamical map of the intracellular space.

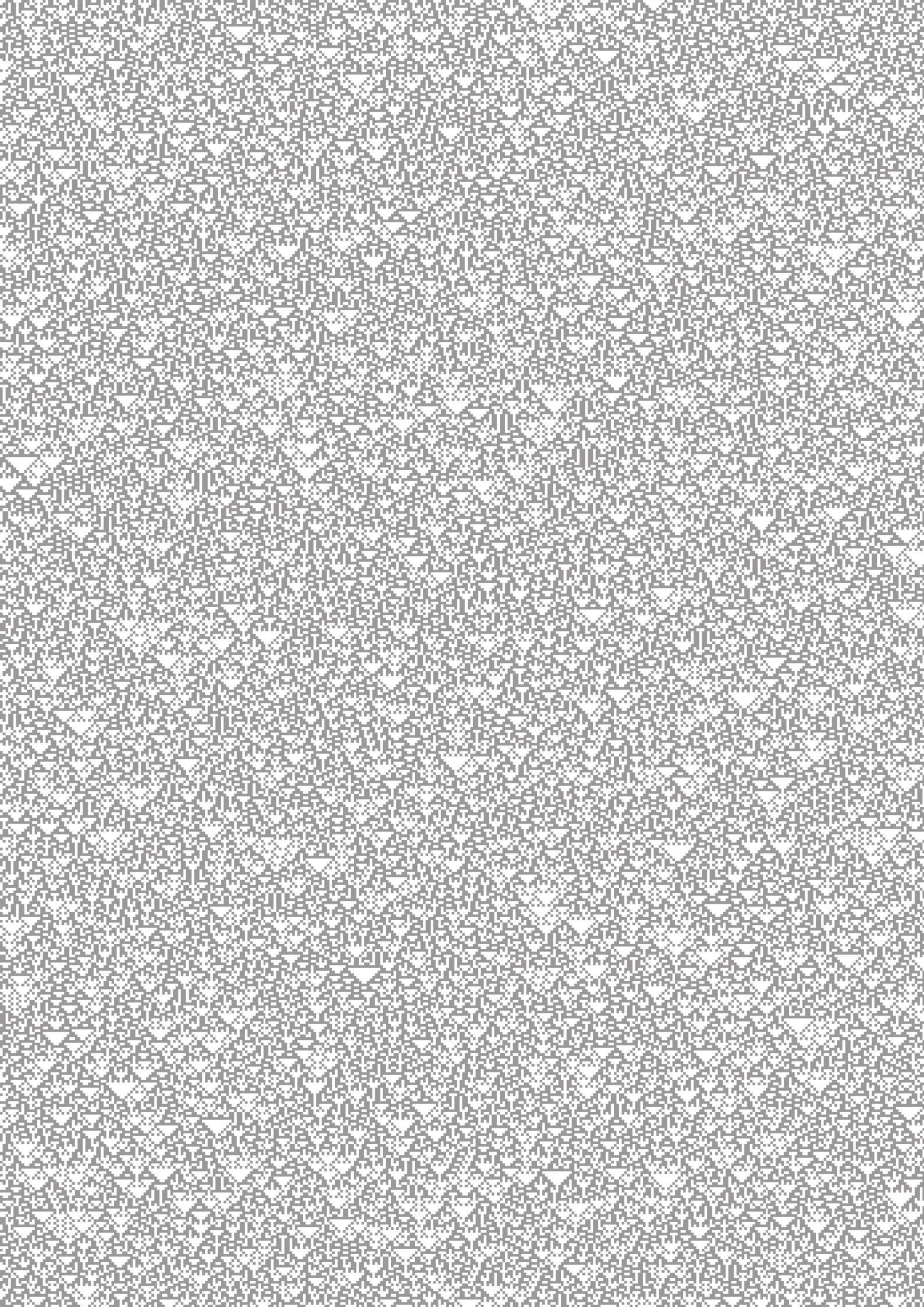
First, the axis corresponding to the average direction of each region is computed; then the endpoints of the regions are connected to other endpoints or potential wells that are found in a small neighbourhood. The reconstructed structure is shown in fig. 2.3.

Sadly, the results of this approach are a bit disappointing: the dynamical map that we were able to reconstruct by linking attractors and pathways is very complex and apparently meaningless. Although the methods to localise attractors and pathways described in sections 2.3 and 2.4 were successful in their scope, the hope of identifying a global dynamics in the data seems vain. The analysis shows that there is no well defined pathway along which the Aβ aggregates are driven nor clear accumulation points. Instead, the molecules follow complex itineraries, bouncing between different regions. The reconstruction of the dynamical structure shown in

fig. 2.3 gives an idea of the mess in the underlying motion. This leads us to formulate a new hypothesis. In fact, the behaviour may suggest that the processing of Aβ is highly delocalised, and molecules are bounced between several loosely distributed processing units. In the next chapter we will introduce a possible explanation for this behaviour, presenting a new model of motion and computing its characteristics.

3

Modelling the ER



3.1 The endoplasmic reticulum

The messy behaviour of Aβ molecules we saw in chapter 2 by the analysis of SPT data suggests a new hypothesis. The complex and delocalised dynamics that we found may involve active transport on the *endoplasmic reticulum*, a network like organelle supporting various functions encompassing protein folding and redistribution. While it is known that the ER acts as an active transportation network, it is not possible to find a directionality and the dynamics of such transport mechanism remains not clear.[18] In a recent work, Holcman et al.[9] have shown that particles moving on the ER follow a two-state process, characterised by alternation of a high-velocity directed motion state associated to flow in the network tubules and a low-velocity diffusing state associated to the nodes. This kind of behaviour, with particles jumping between areas of slow motion, resembles the results of the analysis performed in section 2.5, motivating the hypothesis that transport of Aβ is linked to the ER.

As the inner workings of the ER are not known, it would be difficult to directly compare the the results of chapter 2 to the ER dynamics. How are molecules redistributed in the ER? Which timescale characterise this transport process? First, we have to understand how this network works. Thus, in this chapter we make a step back and we focus on a model of motion on the ER, building on the results of Holcman et al., and presenting numerical simulations to understand the model characteristics.

[18] Nehls et al. *Dynamics and retention of misfolded proteins in native ER membranes* (2000)

[9] Holcman et al. *Single particle trajectories reveal active endoplasmic reticulum luminal flow* (2018)

3.2 Active network model

The ER is an organelle that spans from the nuclear envelope to the cell periphery. It has a fundamental role in the production, maturation and trafficking of proteins and lipids. A recent study based on super resolution imaging[19] has revealed hat its structure is made almost exclusively of tubules at varying densities. The tubules are usually connected by three-way junctions at roughly 120 degrees[28] (see also fig. 3.1). Since we only have 2 dimensional images we will refer to a planar representation of the ER network. This approximation is justified by the fact that in many kind of cells the ER is very thin with respect to its width. Given its regular three-way junctions, the simpler planar graph representing the ER is an hexagonal lattice.

Some characteristics of motion on the ER have been recently described through the analysis of SPT trajectories[9]. First, luminal proteins follow distinct transport mechanisms depending on topology: they move at fast velocity along the network tubules, and they instead move with dominant diffusive component while inside the junctions. Moreover, the luminal flow along tubules was observed to invert its direction at random time, possibly due to tubule constrictions. We introduce a model of motion that takes into account these effects.

We consider a transported molecule as a random walker on a directed graph, where edges invert their direction at a Poissonian rate (λ). Each edge will keep a directionality for an exponential time (timescale $\tau_{switch} = \frac{1}{\lambda}$), then reverse it, and so on and so forth. Moreover, the random walker spends an exponential time in each node (with timescale τ_{node}). This takes into account the time required to escape from a node, given the diffusive motion inside the junction.[26] After having waited in the node, the walker jumps randomly through one of the available outward directed

[19] Nixon-Abell et al. *Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER* (2016)

[28] Terasaki, Chen and Fujiwara. *Microtubules and the endoplasmic reticulum are highly interdependent structures.* (1986)

[9] Holcman et al. *Single particle trajectories reveal active endoplasmic reticulum luminal flow* (2018)

[26] Schuss, Singer and Holcman. *The narrow escape problem for diffusion in cellular microdomains* (2007)

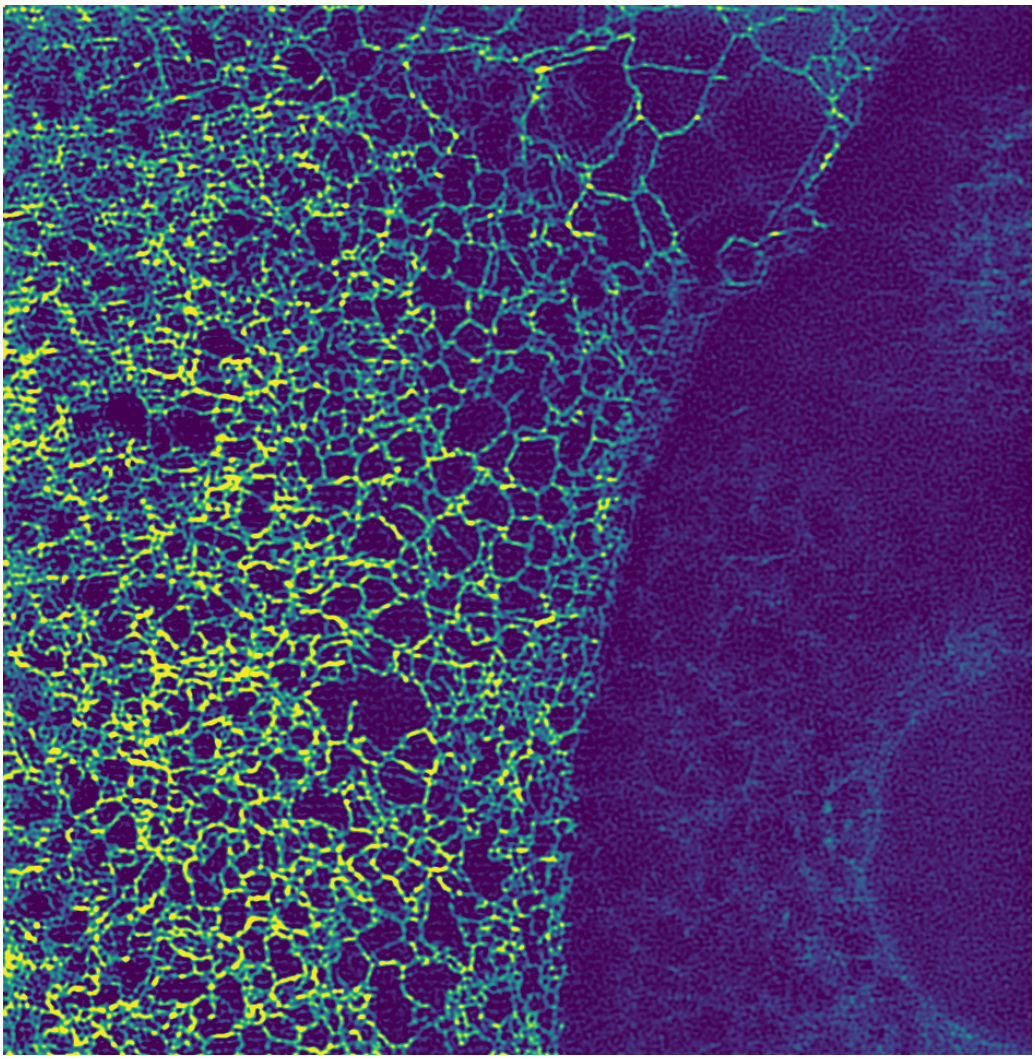


Figure 3.1 The endoplasmic reticulum of a HEK 293T cell imaged by fast SIM.

edges and continues its travel. If all the edges happen to be directed inwards, the particle cannot escape and has to wait another exponential time in the node. An example of the network model is shown in figure 3.2

3.3 Timescale of transport

We evaluate now the timescale of the transport on an active network. We consider two timescales that have important biological implications: the mean first passage time (MFPT) through a given node for a single particle and the average time required for the first particle of a group to reach a given node. This second case models an activation process where many particles are released from a source but one is sufficient to activate a receptor located in a far away node.

Mean first passage time

Considering a particle moving on graph starting from node S, the *first passage time* (or *hitting time*) for node T is the time at which the particle first hits the target T.

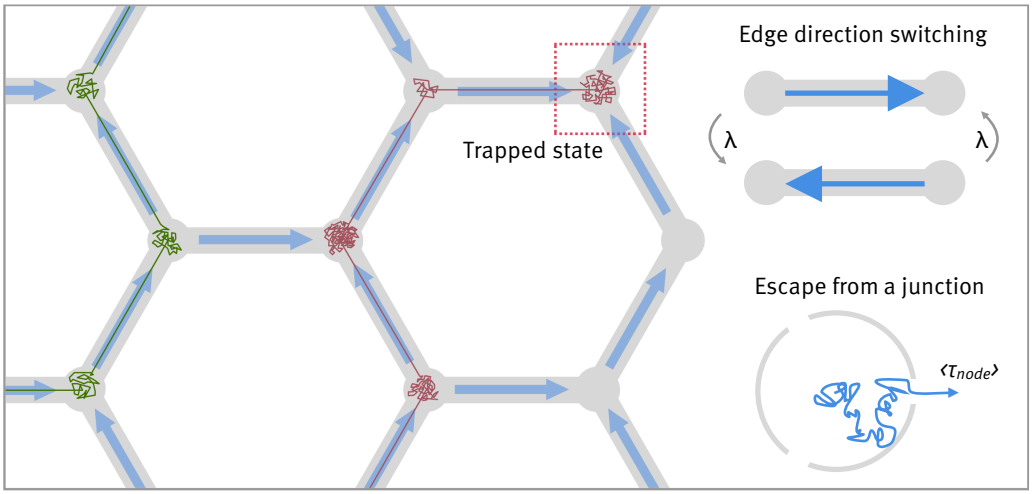


Figure 3.2 Active network model. Particles diffuse inside the junctions and quickly jump along the edges. The direction of the edges alternates with rate λ and the particles take an exponential time to escape from a junction by Brownian motion (timescale τ_{node}). A trapped state (no outward flux from a junction) is marked by the dashed red rectangle.

More formally,

$$\tau_{S\to T}(x) = \inf\{t : x(t) = T \mid x(0) = S\} \tag{3.1}$$

where $x(t)$ denotes the location of the particle at time t. It follows that, if $x(t)$ describes a stochastic motion, we can define the mean first passage time as

$$\bar{\tau}_{S\to T} = \mathbb{E}_x\left[\tau_{S\to T}(x) \mid x(0) = S\right] \tag{3.2}$$

where the expectation is taken over many realisations of the process.

Extreme statistics for passage time

We now consider the case of N particles moving simultaneously on the network, all starting at the same time from the same source node S. In this case, we are interested in knowing the average time required for the first among the N particles to hit the target T. Again, this is a common situation in biological systems, where a single particle may be sufficient to activate a receptor (for example, in a synapse). We define this as the minimum hitting time among N independent processes:

$$\tau^{ex}_{S\to T}(N) = \min\{\tau_{S\to T}(x_1), \tau_{S\to T}(x_2), \ldots, \tau_{S\to T}(x_N)\} \tag{3.3}$$

We can then take the average over many realisations of this multiple-particle process to define

$$\bar{\tau}^{ex}_{S\to T}(N) = \mathbb{E}_{\{x_1,x_2,\ldots,x_N\}}\left[\tau^{ex}_{S\to T}(N)\right] \tag{3.4}$$

We name this last quantity *extreme first passage time* (EFPT).

Numerical simulations and results

Numerical simulations of the model were performed to measure mean and extreme first passage time on both a synthetic hexagonal lattice and a reconstruction of a real network from SIM imaging data. The model parameters used are based on the timescales observed in the SPT data analysed by Holcman et al. and are summarised in table 3.1.

The results for the MFPT (figs. 3.3a and 3.3b) show a saturating behaviour for nodes at in the bulk of the network. Peripheral regions of the network that are

τ_{node}	100 ms
τ_{switch}	30 ms

Table 3.1 Values of model parameters.

weakly connected to the bulk are instead characterised by an exponentially increasing MFPT (figs. 3.3a and 3.3b). Introducing periodic boundary conditions removes this effect, as shown in fig. 3.3c. The saturation behaviour is reasonable: to visit a node that is far from the source, a particle has probably explored most of the network. The difference in MFPT between two nodes both located far from the source becomes negligible and the MFPT saturates to the *cover time* (the time required to visit all the nodes in the graph), independently of the node distance.

One interesting effect of the active network is shown in fig. 3.3d. One may think that the formation of traps in junctions (see fig. 3.2) would cause an increase of the MFPT. A trapped particle has to wait a time τ_{escape} until one of the edges switches, allowing to leave the node. Since the edges switch with rate λ,

$$\mathbb{P}\{\text{any of the 3 edges switches in } (t, t+dt)\} = 3\lambda dt \tag{3.5}$$

$$\langle \tau_{escape} \rangle \sim \int_0^\infty dt\, t\, 3\lambda e^{-3\lambda t} = \frac{1}{3\lambda} = \frac{\tau_{switch}}{3} \tag{3.6}$$

where we neglected the time for diffusion τ_{node}. We thus expect an increase in MFPT that is linear in τ_{switch}. The result of numerical simulations shows that this is indeed the case, but only for relatively large values τ_{switch}. In a range of relatively small τ_{switch} the slowdown due to traps is compensated by faster exploration of the network, making the timescales of the active graph equivalent to those of an undirected model. In fact, the directionality of edges force the particles to move forward, without immediately going back to the node they came from. The parameters range in which this phenomenon appears is compatible with the one found by Holcman et al. and it is thus plausible that the ER operates in this regime.

The value for the MFPT found in the simulations is surprising. On a graph with radius 40, a particle takes around 1000 s (on average) to arrive in a node that is just 10 steps away. This means a timescale of transport in the order of tens of minutes, or hours to reach the peripheral regions. While it is hard to say whether biological processes on the ER actually work at this slow timescale, we can speculate that since most of the cell functionings are faster, the MFPT is not a sufficient parameter to characterise them. When we instead consider the time to arrive for the first particle among a group of N (EFPT), the timescale drops to the order of seconds even for a group of 100 particles. With N = 1000, the time for the first to reach a target located 10 steps away is just slightly more than 1 s. This seems a much more reasonable timescale for an efficient biological process. We speculate that this kind of extreme statistics must have an important role in biology, as already noted in [21, 1, 2].

An asymptotic formula for the EFPT has been described, for the continuous space, in *Asymptotics of Elliptic and Parabolic PDEs*, p. 330.[8] The asymptotics in 2 dimensions reads:

$$\tau^{ex}_{S\to T}(N) \sim \frac{\delta^2}{D_{eff} \log N}, \quad N \gg 1 \tag{3.7}$$

where δ is the distance between S and T, D_{eff} is an effective diffusion coefficient, and N is the total number of particles. Using the graph distance $\delta = d(S,T)$, we can fit this behaviour to the results of the simulations on the active graph. As shown in fig. 3.4a, the simulations match very well the asymptotic behaviour in the bulk of the graph, confirming the validity of the analytical results.

[21] Reynaud et al. *Why so many sperm cells?* (2015)

[1] Basnayake et al. *Asymptotics of extreme statistics of escape time in 1,2 and 3-dimensional diffusions* (2017)

[2] Basnayake et al. *Extreme Narrow Escape: shortest paths for the first particles to escape through a small window* (2018)

[8] Holcman and Schuss. *Asymptotics of Elliptic and Parabolic PDEs* (2018)

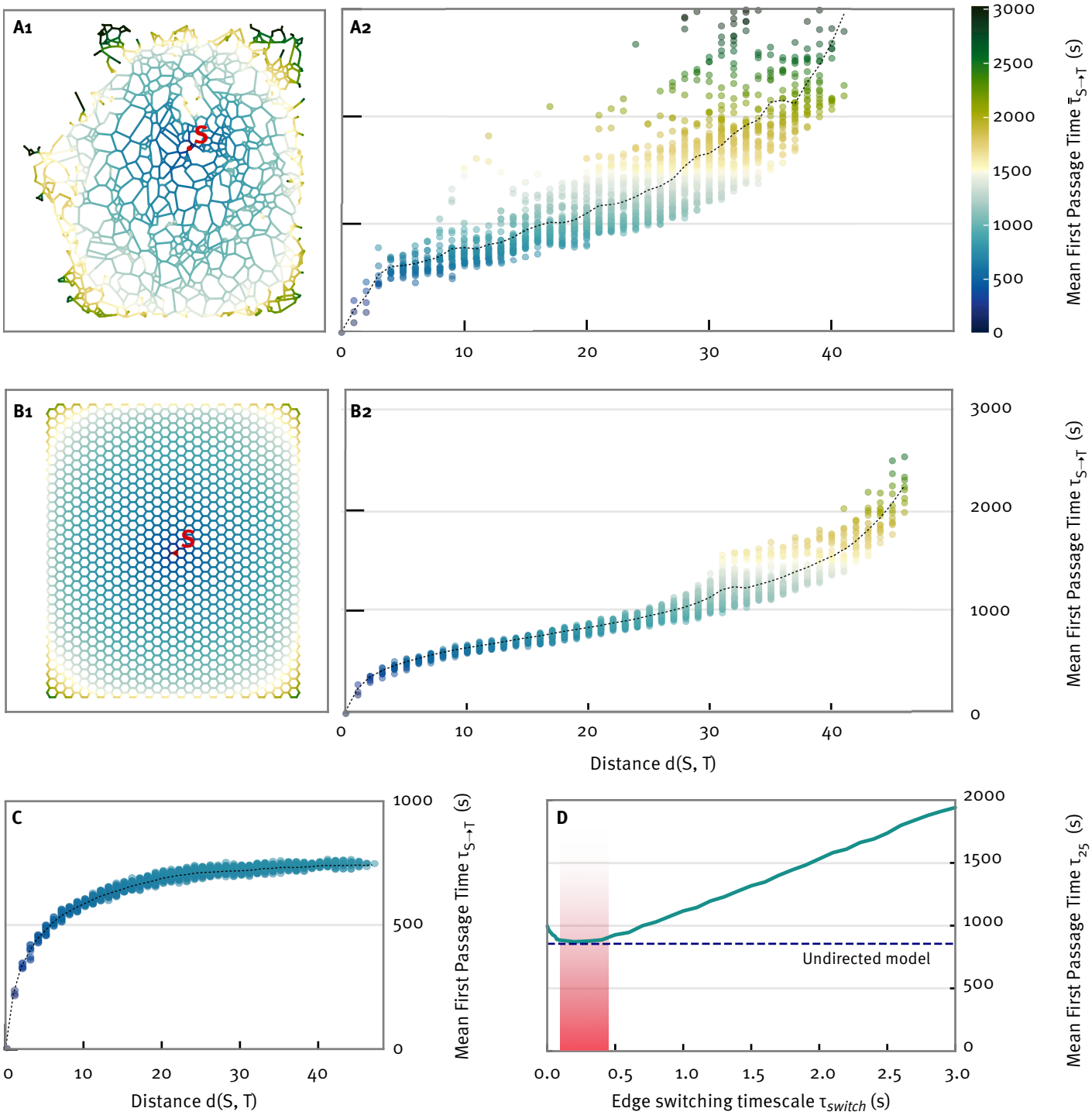


Figure 3.3 Timescales in the active network. **A** MFPT (heatmap and distance plot) on the network reconstructed from SIM imaging, **B** MFPT on an equivalent hexagonal lattice. Boundary effects are visible at the extremal regions that are weakly connected to the bulk of the network. **C** MFPT distance plot for the hexagonal lattice introducing periodic boundary conditions, showing the saturation behaviour. **D** effect on the MFPT of the switching timescale τ_{switch} for $d(S,T) = 25$ for the hexagonal lattice. The red region indicates the range of biologically plausible values of τ_{switch} taken from [9]. The dashed line represents motion on the same network with undirected edges. In the parameter range of interest, the slowdown effect due to the particles being trapped in junctions is compensated by the faster exploration rate due to edge directionality (particles cannot immediately go back to the node they come from), obtaining a MFPT equivalent to the undirected model.

Finally, the simulations show that fastest trajectories follow the optimal path on the graph. This may seem naive, but one has to consider that in principle the first particle to arrive may be the one that goes faster through a longer trajectory (for example, avoiding trapping nodes). In our case, that would mean a deviation from the exponential distribution of waiting times in nodes. Simulations results prove that, while this deviation exists when considering the fastest particles for large value of N ($\sim 10^4$), it is in general negligible with respect to the optimisation of the trajectory path. Indeed, fastest trajectories tend to follow (spatially) optimal paths (see fig. 3.4b). This confirms previous results obtained for Brownian motion in 2 dimensions.[2]

[2] Basnayake et al. *Extreme Narrow Escape: shortest paths for the first particles to escape through a small window* (2018)

3.4 Packet motion

In a standard random walk on a graph, the probability of finding the walker in a given node at equilibrium tends to an uniform distribution[15]. This means that if we release a group of particles from a source, given a sufficiently long time, they will distribute evenly on the network with a timescale proportional to the largest eigenvalue of the stochastic matrix representing the transition probabilities between nodes. What happens instead on the active network model? When considering a fast edge switching rate (τ_{switch} = 30–300 ms), the equilibrium behaviour is preserved, leading to a roughly constant number of particles per node (figs. 3.4c1–2). For a slow switching timescale instead ($\tau_{switch} = 3\,\mathrm{s}$), a novel mechanism of transport emerges. Particles group in lumps, synchronising their motion and arriving in nodes in *packets* (see fig. 3.4c3). These packets are not stable: they are continuously formed (and unpacked) by splitting and merging of different particles or smaller packets. This behaviour undermine the realisation of a even distribution, as the edge switching keeps the system out of equilibrium. This mechanism allows for a delivery in redundant groups, which seems a way to guarantee robustness to a biological transport process.

[15] Lovász. *Random walks on graphs: A survey* (1993)

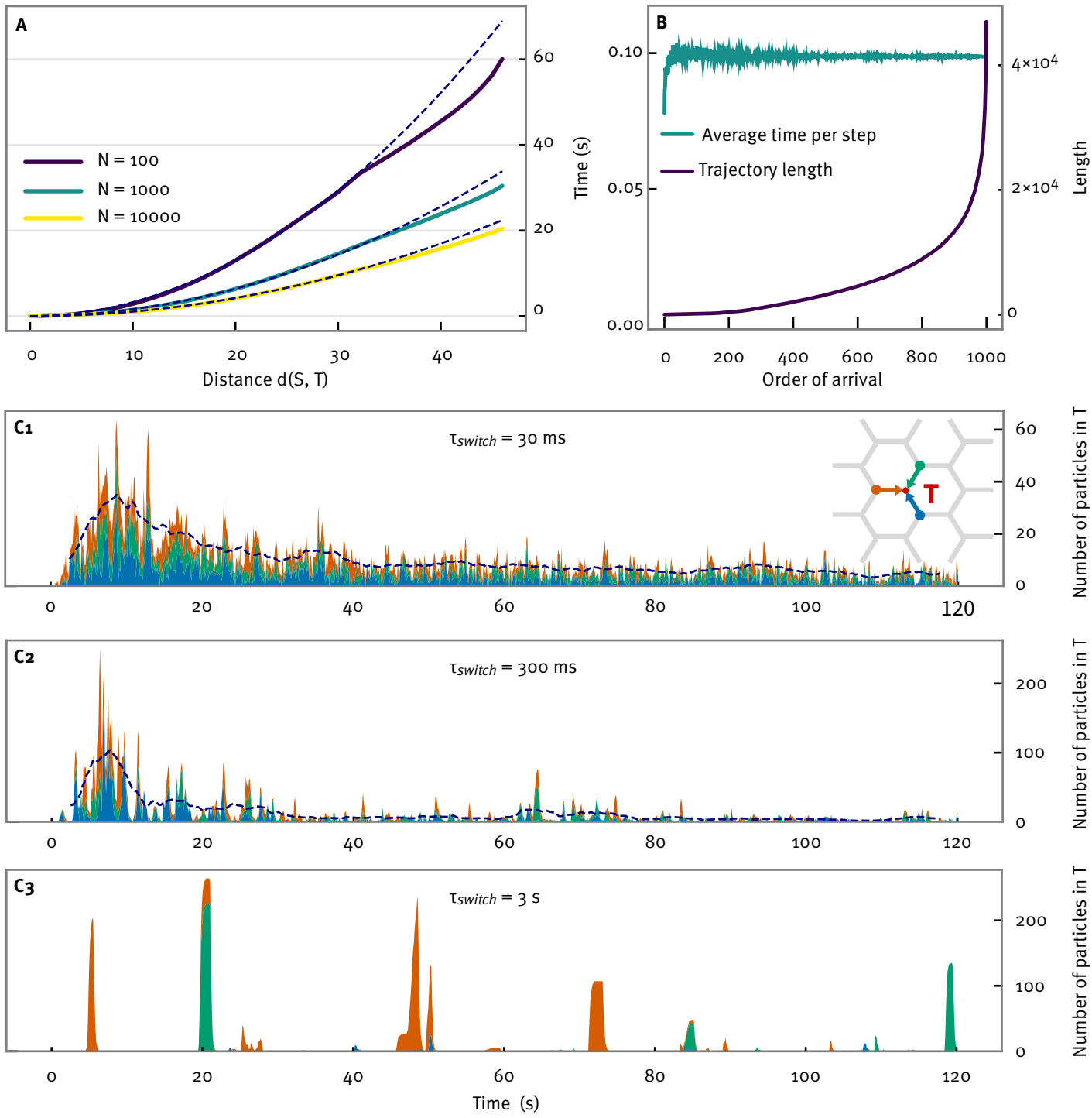
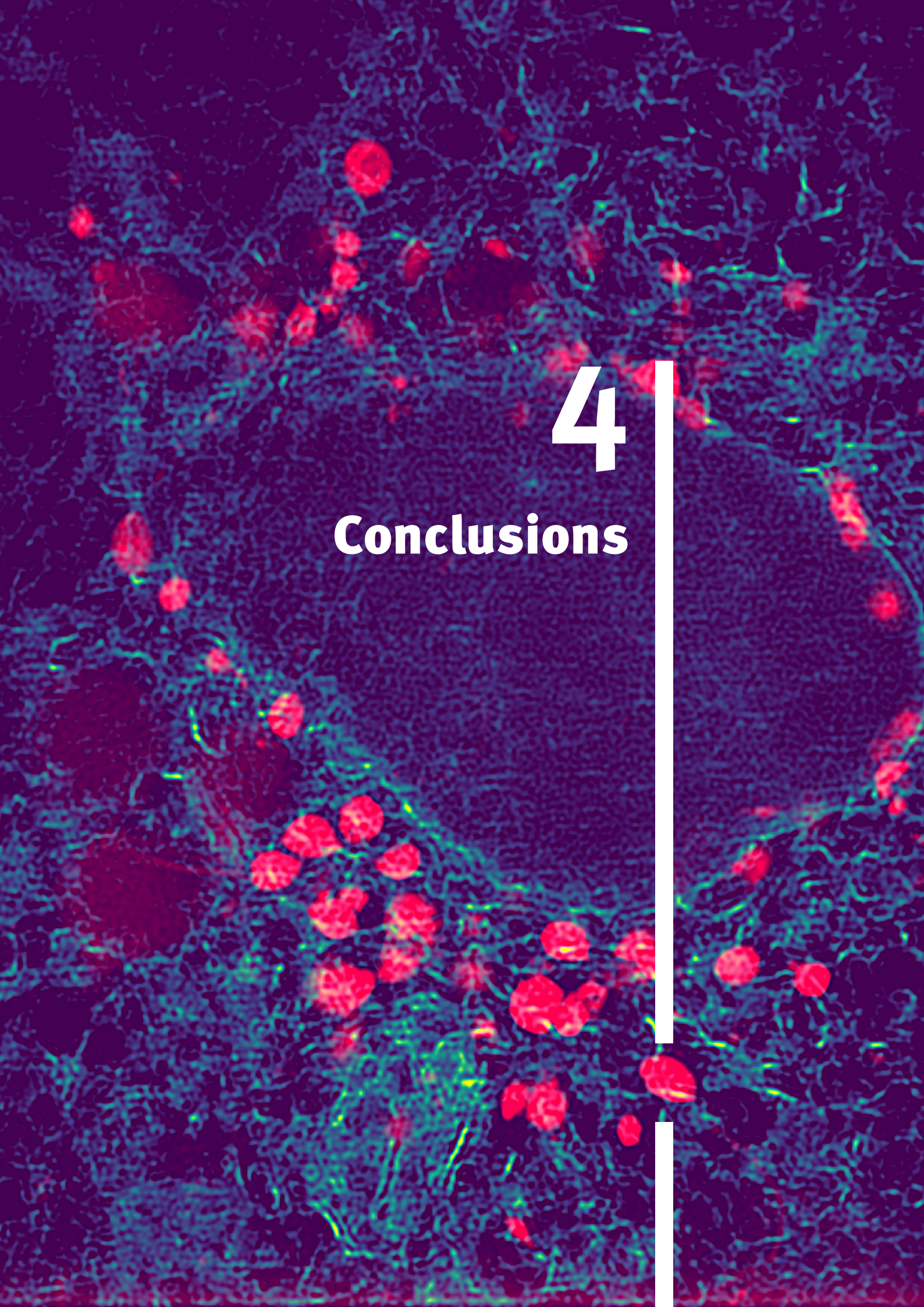
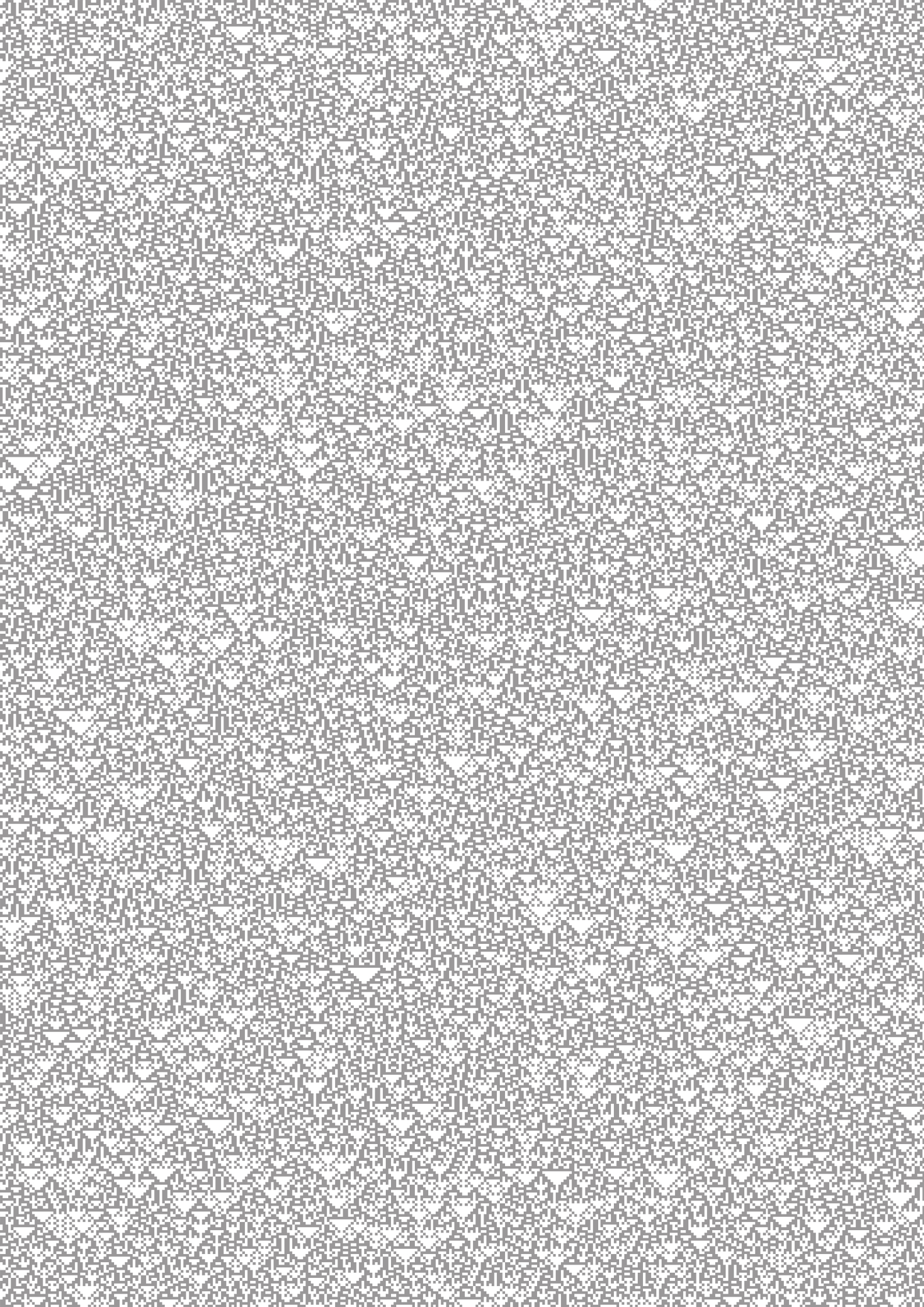


Figure 3.4 Extremes statistics and packet motion. **A** EFPT (time for the first to arrive(on the hexagonal network for different values of N (total number of particles). The dashed line shows the fit for $c_1 \frac{\delta^2}{c_2 + \log N}$ with $c_1 = 0.07$ and $c_2 = -2.39$. **B** characteristics of the trajectories by order of arrival (1000 out of 10^4 trajectories are considered). The plot shows how the trajectory length has a greater impact on the time to arrive than a deviation from the mean time per step. **C** number of particles in a target node T as a function of time. The area plot is coloured based on the edge particles are coming from. For small values of τ_{switch}, an uniform equilibrium distribution is reached after large enough time (figs. C1–2). Instead, for slow switching, a new kind of motion emerges. Particles arrive in the node in lumps (*packets*), at random times (fig. C3).



4
Conclusions



4.1 Perspective and open problems

By means of statistical methods and modelling we have tackled Alzheimer's disease from the molecular point of view, building a structured representation of the amyloid-β dynamics starting from single-particle tracking data. Unfortunately, the structure that this analysis has revealed is not of immediate nor simple interpretation. Regions of coherent motion and attractors form a complex framework, suggesting that amyloid-β aggregates are processed by a delocalised system. New results for the dynamics of the endoplasmic reticulum,[9] also based on SPT analysis, provided hints that this delocalised system may be represented by the ER. These findings were exploited to build a model of motion on the ER and, thanks to that, we have been able to provide an estimate for the timescale of the ER transport mechanism through numerical simulations. However, values for the mean first passage time between nodes of the network found in the simulations are unexpectedly high (tens of minutes), hinting at the fact that MFPT may not be the right parameter to consider when evaluating the redistribution timescale. Efficient transport in the context of biological processes should probably use other means. Speculating, we propose that one should look at extreme statistics to highlight the characteristics of this transport mechanism. Considering an activation process where multiple particles are released from a source but few (or even just one) are sufficient to enable receptors, we focused on the time required for the first particle to arrive in a target node and found a much shorter timescale (in the order of seconds), that seems compatible with the class of biological processes we want to describe. Hence, we compared our results with previous works concerning an equivalent Brownian process,[1,2,8] finding that the asymptotics for the first arrival time in continuous space also hold in the case of motion on a network, and providing new evidence supporting the idea that fastest trajectories follow the spatially optimal path. Our simulations of the active network model also revealed a novel transport mechanism which causes the particles to be delivered in redundant packets, supporting the idea that redundancy naturally emerges as a way to guarantee efficiency and robustness of biological processes, as it has already been pointed out in previous works.[21,25] Overall, we have explored a vast and hitherto little known domain, combining data analysis, modelling and numerical simulations, but most of our results remain open hypotheses: more experimental data is needed to verify the validity of the ER model we proposed and to highlight possible correlation between amyloid-β and ER dynamics; yet we offered a new direction and a new point of view to tackle these questions. Lastly, we hope that a better understanding of the ER transport mechanism will shed new light on amyloid plaques and their clinical implications.

[9] Holcman et al. *Single particle trajectories reveal active endoplasmic reticulum luminal flow* (2018)

[1] Basnayake et al. *Asymptotics of extreme statistics of escape time in 1,2 and 3-dimensional diffusions* (2017)

[2] Basnayake et al. *Extreme Narrow Escape: shortest paths for the first particles to escape through a small window* (2018)

[8] Holcman and Schuss. *Asymptotics of Elliptic and Parabolic PDEs* (2018)

[21] Reynaud et al. *Why so many sperm cells?* (2015)

[25] Schuss, Basnayake and Holcman. *Redundancy principle for optimal random search in biology* (2017)

4.2 Final remarks

The current work terminates in this hanging stage. We started from Alzheimer's and ended in modelling the endoplasmic reticulum, discovering on the way some of its peculiar characteristics and consolidating the existing statistical framework for the analysis of SPT data. Of course, the circle is not closed: we should use these few bits of novel knowledge to look back at the dynamics of amyloid-β. But this will be another long story.

Bibliography

[1] Kanishka Basnayake et al. *Asymptotics of extreme statistics of escape time in 1,2 and 3-dimensional diffusions*. In: *arXiv* 1711.01330 (2017). Preprint.

[2] Kanishka Basnayake et al. *Extreme Narrow Escape: shortest paths for the first particles to escape through a small window*. In: *arXiv* 1804.10808 (2018). Preprint.

[3] Eric Betzig et al. *Imaging intracellular fluorescent proteins at nanometer resolution*. In: *Science* 313.5793 (2006), pp. 1642–1645.

[4] D Eisenberg and M Jucker. *The Amyloid State of Proteins in Human Diseases*. In: *Cell* 148.6 (2012), pp. 1188–1203.

[5] John M Guerra. *Super resolution through illumination by diffraction-born evanescent waves*. In: *Applied physics letters* 66.26 (1995), pp. 3555–3557.

[6] John Hardy and Dennis J Selkoe. *The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics*. In: *science* 297.5580 (2002), pp. 353–356.

[7] David Holcman, Nathanael Hozé and Zeev Schuss. *Analysis and interpretation of superresolution single-particle trajectories*. In: *Biophys. J.* 109.11 (2015), pp. 1761–1771.

[8] David Holcman and Zeev Schuss. *Asymptotics of Elliptic and Parabolic PDEs*. New York: Springer, 2018.

[9] David Holcman et al. *Single particle trajectories reveal active endoplasmic reticulum luminal flow*. In: *Nature cell biology* (2018), p. 1.

[10] LS Honig et al. *EXPEDITION 3: a phase 3 trial of solanezumab in mild dementia due to Alzheimer's disease*. In: *CTAD http://www. ctad-alzheimer. com/live-expedition-3-webcast* (2016).

[11] Nathanael Hozé and David Holcman. *Residence times of receptors in dendritic spines analyzed by stochastic simulations in empirical domains*. In: *Biophys. J.* 107.12 (2014), pp. 3008–17.

[12] Nathanael Hozé and David Holcman. *Statistical methods for large ensembles of super resolution stochastic single particle trajectories in cell biology*. In: *Annual Review of Statistics and Its Application* 4.1 (2017), pp. 189–223.

[13] Nathanael Hozé et al. *Heterogeneity of receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging*. In: *PNAS* 109.42 (2012), pp. 17052–57.

[14] Lancet. *Alzheimer's disease: expedition into the unknown*. In: *Lancet (London, England)* 388.10061 (2016), p. 2713.

[15] László Lovász et al. *Random walks on graphs: A survey*. In: *Combinatorics, Paul erdos is eighty* 2.1 (1993), pp. 1–46.

[16] Simon Makin. *The amyloid hypothesis on trial*. In: *Nature* 559.7715 (2018), S4.

[17] S Manley et al. *High-density mapping of single-molecule trajectories with photoactivated localization microscopy*. In: *Nature Methods* 5.2 (2008), pp. 155–157.

[18] Sarah Nehls et al. *Dynamics and retention of misfolded proteins in native ER membranes*. In: *Nature cell biology* 2.5 (2000), p. 288.

[19] Jonathon Nixon-Abell et al. *Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER*. In: *Science* 354.6311 (2016).

[20] Pierre Parutto and David Holcman. *Detection and interpretation of high density nanometer regions in super resolution single particle trajectories*. Manuscript in preparation. 2018.

[21] K Reynaud et al. *Why so many sperm cells?* In: *Communicative & Integrative Biology* 8.3 (2015).

[22] Michael J Rust, Mark Bates and Xiaowei Zhuang. *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)*. In: *Nature methods* 3.10 (2006), p. 793.

[23] MJ Saxton. *Single-particle tracking: connecting the dots*. In: *Nature Methods* 5 (2008), pp. 671–672.

[24] Zeev Schuss. *Theory and applications of stochastic processes: an analytical approach*. New York: Springer, 2010.

[25] Zeev Schuss, Kanishka Basnayake and David Holcman. *Redundancy principle for optimal random search in biology*. In: *bioRxiv* (2017). Preprint.

[26] Zeev Schuss, A Singer and David Holcman. *The narrow escape problem for diffusion in cellular microdomains*. In: *PNAS* 104.41 (2007), pp. 16098–16103.

[27] Christopher A Taylor et al. *Deaths from Alzheimer's Disease—United States, 1999–2014*. In: *MMWR. Morbidity and mortality weekly report* 66.20 (2017), p. 521.

[28] Mark Terasaki, Lan Bo Chen and Keigi Fujiwara. *Microtubules and the endoplasmic reticulum are highly interdependent structures*. In: *The Journal of cell biology* 103.4 (1986), pp. 1557–1568.

[29] Jean-Yves Tinevez et al. *TrackMate: An open and extensible platform for single-particle tracking*. In: *Methods* 115 (2017), pp. 80–90.

[30] Alexandra Zidovska, David A. Weitz and Timothy J. Mitchison. *Micron-scale coherence in interphase chromatin dynamics*. In: *PNAS* (2013). ISSN: 0027-8424.