POLITECNICO DI TORINO

Faculty of Engineering
Master's Degree in Mathematical Engineering

MASTER'S THESIS

Understanding the mobility of T-cells "in vitro" esperiments

Supervisors:

Prof. Marcello Delitala
Prof. Federico Frascoli

Candidate:

LORENZO HUANG

Academic year 2023-2024

To my grandparents

Contents

1 Introduction **1**

2 Literature Review: Part 1 **5**
2.1 T-cell: A Biological Overview 5
2.2 Fluorescence microscopy 8
2.3 Convolutional Neural Network 9

3 Literature Review 2 **13**
3.1 Lévy flight: a mathematical introspective 13
3.2 Lévy Flight Hypothesis 14
3.3 The Lévy Flight Paradigm 16
3.4 Beyong the Lévy Flight Hypothesis: Foraging bumblebees . 18
3.5 Diluting Lévy Flight Paradigm in T cells motility 21
3.5.1 Anomalous diffusion: fractional Klein-Kramers model 21
3.5.2 The Speed-Persistence coupling model 23

4 Experimental Set-up **25**
4.1 Animals preparation 25
4.2 Cell culture 25
4.3 T-Cells in microwells 26
4.4 Microscopy of choice 26

5 Cells Tracking and Data Acquisition **29**
5.1 Video Labeling 29
5.2 DeepKymoTracker 30

6 Mathematical Concepts **37**
6.1 Cell - track 37
6.2 Speed 38
6.3 Angle 38
6.4 Random Walk 39

6.4.1 Generalization of Random Walk 44
6.5 Speed autocorrelation function 45

7 Numerical results **47**
7.1 Complications 47
7.2 Observations of the Video 47
7.3 Position 49
7.3.1 Special case: Anchored cell 51
7.4 Angles 52
7.5 Speed 53
7.5.1 Auto-correlation speed 57
7.6 Average life span of **Cell1** 58
7.7 Average-time mitosis 59
7.8 MSD - Mean Square Displacement 59

8 Modelling some behaviours seem in experiments **65**
8.1 Random Walk with real data 65
8.2 Random Walks - α models 68
8.2.1 Unrestricted model 70
8.2.2 Restricted models 71

9 Conclusion and future ideas **77**

Matlab and Tables **79**

Chapter 1

Introduction

T cells play a major role in our everyday lives in fighting pathogens and preventing anomalous cells from proliferating and becoming carcinogenic. If we understand T lymphocytes better mathematically or biologically, we will take a step closer to curing or even preventing illnesses like cancer. The main focus of this thesis is the analysis of real data on T cells' movement. Such living cells were placed into a micro-well and were recorded by a spinning disk confocal laser microscope. The raw images would then be processed with a program called DeepKymoTracker with the result of numerical data of interest. The results of the data analysis, which consists of examining data statistics and data from simple model simulations, clarify some key aspects of the type of motility the cells might have.

Chapter 2 contains a literature review across various scientific fields that contribute to the study of T-cells. The first part specifically addresses what T-cells are, their locations within the human body, and their major roles. Studying T-cells *in vivo* is challenging due to numerous constraints, such as observing the cells or tracking their movement without interfering with an organism's life. The aging of humans would also modify the internal dynamical and biological systems, complicating our complete understanding of biology. These and other factors have led researchers to take several steps back, particularly by analysing specimens *in vitro*. When a detailed specimen analysis is desired, the need to isolate and observe single cells arises. The second part of the revire is dedicated to a brief discussion of fluorescent microscopy, the technology that enables us to see living single cells through artificial lenses. A significant amount of biological information can be generated by capturing different videos of the cells in a suitable environment. However, for this thesis, an additional step is required: the extraction of numerical data for a more rigorous analysis. To achieve this, various approaches are possible. One such approach is Deep

Learning Analysis, which is one of the latest trends for handling large quantities of data and can produce results very efficiently. This makes it the third and final topic reviewed in the chapter.

Chapter 3 is the literature review of the mathematical biology field. The concept of Lévy flight and walks are introduced and the different paradigms are listed and implemented in experiments in various forager's dynamics on searching food. Cells's motility can be categorized as a particular case of such dynamics.

Chapter 4 summarises the materials used in the experiments and the origins of the cells and equipment. Additionally, it provides detailed information on how the cells are maintained and nourished.

Chapter 5 discusses the program developed by Khelina Fedorchuk, who recently completed her postdoctoral fellowship at Swinburne University of Technology. The program, named DeepKymoTracker, processes videos captured by the microscope to output tracked cells moving inside micro-wells and provides the desired numerical data for statistical analysis. Further details about its functionality are provided, including additional information on video processing and the T cells' labelling logic. The chapter ends with a brief examination of the program's limitations and missing features, identified during the DeepKymoTracker's testing phase, which was conducted thoroughly by myself.

Chapter 6 presents all the mathematical concepts implemented for the analysis of cells' motility, such as position, angle, speed, MSD, and autocorrelation function.

Chapter 7 shows the results found by implementing the mathematical concepts of Chapter 6 through modelling via the software Matlab ver. 2023a. Interesting results are found by analysing even the simplest statistics such as the position, which shows that T-cells prefer to split close to the wall.

Chapter 8 discusses some simple models. The first part recreates the walks with the real data in a two-dimensional lattice to provide their statistics even with the approximation of direction, while the second part explores possible models that may reproduce the subdiffusive behaviour of some real cell.

Chapter 9 summarises the work done in this experience with some ideas of how some steps throughout the Thesis can be optimized.

Chapter 2

Literature Review: Part 1

In this chapter, a brief introduction and overview of the three important paradigms (Cell Biology, Fluorescent Science and Data Science) are presented. At first glance, they seem to be disconnected from one another, but each of them is an essential piece of the puzzle that makes the study of lymphocytes possible, especially from a mathematical biology perspective.

2.1 T-cell: A Biological Overview

It is impossible to analyse the T-cells without knowing what they are, which is why this subsection contains the basics of T cells in a nutshell.
T cells, or T lymphocytes, are a subclass of white cells that have a major role in the adaptive immune response of the human body; whether there is an external pathogen such as virus-infected cells or an internal one such as cancer cells, T cells will try and neutralise them. The term "adaptive" indicates the capability of the system to get used to new pathogens and new types of antigens.

There are different types of T cells and the process of their maturation and differentiation is part of hematopoiesis, which is described as follows [40]. It all starts in the bone marrow where hematopoietic stem cells (HSC) migrate to the thymus where they mature into either myeloid or lymphoid cells [17]. Only the latter (also known as the common lymphoid progenitor, CLP) can differentiate into T, B or NK cells [17]. Such T cell progenitors express neither CD4 nor CD8 co-receptor and are commonly known as double-negative (DN) thymocytes ($CD4^{-}CD8^{-}$ TCR^{-}). To become proper T cells, DN thymocytes need to go through multiple stages of subdivisions and selections (positive and negative selections [42]) which

occur in different locations within the thymus and are mediated by specific types of chemokines receptors. Each mature T cell will then contain a unique T-cell receptor (TCR) enabling the immune system to recognize a large variety of different pathogens. The selected thymocytes for proliferation have the properties of low affinity to "self" antigens and useful TCR expression. In addition, only 2% of thymocytes pass the selection process and become immunocompetent single-positive T cells, which are $CD4^+$ ($CD4^+CD8^-TCR^+$) or $CD8^+$ ($CD4^-CD8^+TCR^+$). Such cells will be then released from the thymus into the peripheral tissues, waiting to be activated. A summary of the differentiation are illustrated in the figure (2.1).

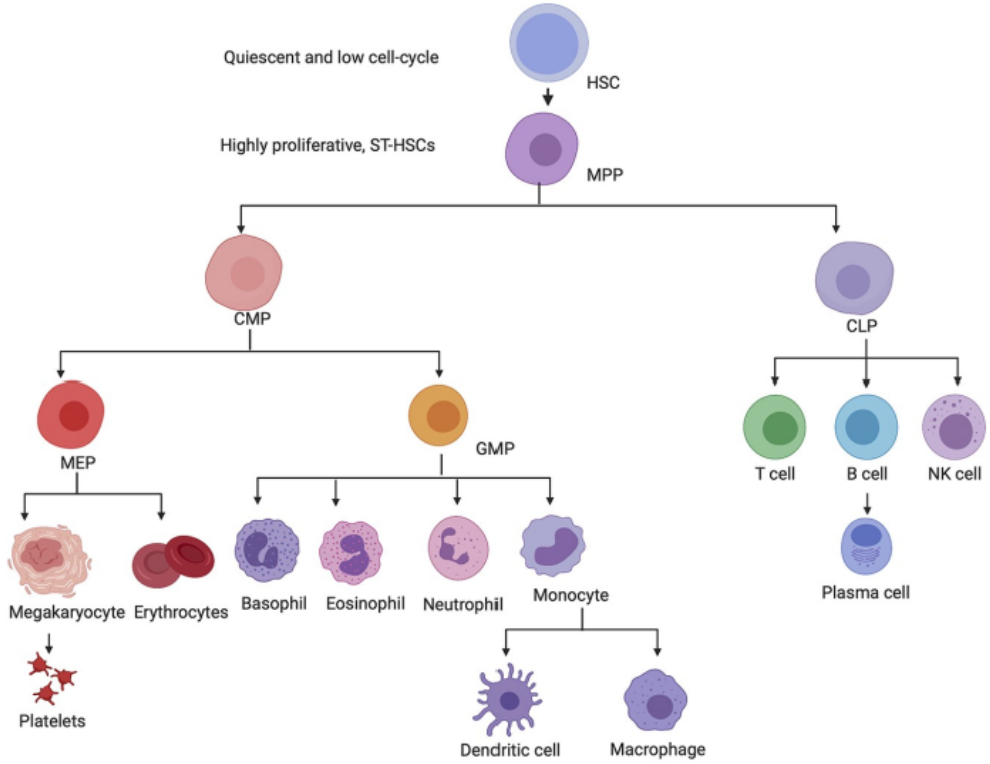

Figure 2.1: Schematic illustration of the hematopoiesis process (image from Ref. [29])

A further sub-classification can be considered based on T cells' function:

- **Helper $CD4^+$ T cell** or simply Helper T cells activate B cells and cytotoxic T cells. The activation of B cells induced the production of

antibodies and macrophages to eliminate microbes, while the activated cytotoxic T cells kill infected target cells [3];

- **Cytotoxic T lymphocytes** or simply $CD8^+$ T cells are specialized in lysing infected or cancerous cells [5];

- **Memory T cells** are usually produced after the neutralization of an acute infection. When an infection occurs, T cells are primed and the expansion phase begins when they acquire the capacity to eliminate the pathogen(s). After that, the contraction phase starts and only the 5-10% of cells survived and differentiated into different types of memory T cells [41];

- **regulatory $CD4^+$ T cells** are essential for the maintenance of self-tolerance and immune homeostasis, and the suppression of autoreactive T cells that evaded the negative selection [27];

- **Innate-like T cells** or unconventional T cells are the class of cells that have a fast immune response and recognise lipids, unlike the ones such as CD4 T helper cells or CD8 cytotoxic T cells which fit the conventional paradigm [1]. This category includes Natural killer T cells (NKT cells), Mucosal associated invariant T cells (MAIT) and Gamma delta T cells ($\gamma\delta$ T cells) [16];

By staying in the peripheral tissues, T cells would be quiescent for years, until they are activated by "professional" antigen-presenting cells, which are usually dendritic cells. This generally occurs in secondary lymphoid tissue such as the lymph nodes [38]. The cycle of T cells will be activated, triggered by signals from TCR, leading to the phase of expansion. Such phase consists of a rapid state of repetitive mitosis division and would dispose of specific pathogens. After successfully neutralising the pathogens, the contraction phase starts and the 95% of the T-cells (called effector cells) die, while the remaining ones become memory T cells, which become dormant and store information for a possible second exposure of the threat.

[1] "Conventional" here is used to indicate the behaviour of T cells distinguishing between self and non-self-entities due to the interaction of the TCR and the Major Histocompatibility Complex (MHC). Unconventional T cells bypass this step and elicit faster immune responses.

2.2 Fluorescence microscopy

Even with the biological knowledge of cells, it is yet not feasible to study their movements. Instruments have to be used to visualise and record with a proper frequency of shots. This subsection contains the basics of fluorescence microscopy, a fundamental instrument capable of investigating human microorganisms through its lens.

The goal of fluorescence microscopy is to provide data at the single-cell level. In the beginning, fluorescence labelling was based on antibody-conjugated fluorophones for immunohistochemistry [7]. A significant development happened in this field when Green Fluorescent Protein (GFP) was found three decades ago, leading to revolutionary non-invasive imaging techniques. GFP can be used to monitor gene expression and protein localisation in living organisms [6]. Different GFP-based variants have been studied and the implementation of small organic fluorescent dyes, nanocrystals, and autofluorescent proteins has deepened the understanding of T cells' movement [15].

The basics of any fluorescence-based technique involve illuminating the specimen with specific excitation wavelengths and analysing the light emitted from the specimen. Fluorescence microscopy is based on the so-called Stokes-shift principle [34], which enables spectral separation and distinguishes emitted fluorescence from the excitation light source. More advanced methods of standard fluorescence microscopy have been developed such as:

- **Confocal scanning laser microscopy**, which gives non-invasive measurements of fluorescent molecules within cells, with the added benefit of depth selectivity. Invented by M. Minsky in 1957, the concept was refined with the principle of Nipgow-Petran Disk and the photoelectric effect [8]: a laser beam is focused into a focal volume and scans a specimen to reconstruct a fluorescence image, the fluorescence accumulated in the focal volume is confocal with a pinhole and passes through the pinhole aperture, then the photons from the emission strike a detector (like a photomultiplier tube), and produce photoelectrons. The location of the specific area is set and synchronised by a computer that also translates the signals into a 2D grey image. Sequential optical slices are necessary to reconstruct a 3D fluorescence of the specimen.
- **Differential Interference Contrast Microscopy**, which uses refractive index gradients to enhance the contrast of imaged specimens.

Also known as Nomarski microscopy, it transmits polarised light through a prism (Nomarski-modified Wollaston prism) that splits the light into two orthogonally polarized rays. After a phase shift and constructive interferences, the two rays converge with enchance bright tones and dark shades [26].

- **SP5 Multispectral Confocal Microscopy**, or in particular, the Leica TCS SP5 was utilized by Dr. Kim Pham to scan multiple sections in high resolution and to record protein localization in motile and dividing T cells. It operates by exciting a piezoelectric crystal with a signal composed of multiple specific frequencies, allowing for the simultaneous diffraction of several wavelengths [37].

- **Spinning disk confocal laser microscopy**, which utilises a rotating perforated disk to scan a specimen simultaneously with thousands of tiny points of light, giving a more time-efficient variant for the confocal microscopy. The spinning disk confocal laser microscope is based on the Nipkow disk and photoelectron effect, with two high-speed spinning disks that interact with focused laser beams and excite fluorescent labels in the specimen. High signal-to-noise images can be obtained at a fast rate of 1000 frames per second [2].

In the experiment that will be analysed further in the discussion of this thesis, a spinning disk confocal laser microscope was used to image T cells during its movement (and division).

2.3 Convolutional Neural Network

After the acquisition of images, there is a necessity to process and extract numerical data, depending on the nature of the experiment. In particular, two tasks have to be attained: segmentation and tracking of each singular cell. The former represents the task of finding the pixels that characterised a cell in a given picture, while the latter requires the proper labelling of the cells in the logical time-sequential of their history of life. Numerous are Software that can manage and give the desired results such as ImageJ [1], Imaris [14] or TACTICS [40] which are all based on the standard module for Segmentation and Tracking. Another approach can be used to track cells and its theory is based on Convolutional Neural Networks. This is the work of Khelina Fedorchuk (PhD, [13]) based on the software TACTICS.

Since the subject itself is quite specific, it is worth spending some lines on the branch it belongs to Machine Learning.

According to Ref. [28], **Machine Learning** can be defined as computational methods using experience to improve performance or to make accurate predictions. It consists of designing efficient and accurate prediction algorithms and has a very broad area of practical applications like document classification, language processing or computer vision. The main idea orbits around finding the best suitable parameters for a function (given by an algorithm) by minimizing a so-called loss function. The main steps are explained as follows[28].

- **Sorting of data**, which starts with a collection of labelled data available and randomly partitions the data into three subsets called training, validation and test sample;
- **Feature selection**, which associates relevant features to the data (also called examples);
- **Training phase**, which uses the selected features to train the learning algorithm $\mathcal{A}$ by tuning the parameters Θ with the training sample;
- **Validation phase**, which selects the best parameters Θ_0 by using the validation sample;
- **Testing phase**, which evaluates the performance of the algorithm by using the test sample in the loss function;

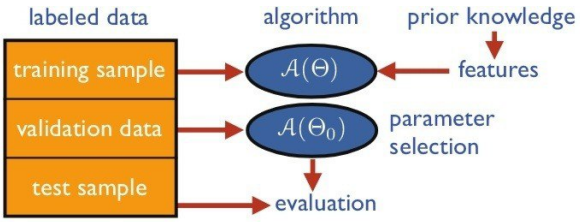

Figure 2.2: Main passages of the learning stages from [28]

Different learning scenarios depend on the type of available data, the order or the methods they are received. The main scenarios are as follows [28]:

- **Supervised learning** utilises labelled data to train algorithms where labelled data and the outcome are already known;

- **Unsupervised learning** utilises unlabelled data to train algorithms where the outcome is not clear;
- **Semi-supervised learning** utilises both labelled and unlabelled data to train algorithms where the outcome is not clear;
- **Online learning** involves multiple rounds where the training and testing phases are intermixed;
- **Reinforcement learning** also has the training and testing phases intermixed. To collect data, an interaction with the environment is necessary and an immediate reward is received. However, this is a short-term system, while in long-term no reward is provided by the environment so "the exploration versus exploitation dilemma [3]" might emerge;

A branch of machine learning is the **deep neural network** [13]. The term "deep" refers to the multiple hidden layers present to learn very elaborate features hierarchically. A deep neural network utilizes the first data as input at the lowest layer and uses them to learn simple features in the first hidden layers, then proceeds to transfer such information to the next layer, leading to a more challenging feature of data based on the information of the previous layer and so on. The layers consist of artificial neurons. Although deep neural network, or more generally deep learning has a great advantage of extracting features without human intervention due to the involvement of "black boxes" [4], it is challenging to train [18] and sometimes hard to interpret too.

A more specific deep neural network is the **convolutional neural networks** or CNNs. The primary distinction between CNNs and conventional feed-forward deep neural networks is their use of learnable filters, which were previously hand-engineered in traditional computer vision before the advent of deep learning. The structure of CNNs, like that of ordinary neural networks, is inspired by biological processes, specifically the architecture of the animal visual cortex. It was discovered that individual neurons in the visual cortex respond only to activation in specific regions of the visual field, known as receptive fields. In CNNs, convolutional learnable filters and the resulting feature maps function as artificial counterparts to the visual cortex: small convolutional filters slide across the input image, covering it region by region, and the computed numerical values (convolutions) for each region are stored as a feature map. To

[3] There is the choice between exploring more and collecting more data or exploiting the information already collected.

summarise, the benefit of using convolutional neural networks is the possibility of acquiring features of image data without human intervention employing learning filters. Furthermore, convolutional neural networks have three essential aspects: local connectivity, parameter sharing and translational invariance.

- **Local connectivity** refers to the possibility that two neurons of two different layers may not be connected;
- **Parameter sharing** refers to the fact that all neurons within a feature map use the same set of weights, which are learned by the filter associated with that feature map. Each neuron in the current feature map is connected to the corresponding neuron in the previous feature map using this identical small set of weights;
- **Translational invariance** means that the features of objects detected by filters in an image will be recognised regardless of the objects' positions within the image.

The combination of the first two can greatly reduce the number of parameters in neural networks, while the last one is essential for the object movements or similarities in objects. In the architecture of a CNN, there is also the inclusion of activation layers and pooling layers which optimise the input of neurons, reduce the size of feature maps, speed up the computational analysis and make possible features more robust.

In the program that Khelina developed, she utilised both 2D and 3D max pooling activations, which are used to reduce the size of feature maps. In particular, for the segmentation, a 2D trained convolutional neural network is used, while on the other hand, for the tracking, a series of 3D convolutional neural networks is implemented.

Chapter 3

Literature Review 2

The study of cells' movement may also be interpreted as a category of efficient research for resources, which is why this chapter is dedicated to the literature review of foragers' efficient research for resource modelling. The main source of this part came from Ref. [23]. Then, it follows some review of other papers attempting to understand cells' motility based on experimental observations (Ref [21]. It is relevant to specify that the state of understanding is still not fully resolved with multiple papers produced yearly (Ref. [39]). Furthermore, the hypotheses themselves are very challenging to verify, and identifying "the best" hypothesis among them is another open matter. Such arduousness comes from a lack of data, mathematical difficulties and other complex factors.

3.1 Lévy flight: a mathematical introspective

Movement Ecology is the discipline studying the foraging strategies of biological organisms. The central question experts in this field are trying to answer is: how does one search efficiently for resources without knowing the position beforehand? In 1906, Karl Pearson (Ref. [33]) proposed to model the movements of biological organisms by a simple random walk. Such a statement implies that the movement lengths of the individual are distributed exponentially according to a Gaussian distribution. However, the distribution failed to explain the frequent long movement lengths that occur when resources are not found close by. As a consequence, another distribution has to be considered. A suitable candidate is a class called **Lévy (α-stable) distributions** ρ_α. Let l_n be the n-th movement length then

$$\rho_\alpha \sim |l_n|^{-1-\alpha} \quad |l_n \gg 1|, \quad 0 < \alpha < 2 \tag{3.1}$$

Such distributions are good candidates since the distributions' tail decay slower than the exponential one, replicating the biological phenomenon that the resources are detected sparsely. A list of properties of Lévy the distribution follows:

- Lévy distribution is in the class of the heavy-tailed distribution;
- the moments of a Lévy distribution do not exist.

$$\left\langle l_n^i \right\rangle = \int_{-\infty}^{+\infty} l_n^i \rho(l_n) dl_n = \infty \tag{3.2}$$

  where $\left\langle l_n^i \right\rangle$ is the i-th moment of the distribution. Such results imply that the standard central limit theorem does not apply. Gnedenko and Kolmogorov proved mathematically that Lévy distributions obey a generalized central limit theorem (Ref. [36]);

- as a consequence of the previous point, these distributions are stable, which means that a linear combination of two random variables sampled independently from the same distribution reproduces the very same distribution, up to some scale factors;
- Lévy stable distribution is scale-invariant, which means the distributions remain unchanged under a rescaling;

In the limit-case that the resources are available everywhere, the Gaussian distribution is a better choice than Lévy, since the forager has no reason to waste energy for long-distance travelling.

3.2 Lévy Flight Hypothesis

Historically, the first group to pioneer into such matter were Viswanathan et al. In their article published in 1996 (Ref. [43]), they traced and studied a flock of albatrosses. They found that the plot of the number of flight time intervals of certain lengths of 5 albatrosses during 19 foraging bouts could be described as a power law of t with exponent $\mu = 2$. The hypothesis, called **Lévy Flight Hypothesis** (LFH), was then formulated. It states that Lévy motion provides an optimal search strategy for sparse, randomly distributed, immobile, re-visitable targets in unbounded domains. Since the moments of a Lévy distribution are infinite, the concept of the **Lévy walk** was introduced to overcome the problem of validating the theory through experimental data. The differences between them are as follows:

- The **Lévy walk formulation** provides that the random walk jumps are drawn randomly from the Lévy distribution, but as a penalty for long jumps, the walker spends a time t_n proportional to the length of the jump to complete it

$$t_n = v l_n \tag{3.3}$$

where $|v|$ is chosen constant and it defines the speed of the Lévy walker. In contrast to the Brownian motion where the mean square displacement $\langle x^2 \rangle$ grows linearly in time, for Lévy walks it grows faster than linear

$$\left\langle x^2 \right\rangle \sim t^{\gamma} \text{ as } t \to +\infty \tag{3.4}$$

where $\gamma > 1$. Generally, when $\gamma \neq 1$, the phenomenon is classified as anomalous diffusion (Ref. [36]). The entire spectrum of anomalous diffusion is observed in a range of natural processes and various aspects of human activities;

- The **Lévy flight formulation** provides that a walker makes a jump of length $|l_n|$ during an integer time sped of duration $\Delta n = 1$. Thus, a Lévy flyer can jump instantaneously over arbitrarily long distances with arbitrarily large velocities, without a penalty;

In experimental literature, the distinction between Lévy walks and Lévy flights is often overlooked when analysing step length frequency distributions. A step length x per unit time is defined as a jump length l_n, so fitting truncated power law $\sim x^{-\mu}$ to these distributions corresponds to using a cut-down form of the jump length distribution with exponent $\mu = 1 + \alpha$, which tests for truncated Lévy flights. This truncation solves the issue of infinite moments in random walks from Lévy flights. However, this approach does not test the Lévy Flight Hypothesis (LFH), which is derived from Lévy walks. Furthermore, the validation of the Lévy walks hypothesis requires a completely different data analysis method, which falls beyond the scope of the literature review, since the thesis will be focusing neither on the Lévy flight nor the Lévy walks.

There are two different types of LFHs:

- **Lévy Search Hypothesis** (LSH) is formulated to obtain a Lévy flight as the optimal search dynamics. It has the following characteristics:

  1. The type forager chosen is a cruise forager, which senses targets whenever it is moving;

2. for a cruise forager, a jump is terminated when it hits a target (this defines a truncated Lévy walk) ;
3. the specification of whether a forager eliminates the resource or not determine if the search is a destructive or non-destructive type;
4. defines the density of the sources;
5. defines if the sources are moving or immobile;

Furthermore, the optimal has to be specified mathematically, for instance, it can minimise the searching time for finding the resources.

- **Lévy environment Hypothesis** (LEH), which suggests that Lévy flight emerge from the interaction between a forager and a source. Such source is considered to have a distribution;

The major difference between the LSH and LEH is that the former, unlike the latter, states that under certain conditions, a forager performs Lévy flights regardless of the food source distribution. In the section 3.3, the concept of the Lévy Flight Paradigm (LFP) will be introduced and it will unify the LSH and the LEH by focusing on power laws in foraging dynamics without considering the specific validity conditions of these hypotheses. Unlike the LSH and LEH, the LFP often lacks clarity regarding the mathematical, physical, and biological origins of the power laws it identifies. However, it offers a broader perspective by encouraging the examination of power laws in animal movement data, beyond just testing for exponential distributions. It widened the scope by emphasizing that one should also check for power laws in animal movement data.

3.3 The Lévy Flight Paradigm

The ideas that Viswanathan's paper (Ref. [43]) started never stopped in the field of ecology, even though the results of the power lawin the original paper were later found to be incoherent by Edwards (Ref. [10]), due to an incorrect interpretation of the time acquired. With a proper correction of the data and the implementation of a truncated power law, the Lévy flights' behaviour disappeared, or to be more specific, only a couple of albatrosses seem to preserve a Levy Flight behaviour. With a more profound study, for instance, the work of Levernier et al. (Ref. [25]), the Lévy walks with power law exponent $\mu = 2$ is optimal in one dimension. In any higher dimension, the optimization by Lévy walks was at best marginal, in

the sense that the existence of any maximum in the search efficiency sensible depends on the choice of parameters controlling the foraging motion. Regardless of this result, the rigorous test of the LFH is very challenging, especially mathematically, which is why a loosen *modus operandi* (i.e. ignoring the mathematical assumption) has been considered. The concept is called **Lévy flight Paradigm**, consisting of looking for power laws in the probability distributions of step lengths of foraging animals ignoring any conditions of validity of LSH or LEH.

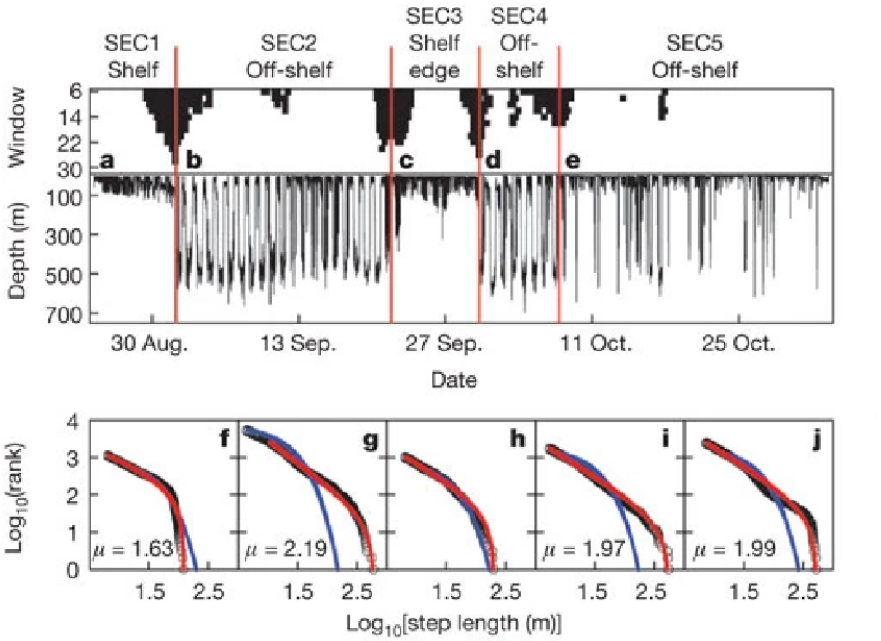

Figure 3.1: Top: Data of the diving depths of a blue shark acquired throughout half a year. The read lines split the data into different sections (**a**-**e**), where the shark dives deep or the diving depth is more constrained. These sections match the shark being off-shelf or on the shelf, respectively. Bottom: double-logarithmic plots of the move sped length frequency as a function of the step length, which is the vertical distance moved by the shark per unit time, with the notation (**f**-**j**) corresponding to the primary data shown in section (**a**-**e**). Clack circles correspond to data, red lines to fits with truncated power laws of exponent μ and blue lines to exponential fits. Figure from Ref. [19]

An instance of such application is found in the work of Humphries et al. [19]) on the diving depths of free-ranging marine predators. It was discovered that in the case of the blue shark, the distributions of the depth length are very well approximated by a combination of exponential distribution and a power-law distribution. When the predators were close to

shallow water, due to a strong dependence on the environment, an exponential distribution was found to be suitable to fit the data. On the other hand, when the shark was in the open ocean, its movement lengths were very well approximated by a truncated power law. Thus, this suggests that the best candidate to describe the data can be a superposition of the two different distributions. Such a combination of 2 types of motion is called **bimodal motion**; it consists of two distinct phases of motion mixed randomly. Fig. 3.1 represents graphically all the avobe considerations.

3.4 Beyong the Lévy Flight Hypothesis: Foraging bumblebees

The LFH from Section 3.2 illustrates the problem of identifying a suitable mathematical model for biologically relevant search strategies. Subsequently, such a hypothesis was then softened to generally look for power laws in the probability distribution of step lengths for foraging animals and it was formulated as the LFP (Section 3.3). By **diluting the LFP hypothesis** even more, one can find an even loosen formulation: foraging dynamics of biological organisms can be understood by analysing probability distributions. An instance of such *modus operandi* was explored by Thomas Ings and Lars Chittka in their report on the laboratory experiment regarding bumblebee dynamics with a predation threat (Ref. [20], [24]). The experiment was set up as follows: bumblebees flew in a cubic arena of $\approx$75 cm side length by foraging on a 4$\times$4vertical grid of artificial yellow flowers on one wall. On the landing platform of each flower, nectar was given to the bumblebees and replenished after consumption.

Artificial spiders were introduced to analyse the possible different foraging behaviours of the bumblebees under the threat of predation. The 3D flight trajectories of 30 bumblebees, tested sequentially and individually, were tracked by two high-frame-rate cameras. Several scenarios were staged during the experiment, but only three of them are relevant to our analysis:

1. spider-free foraging;
2. foraging under predation risk;
3. a memory test one day later;

In scenario one, the bumblebees harvest nectar from artificial flowers freely. In preparation for the other two scenarios, spider models were

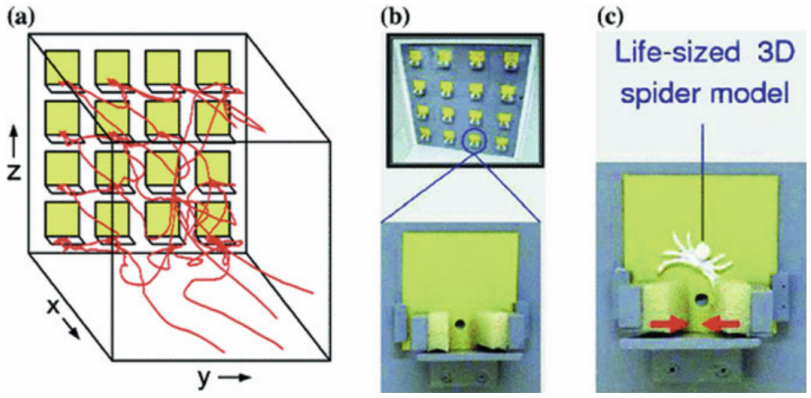

Figure 3.2: bumblebees experiment set-up. On the left (a) is an image of possible trajectories the bumblebees may trace in the confined space. The yellow squares are artificial flowers to collect nectar. (b) is the close-up to show the real artificial flower and (c) is the picture of the plastic spiders to simulate a predation threat. Picture from [23]

placed randomly among the artificial flowers. The plastic threats would hold the bumblebee momentarily to simulate a predation attempt. Scenario two is the recording of bumblebees where the spider models were left inactive in the arena, these are the same bumblebees that have already experienced the threats when activated spider models were placed in the same flowers. Scenario three is the copycat of scenario two one day later, testing the possible long influence of previous experience with predation risk on the bumblebees' flight dynamics. This experiment can test neither the LSH nor the LEH, as the flight arena is too small: the bumblebees always sense the walls and may adjust their flight behaviour accordingly. As discussed earlier, the less stringent guiding principle derived from the LFP further suggests that the key to understanding foraging dynamics may lie primarily in the probability distributions of flight step lengths. The best fit seems to be a bimodal motion when analysing the single velocity distribution in each stage: one distribution fits the data well near the flowers while another distribution is more suitable when bumblebees are far from the food source. However, the comparison of the velocity distribution of the 3 different stages of single bumblebees seems to be the same, qualitatively and quantitatively. To properly analyse the changes in bumblebees' flights due to changes in environmental conditions, another statistic was implemented, called the velocity autocorrelation function, as one may infer from our diluted LFP principle. The definition of this statistic is given in Eq. 6.28. By plotting the autocorrelation function for the velocity along

the y-axis of the three scenarios (as one may see in Fig. 3.3), there is a significant memory[1] decay of the function into negative values for the blue and green dotted lines, representing scenarios two and three. Such description matches coherently with the biological behaviour of the bumblebees. According to the data, whenever they go close to a flower they have previously experienced a capture, they try to avoid it even after a day. In contrast, the red dotted line, which represents the statistic of scenario one, exhibits a more gradual memory decay before going into negative values. Thi is an expected result since the bumblebees should be careless in scenario one where no predation risks were present. To reproduce theoretically such changes, the model chosen was a **Lanvegin equation**, also known as Newton's Law of stochastic physics. The equation is as follows:

$$\frac{dv_y(t)}{dt} = -\eta v_y(t) - \frac{dU(y(t))}{dy} - \xi(t) \tag{3.5}$$

where the term in the LHS of Eq. 3.5 is the acceleration along y, while in the RHS of Eq. 3.5 there are the friction, a potential and a stochastic term, respectively. The friction term considers the damping of the surrounding fluid, the potential term mimics the interaction between bumblebees and spiders, and the stochastic term includes any random behaviour of the bumblebees. The results can be seen in the Fig 3.3. The green dotted line is the simulation of the Langevin equation when the potential term is considered, while the red dotted one only has the friction and stochastic term. The predictions given by the autocorrelation function fit the biological phenomena: the animal tends to fly back when it knows a possible threat is nearby.

These results demonstrate that velocity correlations can provide essential insights into foraging and search dynamics, particularly through the emergence of complex correlation decay driven by interactions between the forager and predators. While this experiment could not test the Lévy Stable Hypothesis (LSH) due to unmet mathematical assumptions, the findings conceptually align with the Lévy Environment Hypothesis (LEH). The interaction between the forager and its environment was theoretically modelled by a repulsive force activated in the presence of predators, qualitatively reproducing the experimental outcomes. Along with the spatially intermittent dynamics observed when approaching food

[1] When a system is memoryless as, in the case of random walk, the expected decay is instantaneous, meaning that the value of the autocorrelation function on lag 1 is close to 0, while for another complex system the memory has different types of decays, where the value at lag 1 can range from -1 to 1.

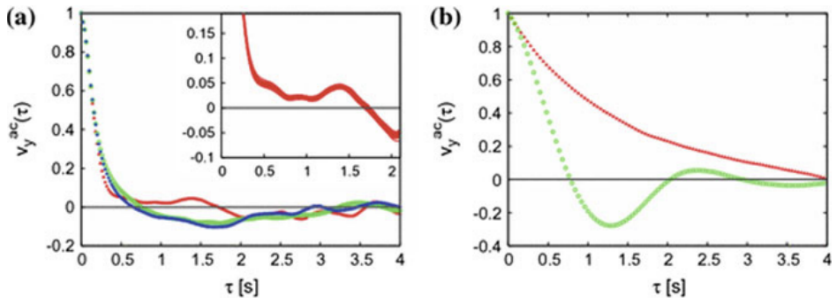

Figure 3.3: Plot of the autocorrelation functions: on the left, the plot of the real data, the red dotted line represents scenario 1, the blue dotted line represents scenario 2 and the green dotted line represents scenario 3. On the right, the plot of the simulated Langevin equation, the green dotted line is the results of Eq. 3.5, while the red one is the simulation of Eq. 3.5 without the potential term. Figure from [23]

sources, these results highlight a complex spatio-temporal adaptation of bumblebees to food availability and predator presence, contrasting sharply with the scale-free dynamics predicted by the Lévy Flight Hypothesis (LFH).

3.5 Diluting Lévy Flight Paradigm in T cells motility

In this section, a brief review of further significant papers is given, to show that the search for proper probability distributions to suit foraging dynamics is in the same spirit of the study of T-cell motility.

3.5.1 Anomalous diffusion: fractional Klein-Kramers model

In the work of Dieterich et al. (Ref. [9]), authors studied the cell movement of mutated epithelial cells (transformed Madin-Darby canine kidney), in particular 2 types: wild-type (NHE^+) and NHE-deficient (NHE^-) cells. A comparison between two models has been implemented: the Brownian motion and the fractional Klein-Krames model. The FKK model satisfied the fractional Flein-Kramer equation without external forces:

$$\frac{\partial P}{\partial t} = -\frac{\partial \left[vP\right]}{\partial x} + \frac{\partial^{1-\alpha}}{\partial t^{1-\alpha}}\gamma_\alpha \left[\frac{\partial \left[vP\right]}{\partial v} + \frac{K_B T}{M}\frac{\partial^2 P}{\partial x^2}\right] \tag{3.6}$$

where P is the probability distribution in position x, velocity v and time t. γ_α denotes the damping term, k_B is the Boltzmann constant, T is the

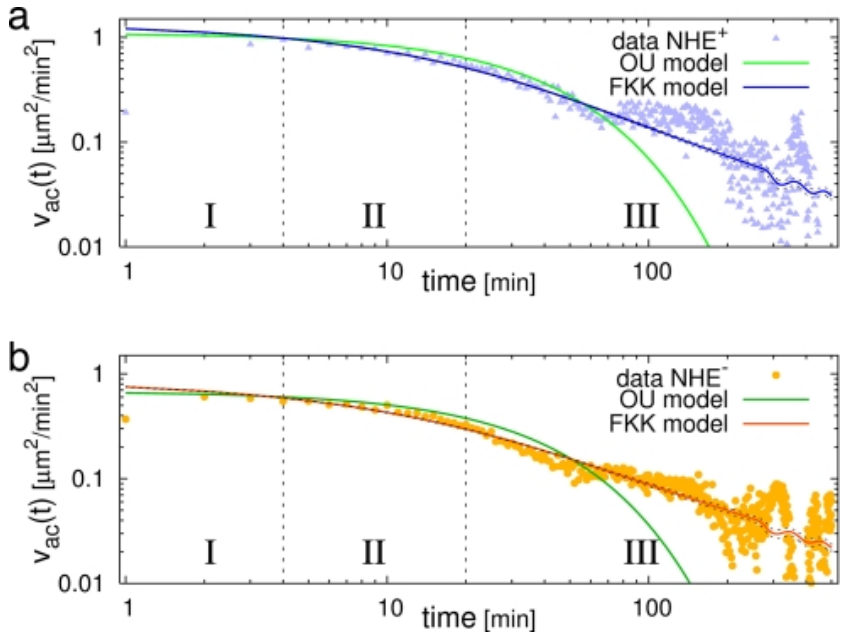

Figure 3.4: Plots of the autocorrelation function. The dotted points are the real data, the green lines are data obtained from the Brownian motion, and the blue and red lines are from the Fractional Klein-Krames model. Figure from [9]

temperature, M is the mass of the particle, and α defines the order of the fractional time derivative. The Brownian motion satisfied Eq. 3.6 in the limit case of pure diffusion.

The various statistical analyses studied in the paper, such as means square displacement, probability distribution over times, kurtosis, and velocity autocorrelation function, suggest that anomalous diffusion is suitable to describe cell motility in the long run, while classical diffusion fits the data better in the short term. This is another instance of bimodal motion.

Physical constraints may induce non-linear diffusion

The dynamics equation of the FKK model was introduced in Dieterich's research, but many other papers would also consider similar approaches. In the anomalous diffusion, the MSD (definition given in 6.20) does not follow a linear dependence on time, so fractional diffusion equations were designed. However, there are instances of anomalous diffusions even from a standard diffusion equation with particular border conditions (Ref. [31]). Similar to our case of study, where the cell is confined to a squared domain (details in Chapter 4).

3.5.2 The Speed-Persistence coupling model

Jerison and Quake (Ref. [21]), have studied the T-cell behaviours *in vivo* through a selective plane illumination microscope, where they could observe native populations of T cells in the live larval zebrafish. One of their primary goals was to find if there is "a statistically consistent behavioural program carried out by these cells" (Jerison and Quake, [21]). A complete standard analysis of the statistics has been proposed, with a clear conclusion that a simple Lévy flight could not predict cell motility in its entirety.

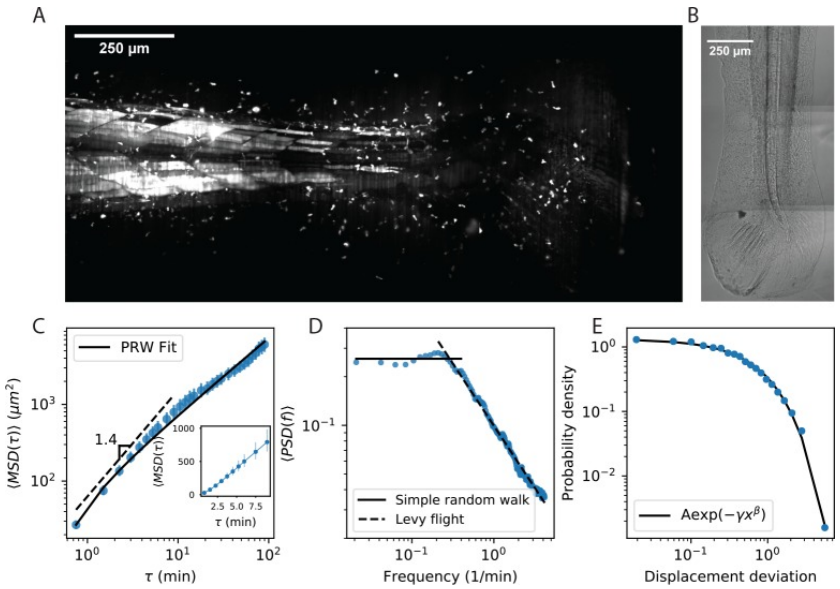

Figure 3.5: Image (a) is the frame of the zebrafish's tail in larvae form. The white dots are the T cells inside the animal. Image (b) is the same image of the zebrafish's tail, but in another channel to show the physical constraints of cell movement. Images (c), (d) and (e) are plots of statistics. Specifically, (c) is the plot of distributional MSD over time in a double logarithmic form, (d) is the plot of the speed over its frequency in a double logarithmic form and (e) is the plot of the frequency of displacement deviation over the d. deviation. Figure from [21]

Even though it seems that the cell has heterogeneous behaviours, like in the case of albatrosses in Section 3.3, the study has brought to model cells' movement as a specific type of Langevin equation, called the persistent Ornstein-Uhlenbeck process, given by

$$\frac{dv_i}{dt} = -\frac{1}{P}v_i + \frac{S}{\sqrt{P}}\xi \tag{3.7}$$

where v_i is the i-th component of the velocity v, S is the speed and P is

the percistence time, or the average time before a cell turns. Jerison and Quake observed that cells when navigating with high speed, tend to have low-angle turns, while low speeds are positively correlated to large-angle turns. As a result, Eq. 3.7 becomes as follows:

$$\frac{dv_i}{dt} = -\frac{1}{\frac{S}{\alpha} + \beta} v_i + \frac{S}{\sqrt{\frac{S}{\alpha} + \beta}} \xi \tag{3.8}$$

where α and β are constant with unites of time. With only a free parameter, the latter equation is called the **Speed-Persistence coupling model**. In their analysis, such an approach seems to describe the effective diffusive behaviour of the cells' trajectories and their scaling at longer times. However, it may not capture all the details of the microscopic dynamics. The study also did not explicitly consider the complex geometry of the area the living cells are (the tail of zebrafish), which may be an important influence on cells' motility.

T-cells, although not as complex as insects or birds, do show a clear biological "behaviour" that is similar: as animals search for food, T-cells search for antigens, by taking inspiration from the Bumblebees experiment, which was set up in a close domain, and considering the loosen formulation of LFH (searching for distribution), the thesis will lean towards verifying if there are some power laws in the probability distributions of step lengths. Furthermore, even if it is not meant for a confined space, an attempt to interpret the MSD with respect to the time will be presented. In addition, by considering the observations made in 3.5.1, one may expect that the majority of the MSD that will be found in our experimental data is going to be sub-diffusive due to the spatial confinement, that is if the cells' lives are long enough to reach the border.

Chapter 4

Experimental Set-up

This chapter is dedicated to discussing the preparation of the experiments, in particular, it will specify where the T-cells come from, where they are placed, how the chosen microscope is set up and how long a single experiment takes (the majority of specifics came from Ref. [32] and [40]).

4.1 Animals preparation

Mice were maintained in a pathogen-free environment with food and water readily available. All procedures involving mice were performed following the National Health and Medical Research Council animal ethics guidelines, approved by the Peter MacCallum Cancer Centre Experimentation Ethics Committee. C57BL/6 and uGFP/OT-1 C57BL/6 mice, which transgenically express both the OT-1 T cell receptor and Green Fluorescent Protein (GFP) at the age of 6 to 30 weeks were utilised. C57BL/6 mice were obtained from Animal Resources Centre (Perth, Australia), while uGFP/OT-1 C57BL/6 mice were generated by breeding OT-1 transgenic mice (acquired from Walter and Eliza Hall Institute, Parkville, Australia) with uGFP mice (acquired from The Jackson Laboratory, Maine, USA).

4.2 Cell culture

- **Dendritic cells**: Cells from the bone marrow were isolated from the hind limbs of C57BL/6 mice and cultured in enriched RPMI medium 1640 containing GM-CSF (10 ng/ml, Peprotech Inc, Rocky Hill, USA). Cells were cultured for 2 days, the media was replenished and cells were cultured for a further 4 days at 37 °C in 5% CO_2 to gener-

ate dendritic cells (CD11c+/CD86lo/MHCIIlo) for use as antigen-presenting cells;

- **Naïve CD8$^+$ T lymphocytes** were isolated from uGFPO/OT-1 C57B-L6 mouse splenocytes by negative selection using the CD8$^+$ T cell isolation kits, as per manufacturer's instructions (EasySep STEM-CELL Technologies). The cells were placed in 14mL polystyrene round-bottom tubes to properly fit into the magnetic carousel. For 1x108 splenocytes, the cells were incubated with Normal Rat Serum (50μL/ mL) and antibody cocktail (50 μL/mL) for 10 minutes and re-suspended in magnetic beads (125 μL/mL) for 5 minutes. The cells were then resuspended in HANKS buffer up to a total volume of 2.5mL. The tube was placed into the magnetic carousel for 2.5 minutes then inverted, pouring off the desired CD8+ T lymphocytes;

Dendritic cells (DCs) and CD8$^+$ T cells were cultured in RPMI medium 1640 enriched with foetal-calf serum (10%v/v), glutamine (1mM, GIBCO-BRL), sodium pyruvate (1nM, GIBCO-BRL), non-essential amino acids (100 nM, GIBCO-BRL), penicillin/streptomycin (100 ng/mL), and minimal essential medium alpha modifications (SAFC Biosciences, Sigma Aldrich).

4.3 T-Cells in microwells

T-cells were placed in microfabricated microwells for cellular studies in vitro. The cell microwells were made of PDMS with dimensions of 125$\times$ 125$\times$45 μm and with well-defined vertical sidewalls and a transparent base. Cell microwells were placed into a well of an 8-well chamber slide (LAB-TEK II, NUNC) sterilized with 100% EtOH and UV light, and rinsed with media prior to use. The gibbon ape leukaemia cell line, MLA-144, eGFP were cultured at 37°C, 10% CO_2 in Dulbecco's Minimal Essential Medium (SECF) supplemented with 10% (v/v) fetal calf serum, L-glutamine (1mM) and 100ng/mL penicillin/streptomycin.

4.4 Microscopy of choice

Time-lapse images were obtained with an IX71 inverted microscope (Olympus, Tokyo, Japan) equipped with a Nipkow disk-type confocal unit (Yokogawa CSU22, Tokyo, Japan) and EM-CCD Andor camera (Model: iXon EM +885, Belfast, Northern Ireland). Images were acquired in both DIC and green channels using a 20x air objective 0.45NA, which corresponded

to a pixel size of 0.33 μm $\times$ 0.33 μm. The working distance was 6.6-7.8 mm. Exposure time was 600 ms for green and 100 ms for DIC. Multiple stage positions were captured (controlled by MetaMorph software version 7.7.11.0) with a sampling rate of 30 seconds for 1-24 hrs and were saved as 8-bit 2-D arrays (380 × 380 pixels).

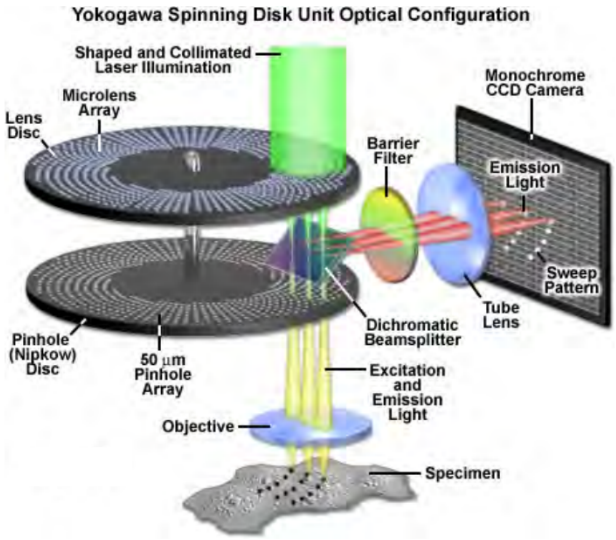

Figure 4.1: Schematic representation of Yokogawa spinning disk confocal laser configuration, image from Ref. [40]

Once everything was ready, a single $CD8^+$ T cell would be activated by a dendritic cell and placed in each well at a fixed time in the morning. Due to the activation, the cell would roam around inside the well and start mitosis. The microscope would take a picture every 30 seconds for several hours. After that, the daughter cells would be collected and the micro-well washed before placing such cells again in the well in a random position and restarting the data collecting process again. The environment is isolated and sterilised. A stream of humid gas is released from a corner of the chamber to guarantee oxygen and carbon dioxide supplies to the cells.

Chapter 5

Cells Tracking and Data Acquisition

This chapter describes the program used to track cells and extract numerical data after the frames of micro-wells have been acquired from the microscope. Eleven independent videos were processed by the program and Excel files were generated for numerical analysis.

5.1 Video Labeling

The acquisition of the video was done more than a decade ago and the history of the original cells is not clear. In addition, the off-springs lineage between videos is not available[1] due to the multiple changes of the files' names. The label for the current thesis is based on the last known change: **Video01**, **Video02**, **Video03**, **Video04**, **Video05**, **Video06**, **Video07**, **Video11**, **Video16** and **Video41**. Each of them contains frames of the cells in the Fluorescent, Bright field and Red channels. For this Thesis, only the Fluorescent and Bright field channels are relevant, while the rest are used for a more biology-centric study. Each video contains information on 4 micro-wells, and each of them contains at least one cell when the microscopy was recorded. To label every single well, the cardinal points are used and each well is identified with NW (North-West), NE (North East), SW (South-West) and SE (South East). For instance, Video01.SE refers to the well on the bottom-right in Video 01.

[1] After a discussion of how important it is to retrace the history of the cells, Sarah Russell, one of the team leaders of our group at Swinburne University of Technology, kindly offered her help to reconstruct the whole lineage. However, the reconstruction is

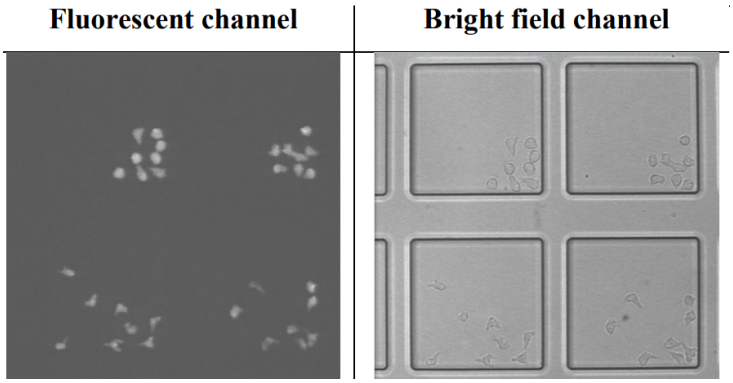

Figure 5.1: Instance of 2 channels of the micro-well: Fluorescent channel on the left and Bright field one on the right. Both of them represent the picture of the same time frame. The blobs in the wells are the T cells.

5.2 DeepKymoTracker

The program used for tracking the cells is called DeepKymoTracker [2], written by Khelina Fedorchuk. It was developed in Python™ ver. 3.6.13, and contains the applied concept of convolutional neural network. Five main steps characterise Khelina's work when I used it:

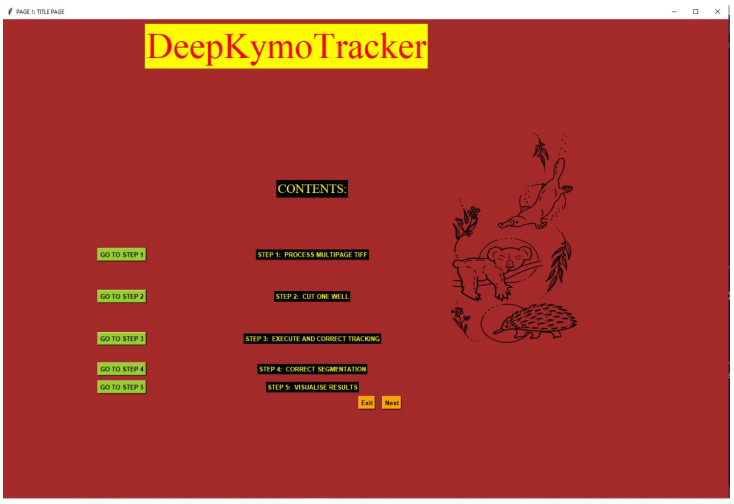

Figure 5.2: DeepKymoTracker Starting Interface

still underway and beyond the scope of this thesis

[2]More detailed information can be found in Khelina's Github/repository: https://github.com/khelina/T-cell-lineages-tracking

- **Step 1** File adjustment: The first step is the adjustment of the file format obtained from the microscope and conversion into a suitable format for cell-tracking;
- **Step 2** Cut one well: this is the step where the frame-cutting of a single well is done. Both fluorescent and bright field channels are cut and stored in a specific file, ready to be used in the **Step 3**;
- **Step 3** Track and Segment T cells: this is the main focus of the program. As soon as the file from **Step 2** is added, the user can start and let the program track the cells in the well. As the cells are recognized in each time frame, a contour is made around it and a label of them is superimposed; Each frame is developed and displayed to the user, with the possibility of manual correction. The analysis of the cell is done by using the fluorescent channel, while the drawing of the contour is made in the bright field one. While the frames with segmentations are produced, a paired image of tree lineage is developed at the same time. At the end of the process (which can not be paused unless for correcting wrong labelling or segmentation), a movie of contoured cells paired with the lineage tree of the cells can be generated;

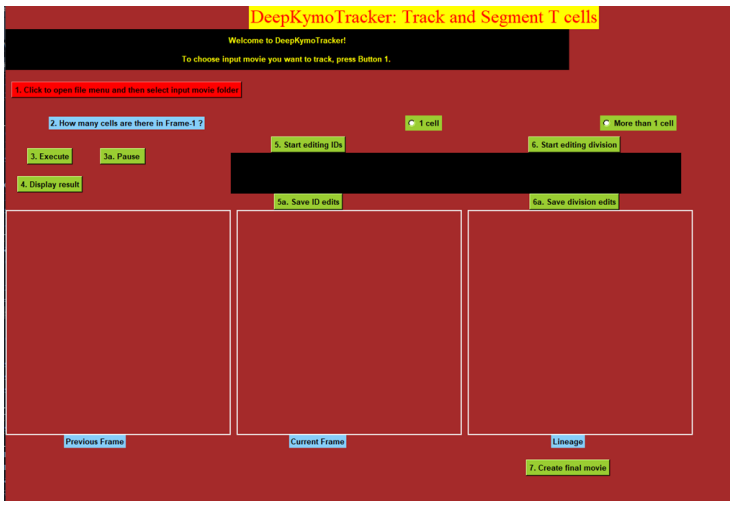

Figure 5.3: DeepKymoTracker's **Step 3** starting interface, image from [12]

- **Step 4** Correct Segmentation[3]: This is a further step to correct possible wrong segmentations which may occur in **Step 3**. Even though the possibility of correcting the segmentation is already implemented in the **Step 3**, the **Step 4**'s feature is more robust and user-friendly: it

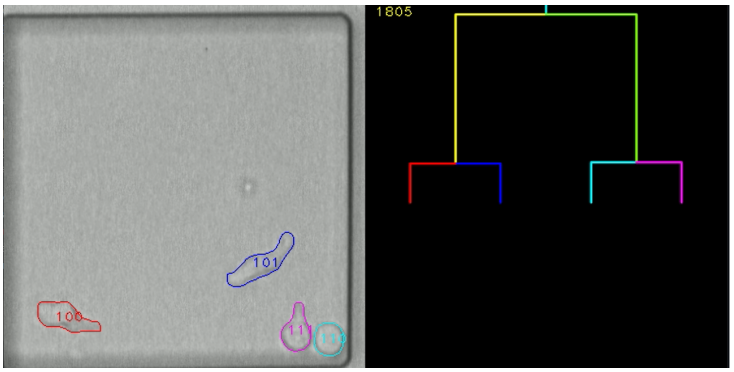

Figure 5.4: Video03.NE frame 1805. On the left panel, there are 4 contoured and labelled cells in the well. On the right panel, the cells' tree lineage is displayed.

provides the option to manually draw the cell's contour and enables the user to do the correction later.

- **Step 5** Excels files extraction: the last passage is the production of numerical data from the contoured cells. Specifically, the features obtained are the coordinates of the centroid, the area, the perimeter, the circularity of the cells;

It is worth noting that the program can generate much more than what has been listed previously. For example, the complete program in **Step 3** can also produce various plots showing changes in area and perimeter over time. However, due to time and computer storage constraints[4], Khelina made personalized adjustments to the code for a faster generation of data, leaving the unnecessary features behind. To give a general idea, generating plots for the Excel file could take anywhere from 5 to 20 hours for a single micro-well video.

More details on T-cells' data

The well, 125×125 μm^2, is displaced as a 380×380 pixels resolution image. The centroid of the cell is calculated as the weighted average of the pixels:

$$\mathbf{x}_c = \frac{\sum_{i \in P} w_i \mathbf{x}_i}{\sum_{i \in P} w_i} \tag{5.1}$$

[3] I was the first "beta tester" of Khelina's work, and some trial and error were involved. Due to time constraints during my stay in Australia, not all steps were thoroughly tested and adjusted for optimal performance. However, since this step was added by Khelina during my testing, I can proudly claim that I contributed to its improvement.

[4] and to keep procedure within workable times.

where w_i are the weighs, P is the set of pixels that defines a single cell, $\mathbf{x}_c$ = (x,y) are the coordinates of the cell's centroid and $\mathbf{x}_i$ is the position of the i-th pixel. The weights are considered all the same equal to 1, $\omega_i = 1$ $\forall i$. Area and perimeter are found with the same pixel-contour logic while circularity can be derived with the following formula:

$$Circularity = 4\pi \frac{Area}{Perimeter^2} \tag{5.2}$$

Boolean digits are used to label the cells in the following way: at the beginning of a new video, a single cell is placed in each micro-well. Such a cell has the label of 1. When mitosis occurs, the two daughter cells would receive the label from the previous cell with an additional boolean number that would distinguish them. The figure (5.5) gives a representation of a generational tree starting from cell 1 until the fourth generation:

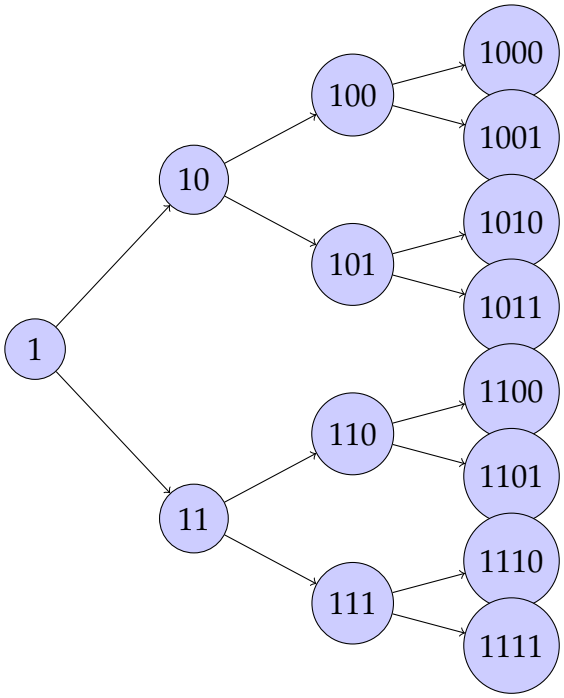

Figure 5.5: A tree that represents the mitosis of T cells and the labels which are given to the daughter cells up to the fourth generation

Whenever it is necessary to refer to a specific cell, the following complete labelling will be used:

(**Video** + number of video) . (cardinal points) . (**Cell** + cell boolean number)

For instance, if it is important to look at cell 100 in the top left corner well on video 04, the label is **Video04.NW.Cell100**.

Limitations and Complications

There are a few aspects of the DeepKymoTracker that impact directly on the numerical analysis of cells' movement.

- The only statistics that directly reflects a progression of cells' motion is centroids, while information about the movement of cells in their entirety is unavailable;
- The density of the cell might be different across its body, which makes the formula 5.1 an approximation of the true cell's centroid;
- The segmentation, when cells are attached (which is guaranteed after mitosis), is an approximation of a proper segmentation. This is mainly due to the impossibility of the program to spot and draw overlapping contours, and its inability to process true 3D images;

Figure 5.6: Video03.NE, **Cell11** and **Cell10** in the frame 405 are displayed. **Cell10** should also have a tail that overlaps with **Cell11**, but due to the limits of DeepKymoTracker, it is not registered for **Cell10**, causing a further error for **Cell10**'s centroid

- In **Step 2** of the program, the frame-cutting logic did not consider possible, spurious, camera movements[5]. As a result, a post-process correction is needed after the extraction of the Excels file at the end of **Step 4** to correct for inaccurate positions due to such movements (See Matlab file for the specifics changes);

[5]In the latest version of DeepKymoTracker, an extra feature was implemented by Khe-

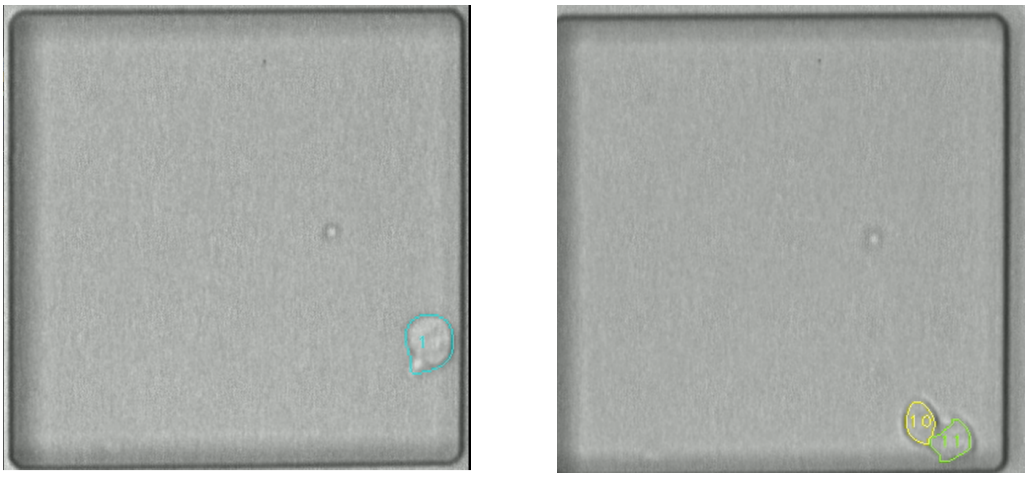

Figure 5.7: Left image: Video03.NE.Cell1 frame 1. Right image: Video03.NE.Cell10 and Video03.NE.Cell11 frame 486. Camera shifting is evident from the left image to the right image, especially in the top left corner

lina to correct the camera's movement. Unfortunately, due to the timing and tool constraints, that option was not available when data was collected for this work.

Chapter 6

Mathematical Concepts

This chapter collects all the mathematical tools that will be essential for the study proposed next. In particular, this part includes the Cell-tracking definition, the Random Walk, the diffusion equations and the auto-correlation function,

An important matter is addressed: when a physical measurement is with a tilde, it refers to its value with respect of micrometres and seconds, while the tilde-less version has the unit of measure with respect to frames and time steps.

$$t[\text{time steps}] \quad \tilde{t}[\text{sec}] \qquad x[\text{pixels}] \quad \tilde{x}[\mu m]$$

The conversion from one measurement to another has the following rules:

$$1 \text{ pixels } = \frac{380}{125}\mu m \qquad 1 \text{ time step } = 30 \text{ sec}$$

6.1 Cell - track

Definition. A 2-dimensional **cell track** is a finite n-sequence

$$\mathrm{T} := [\ (\mathbf{x}(1),\ \tilde{t}(1)),\ \ldots,\ (\mathbf{x}(k),\ \tilde{t}(k)),\ \ldots,\ (\mathbf{x}(n),\ \tilde{t}(n))\ ] \tag{6.1}$$

where $\mathbf{x}(\mathrm{k}) \in \mathbb{R}^2$ is the k-th position and $\tilde{t}(\mathrm{k}) \in \mathbb{R}$ is the k-th time in second. The value in parentheses is the time in time steps.

For population-based cell track analysis, it is common to align a set of tracks to the same starting point.

Definition. For a cell track T, the track that results from subtracting the first element from all elements in the sequence i.e.

$$\begin{aligned} \mathrm{T}^0 :=& [\ ([0,0],0),\ (\mathbf{x}(2)-\mathbf{x}(1),\ \tilde{t}(2)-\tilde{t}(1)),\ \ldots, \\ & (\mathbf{x}(k)-\mathbf{x}(1),\ \tilde{t}(k)-\tilde{t}(1)),\ \ldots,\ (\mathbf{x}(n)-\mathbf{x}(1),\ \tilde{t}(n)-\tilde{t}(1))\ ] \end{aligned} \tag{6.2}$$

is called **zero-aligned cell track set**, or zero-aligned version of T.

6.2 Speed

The speed of a cell with respect to T_0 with length n is defined as follows:

$$\begin{aligned} V^0 :=& \Big[\ 0,\ \sqrt{V_x^0(1)^2+V_y^0(1)^2},\ \sqrt{V_x^0(2)^2+V_y^0(2)^2},\ldots, \\ & \sqrt{V_x^0(k)^2+V_y^0(k)^2},\ \ldots,\ \sqrt{V_x^0(n-1)^2+V_y^0(n-1)^2}\ \Big] \\ =& [\ 0,\ V^0(1),\ V^0(2),\ \ldots,\ V^0(k),\ \ldots,\ V^0(n-1)\ ] \end{aligned} \tag{6.3}$$

where V_x^0 and V_y^0 are defined as follows, with **x**(k) = (x(k), y(k)):

$$\begin{aligned} V_x^0 :=& \left[\ 0,\ \frac{x(2)-x(1)}{2-1},\ \ldots,\ \frac{x(k)-x(k-1)}{k-(k-1)},\ \ldots,\ \frac{x(n)-x(n-1)}{n-(n-1)}\ \right] \\ =& [\ 0,\ x(2)-x(1),\ \ldots,\ x(k)-x(k-1),\ \ldots,\ x(n)-x(n-1)\ ] \\ =& [\ 0,\ V_x^0(1),\ \ldots,\ V_x^0(k),\ \ldots,\ V_x^0(n-1)\ ] \end{aligned} \tag{6.4}$$

$$\begin{aligned} V_y^0 :=& \left[\ 0,\ \frac{y(2)-y(1)}{2-1},\ \ldots,\ \frac{y(k)-y(k-1)}{k-(k-1)},\ \ldots,\ \frac{y(n)-y(n-1)}{n-(n-1)}\ \right] \\ =& [\ 0,\ y(2)-y(1),\ \ldots,\ y(k)-y(k-1),\ \ldots,\ y(n)-y(n-1)\ ] \\ =& [\ 0,\ V_y^0(1),\ \ldots,\ V_y^0(k),\ \ldots,\ V_y^0(n-1)\ ] \end{aligned} \tag{6.5}$$

6.3 Angle

The direction of cells' displacement along the x and y-axis has been analysed because it aids the characterization of cells' motility regimes. For each zero-aligned cell track set T^0 of length n, a set of angles of length n with respect to T_0 is calculated in the following way:

$$\begin{aligned}\theta^0 := \; & [0, \; f(V_x(1), V_y(1)), \; f((V_x(2), V_y(2)), \; \ldots, \\ & \qquad f((V_x(k), V_y(k)), \; \ldots, \; f((V_x(n-1), V_y(n-1))] \\ = \; & \Big[\; 0, \; \theta^0(1), \; \theta^0(2), \; \ldots, \; \theta^0(k), \; \ldots, \; \theta^0(n-1) \; \Big] \end{aligned} \tag{6.6}$$

where $f : \mathbb{R}^2 \to [0, 2\pi)$ and $\mathrm{f}(\mathrm{V}_x, \mathrm{V}_y) \in [0, 2\pi)$

$$f(x,y) := \begin{cases} 0 & \text{if } V_x \geq 0 \wedge V_y = 0 \\ \arctan\left(\frac{V_x}{V_y}\right) & \text{if } V_x > 0 \wedge V_y > 0 \\ \frac{\pi}{2} & \text{if } V_x = 0 \wedge V_y > 0 \\ \arctan\left(\frac{V_x}{V_y}\right) + \pi & \text{if } V_x < 0 \wedge V_y > 0 \\ \pi & \text{if } V_x < 0 \wedge V_y = 0 \\ \arctan\left(\frac{V_x}{V_y}\right) + \pi & \text{if } V_x < 0 \wedge V_y < 0 \\ \frac{3\pi}{2} & \text{if } V_x = 0 \wedge V_y < 0 \\ \arctan\left(\frac{V_x}{V_y}\right) + 2\pi & \text{if } V_x > 0 \wedge V_y < 0 \end{cases} \tag{6.7}$$

For a lighter notation, as in the sequences of V^0 and θ^0, the apex 0 is implied but not specified since the zero-align cell track set will always be considered for the calculus of the statistics and measurements.

6.4 Random Walk

A random walk is a path derived from a sequence of random steps on some mathematical space. In our case, which is biology-based, the walkers can be the cells and the walk is on a 2D square lattice. The space is divided into squares and at each time every single cell can move onto the next possible "tile", a neighbour of the current square the cell is in. For the sake of simplicity, no cell division is present in the movies we consider. In our case, the mass is referred to the number of the cells.

Let $\mathrm{c}_{i,j}$(t) be the concentration of cell in the (i,j)-tile at time t. According to figure 6.1, the concentration in the square (i,j) at time $t + \tau$ is the sum of three contributions: the concentration in the square (i,j) at time t, the positive inflow of cells from the neighbours "tiles" and the negative outflow of cells from the square (i,j) to the neighbours "tiles". Mathematically:

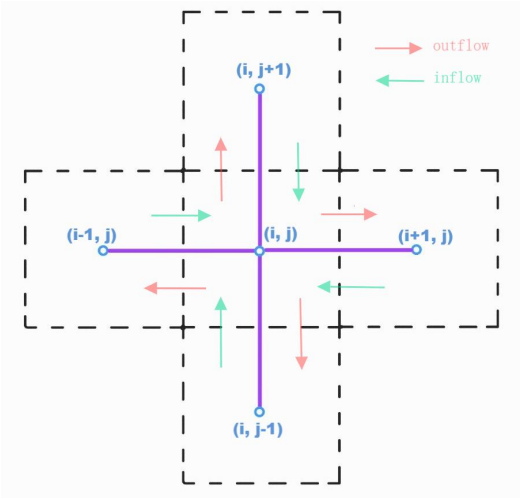

Figure 6.1: Random Walk in a square lattice: the purple lines are the step length a cell can take in a unit of time. A possible jump from one tile to another is modelled to happen from one tile's centre to another. The red arrows represent the outflow probability and the green ones represent the inflow probability.

$$c_{i,j}(t+\tau) = c_{i,j}(t) + \Big(p_{(i+1\to i)}c_{i+1,j}(t) + p_{(i-1\to i)}c_{i-1,j}(t) + p_{(j+1\to j)}c_{i,j+1}(t) + p_{(j-1\to j)}c_{i,j-1}(t)\Big) - \Big(p_{(i\to i+1)}c_{i,j}(t) + p_{(i\to i-1)}c_{i,j}(t) + p_{(j\to j+1)}c_{i,j}(t) + p_{(j\to j-1)}c_{i,j-1}(t)\Big) \tag{6.8}$$

By taking the first term of the RHS of the equation 6.8 to the LHS, we get the variation of cell in the (i,j)-tile in a τ time as the displacement of cells' flows

$$c_{i,j}(t+\tau) - c_{i,j}(t) = \text{inflow of cells} - \text{outflow of cells} \tag{6.9}$$

Which respects the conservation of mass. Now, it can be shown that one may get from the equation 6.8 a macro-equation such as the classical diffusion equation, by substituting as follows:

$$p_{i\to\hat{i}} = \frac{D\tau}{\Delta x^2} \quad p_{j\to\hat{j}} = \frac{D\tau}{\Delta y^2} \tag{6.10}$$

If eq. 6.10 is implemented into eq. 6.8 the following is obtained:

$$\frac{c_{i,j}(t+\tau)-c_{i,j}(t)}{\tau} = D\left(\frac{c_{i+1,j}(t)-2c_{i,j}(t)+c_{i-1,j}(t)}{\Delta x^2} + \frac{c_{1,j+1}(t)-2c_{i,j}(t)+c_{i,j-1}(t)}{\Delta y^2}\right) \tag{6.11}$$

As τ, Δx^2, $\Delta y^2 \to 0$ and c $\in C^2$

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial x^2}+\frac{\partial^2 c}{\partial y^2}\right) = D\Delta c \tag{6.12}$$

Where $\Delta(.) := \left(\frac{\partial^2(.)}{\partial x^2}+\frac{\partial^2(.)}{\partial y^2}\right)$ is the 2D Laplace operator. Eq. 6.12 is the classic diffusion equation. Alternatively, one may obtain the same result by starting from 6.8, setting all the probability as p and c$\in$ C^2, and using the Taylor expansion as the jump from one tile to the neighbour is infinitesimal (Δx and $\Delta y \to 0$):

$$\begin{aligned}
c_{i+1,j} &\approx c_{i,j} + \frac{\partial c_{i,j}}{\partial x}\Delta x + \frac{\partial^2 c_{i,j}}{\partial x^2}\frac{\Delta x^2}{2}\\
c_{i-1,j} &\approx c_{i,j} - \frac{\partial c_{i,j}}{\partial x}\Delta x + \frac{\partial^2 c_{i,j}}{\partial x^2}\frac{\Delta x^2}{2}\\
c_{i,j+1} &\approx c_{i,j} + \frac{\partial c_{i,j}}{\partial y}\Delta y + \frac{\partial^2 c_{i,j}}{\partial y^2}\frac{\Delta y^2}{2}\\
c_{i,j-1} &\approx c_{i,j} - \frac{\partial c_{i,j}}{\partial x}\Delta y + \frac{\partial^2 c_{i,j}}{\partial y^2}\frac{\Delta y^2}{2}
\end{aligned}$$

Einstein obtained the same formula in his paper on the theory of diffusion [11] by studying the phenomena previously reported by Robert Brown in 1828 of small granules in still water. If a cell is released at the origin at time t = 0 in an unbounded space, the solution of the diffusion equation is the normalized Gaussian PDF:

$$c(\mathbf{x},t) = \frac{1}{4\pi D t}e^{-\frac{\|\mathbf{x}\|^2}{4Dt}} \tag{6.13}$$

Einstein remarks that this solution *is that of the fortuitous error, which was to be expected*. By knowing the expression of c(**x**,t) 6.13 we can obtain the mean value and the covariance matrix (μ and Σ) of **x**, with an important reminder:

$$\frac{1}{4\pi Dt}e^{-\frac{\|\mathbf{x}\|_2^2}{4Dt}} = \frac{1}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}\frac{1}{\sqrt{4\pi Dt}}e^{-\frac{y^2}{4Dt}} = c(x,t)c(y,t) \tag{6.14}$$

where c(x,t) is the concentration of cells only along the x-axis at time t and c(y,t) is the concentration of cells only along the y-axis at time t.

$$\mu := (< x >, < y >) \tag{6.15}$$

$$\Sigma := \begin{pmatrix} < x^2 >, 0 \\ 0, < y^2 > \end{pmatrix} \tag{6.16}$$

Where

$$\begin{aligned} < x >= \mathbb{E}[x] &:= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xc(\mathbf{x},t)dxdy = \int_{-\infty}^{+\infty} xc(x,t)dx \\ &= \int_{0}^{+\infty} xc(x,t)dx - \int_{0}^{+\infty} xc(x,t)dx = 0 \end{aligned}$$

$$\begin{aligned} < y >= \mathbb{E}[y] &:= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yc(\mathbf{x},t)dxdy = \int_{-\infty}^{+\infty} yc(y,t)dx = \\ &= \int_{0}^{+\infty} yc(y,t)dx - \int_{0}^{+\infty} yc(y,t)dx = 0 \end{aligned}$$

$$< x^2 >= \mathbb{E}[x^2] := \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2c(\mathbf{x},t)dxdy \overset{(*)}{=} 2Dt \tag{6.17}$$

$$< y^2 >= \mathbb{E}[y^2] := \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2c(\mathbf{x},t)dxdy \overset{(*)}{=} 2Dt \tag{6.18}$$

$$\begin{aligned} (*) = \int_{-\infty}^{+\infty} x^2c(x,t)dx &= \int_{-\infty}^{+\infty} xxc(x,t)dx \\ &\overset{6.13}{=} -\frac{x}{2}4Dtc(\mathbf{x},t) + 2Dt\int_{-\infty}^{+\infty} c(\mathbf{x},t)dx \overset{6.13}{=} 2Dt \end{aligned} \tag{6.19}$$

The distributional mean square displacement (MSD) is defined in the following way:

$$< \delta^2(t) >:= \mathbb{E}[\|\mathbf{x}(t) - \mathbf{x}(0)\|^2] = \mathbb{E}[(x(t) - x(0))^2 + (y(t) - y(0))^2] \tag{6.20}$$

In a standard diffusion process, the MSD is linearly dependent over t:

$$< \delta^2(t) > \overset{(**)}{=} 4Dt \tag{6.21}$$

$$\begin{aligned}
(**) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x^2+y^2)\frac{1}{4\pi Dt}e^{-\frac{x^2+y^2}{4Dt}}dxdy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2\frac{1}{4\pi Dt}e^{-\frac{x^2+y^2}{4Dt}}dxdy + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y^2\frac{1}{4\pi Dt}e^{-\frac{x^2+y^2}{4Dt}}dxdy \\
&= 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2\frac{1}{4\pi Dt}e^{-\frac{x^2+y^2}{4Dt}}dxdy \\
&= 2\int_{-\infty}^{+\infty}x^2\frac{1}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}dx\int_{-\infty}^{+\infty}\frac{1}{\sqrt{4\pi Dt}}e^{-\frac{y^2}{4Dt}}dy \\
&= 2\cdot 2Dt = 4Dt
\end{aligned} \tag{6.22}$$

Observation: It is direct to extend 6.21 to a n-dimension case and the formula would be

$$< \delta^2(t) >= 2nDt \tag{6.23}$$

Around 9 years after Einstein's paper, Ivar Nordlund came up with the idea of studying mercury droplets [30] by performing the analysis of every single trajectory. The mathematical tool was the time-averaged MSD:

$$\overline{\delta^2(\Delta)} = \frac{1}{t-\Delta}\int_0^{t-\Delta}\|\mathbf{x}(t'-\Delta) - \mathbf{x}(t')\|^2 dt' \tag{6.24}$$

where Δ is called the lag time, the width of the window slid along the cell track with typically $\Delta \ll t$ for good statistics. In the classical 2D random walk

$$\overline{\delta^2(\Delta)} = 4Dt \tag{6.25}$$

which is the same as eq. 6.21, implying the property of ergodicity: ensemble averages and long-time averages are equivalent in the limit of long measurement times. However, this is not true for our phenomena (as we will see).

Estimators of distributional MSD and time-averaged MSD

The MSD(t) can be estimated through $D^2(t)$

$$\hat{D}^2(t) = \frac{\|x_1(t)\|^2 + \cdots + \|x_n(t)\|^2}{n} \tag{6.26}$$

where $x_1(t)$, ..., $x_n(t)$ are the positions of n zero-aligned cell tracks at some fixed time t.

The MSD*(t) can be estimated through $\hat{D}^*(\tau = mt_{int})$

$$\hat{D}^*(\Delta = mt_{int}) = \frac{1}{N-m} \sum_{s=1}^{N-m} \|\mathbf{x}(m+s) - \mathbf{x}(s)\|^2 \tag{6.27}$$

where N is the total number of timesteps and t_{int}.

6.4.1 Generalization of Random Walk

It is interesting to note that with different hypotheses of probability $p_{k\to\tilde{k}}$, $k \in i, j$, one may obtain different variations of the diffusion equation: for example from anisotropic diffusion equations to anisotropic convection-diffusion equations. A further generalisation can be implemented by considering rectangle or hexagonal tiles instead of square tiles or considering a step length that is different from one. The following is implemented for the second part of Chapter 8.
In the case of the square 2D lattice, the direction probability is given by (lattice Fig. 6.1):

$$p_{\text{East}} = \frac{n(]-\frac{\pi}{4}, \frac{\pi}{4}])}{n(]0, 2\pi])} \qquad p_{\text{North}} = \frac{n(]\frac{\pi}{4}, \frac{3\pi}{4}])}{n(]0, 2\pi])}$$

$$p_{\text{West}} = \frac{n(]\frac{3\pi}{4}, \frac{5\pi}{4}])}{n(]0, 2\pi])} \qquad p_{\text{South}} = \frac{n(]\frac{5\pi}{4}, \frac{7\pi}{4}])}{n(]0, 2\pi])}$$

In the case of the hexagon 2D lattice, the direction probability is given by (lattice Fig. 6.2):

$$p_{\text{East}} = \frac{n(]-\frac{\pi}{6}, \frac{\pi}{6}])}{n(]0, 2\pi])} \qquad p_{\text{North-East}} = \frac{n(]\frac{\pi}{6}, \frac{3\pi}{6}])}{n(]0, 2\pi])} \qquad p_{\text{North-West}} = \frac{n(]\frac{3\pi}{6}, \frac{5\pi}{6}])}{n(]0, 2\pi])}$$

$$p_{\text{West}} = \frac{n(]\frac{5\pi}{6}, \frac{7\pi}{6}])}{n(]0, 2\pi])} \qquad p_{\text{South-West}} = \frac{n(]\frac{7\pi}{6}, \frac{9\pi}{6}])}{n(]0, 2\pi])} \qquad p_{\text{Sourth-East}} = \frac{n(]\frac{9\pi}{6}, \frac{11\pi}{6}])}{n(]0, 2\pi])}$$

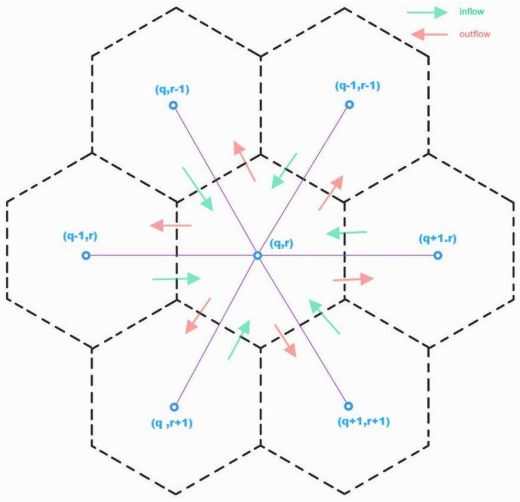

Figure 6.2: Random Walk in a hexagon lattice: the purple lines are the step length a cell can take in a unit of time. A possible jump from one tile to another is modelled to happen from one tile's centre to another. The red arrows represent the outflow probability and the green ones represent the inflow probability.

Where $n(]a,b])$ is the function that counts the number of angles from a to b of the vector θ^*.
The step length taken at each time is a "bootstrap" of the corresponding V^* distribution. For instance, if the **Cell1** of interest is **video04.SE.Cell1** for the simulation, a random pick from the pool of V^{04SE1} with respect to the vector of angles θ^{04SE1}, is considered.

6.5 Speed autocorrelation function

The **autocorrelation function** of cells' speed can have the following expression

$$\begin{aligned} v^{AC}(t_0) &= \frac{E[(V(t+t_0) - E[V(t+t_0)])(V(t) - E[V(t)])]}{\sqrt{(Var(V(t+t_0))Var(V(t)))}} \\ &= \frac{\text{Cov}(V(t+t_0), V(t))}{\text{Var}(V(t))} \end{aligned} \tag{6.28}$$

Estimator of the speed auto-correlation coefficient

The estimator of the autocorrelation coefficient is obtained as follows:

$$v^{\hat{AC}}(l) = \frac{\sum_{i=1}^{n-l}(V_i - \hat{\mu}_V)(V_{l+i} - \hat{\mu}_V)}{\sum_{i=1}^{n}(V_i - \hat{\mu}_V)^2} \tag{6.29}$$

where l is the time lag, the discretized time between $t - t_0$, and $\hat{\mu_V}$ is the sample average of V.

Chapter 7

Numerical results

This chapter contains all the numerical analysis and results from the statistics defined in Chapter 6. A few complications are considered, due to the limited data available and quality of information. Further details are in the section 7.1. All the calculations and data extraction from the Excel files provided by the DeepKymoTracker are processed in the software Matlab 2023a whose specifics are found in the Appendix.

7.1 Complications

The experiment and the data acquisition of the T-cells were not meant to be studied for the specific goal of this thesis. The design of the experiment had other motives, much more biological-based and to answer others equally, if not more important and intriguing questions like:

- which is the generation that decided to manifest a certain gene [44] in its off-springs;
- is it possible to design a precise cell-tracker program that can perfectly capture cell division and the cell's shape right after the mitosis [13] [40];

7.2 Observations of the Video

In the visual analysis of the videos, some biological behaviours have been analysed and mentioned in this section. A more detailed explanation is given in the next sections with more mathematical-based reasoning.

- It seems that every single cell in the well has two main phases: the first is the "wandering" phase and the second is the "division" phase. In the former, the cell focuses more on exploring the surrounding environment, while in the latter, the cell would slow down, acquire a round shape and become slower before the proper splitting into two daughter cells;
- The mitosis seems to occur in the proximity of the wall, which makes one wonder if the T-cell would naturally search for a firm place to split itself, or if the remaining time is enough for it to wander close to the wall;
- If the previous point has a validity of some sort, which is the natural behaviour of T-lymphocytes to desire a firm place to split, then by observing some cells, a new hypothesis may arise: T-cells may have failed attempts of mitosis. This hypothesis is driven by the fact that sometimes, the cell would close to the wall and assume the typical round shape for mitosis;
- From time to time, some cells happened to have a part of them stuck in a single point of the ECM. The nature of such phenomena is unknown, but it does interfere actively in the cells' movement, so in the numerical analysis of position and speed such case has to be taken into consideration;

Unfortunately, the data available during the data acquisition in my stay at the Swinburne University of Technology were not sufficient, since they were limited in numbers and Khelina's program had not been fully tested, as mentioned in Chapter 5. Even though the responsible for the experiment was meticulous with his routine, that did not prevent the fluctuational average lifetime of the T-cells. In some cases, the lifetime of **Cell1** is over a thousand frames, other times it is under five hundred. The first case is labelled as "good data" while the second is labelled as "bad" data, If the number of frames is under two hundred then it is considered "very bad" data. Furthermore, the entire history of such cells is unavailable, even with thousands of frames per single **Cell1**, due to the unknown tree lineage between the cells in different videos. Lastly, no rigorous action was considered when a new experiment was started, resulting in an incomplete record of the entire cells' life by the microscope. Filtering the incomplete data was attempted, resulting in even fewer data to study.

Before we dive into the first statistics, it is crucial to mention the two main influences[1] of **Cell1**'s movement during its entire life-span: the wall of the wells and the internal biological changes.

- **Walls**: The influence of the T-cells' activity by the wall is somewhat known. In every video available, the cell seems to have some "awareness" of such physical constrain, since the collision of any cells toward the obstacles never occurred and the possibility of attraction is present. However, the walls certainly limit the movement of the living cells.

- **Biological parameters**: Biological parameters refer to the different internal component of the cells that changes during the cells' lives. Internal forces and internal movement of cells are not detectable in this study and may still influence the statistics considered. An instance of it can be seen in mitosis; part of this very complex intercellular process is the rearrangement of cytoskeletons in the cytoplasm of the cell [35]. The density of the cell may vary and the cells' centroid should not be calculated as in 5.1. Another instance of biological parameter change is the manual separation of cells in a well by the experimentalist. This may cause an active response of the cells to change their superficial and internal state, resulting in a variation in its motility from the one it could have without human intervention.

7.3 Position

With its simplicity, the cell's centroids in the wells could give some very informative results. By plotting all the positions of the available **Cell1**'s centroids, it is possible to see the uneven spread of positions, with a considerate preference for staying close to the wall. This may be caused by the attraction of cells toward the wall for its stiffness. Another equally viable hypothesis is that the wall is an obstacle for the cells, unable to proceed further. In addition, since the cells would not collapse onto the wall, they seem to be aware of the physical constraints.

The starting position of **Cell1**s is evenly spread all over the well due to the manual separation, which could bring more complex initial conditions. On the other hand, the ending point of every **Cell1** seems to be close to the wall (figure 7.2), leading to an additional possibility that they may prefer to stay close to the wall for the incoming mitosis. Note that the videos stop when **Cell1** divides.

[1]These are the only 2 main factors for $\mathbf{Cell}_1$; if the other cells are present in the well, then a third major influence has to be considered which is the presence of multiple cells, not considered here.

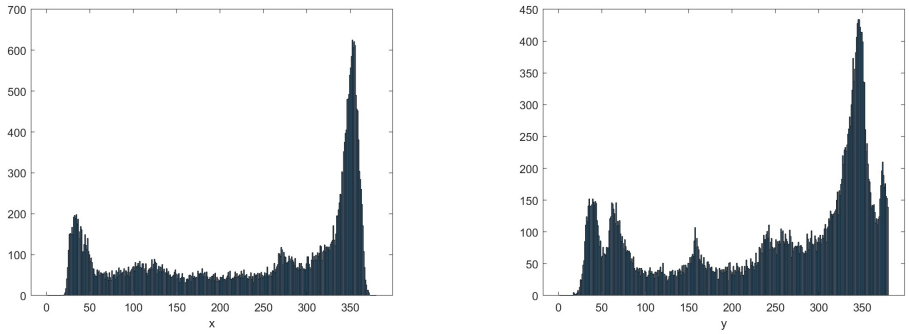

Figure 7.1: Histograms of positions along the x and y-axis. On the abscissa, there are the classes of positions in the well. On the ordinate, there are the counts of each class of position.

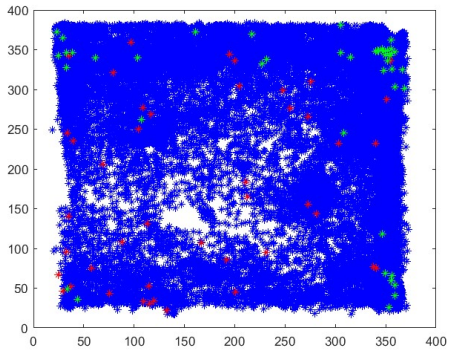

Figure 7.2: All recorder positions are plotted in blue, the red dots are the starting point of the cells, while the green ones are the ending positions

By analysing the positions of each single cell as a path, different types of motions seem to be present. There is the type of path that makes the cell wander across all the well, even when they might need all their life span to do it (figure 7.3, **Video01.NW.Cell1**) or they might only need a fraction of their life (figure 7.3, **Video11.SW.Cell1**). In contrast, there are cases where the cells' movement is not widespread and remains in a particular part of the well, close to the wall (**Video03.SW.Cell1**), suggesting once again the attraction of cells towards the wall. Another particular feature that can be observed in **Video01.SE.Cell1** is the persistency of motion towards a direction for a considerate amount of frames.

Whether the roaming is across all over the well or in a very concentrated area close to the well, they all do not resemble a typical 2D random walk. The random walk doesn't have a particular direction of choice, while the

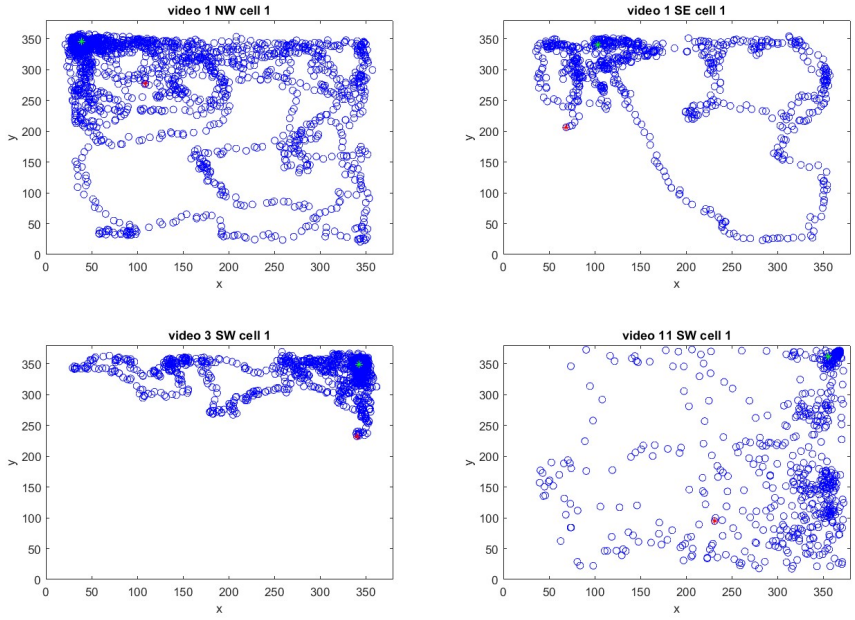

Figure 7.3: Plots of 4 Cell1s' paths. On the top left **Video01.NW** (frames 1761), on the top right **Video01.SE** (frames 713), on the bottom left **Video03.SW** (frame 1121), on the bottom right **Video11.SW** (frames 594)

Cell1 seems to have a preference of direction from time to time.

7.3.1 Special case: Anchored cell

Among the different cells' behaviour, a curious phenomenon happened in some cases, precisely in the following video

- **Video01.NE.Cell1**
- **Video02.NE.Cell1**
- **Video04.NE.Cell1**
- **Video04.SE.Cell1**

The name given to this behaviour is "cell anchoring": for some unknown reason, the wandering cell in the well could nail itself in a specific point and get stuck for a while. The resulting shape of the cell is similar to a pointy island, while it tries to move around, it cannot properly escape due

to a fixed point. Eventually, the living cell would manage to free itself. As one may notice, most of the cases happened in the same well, the **SE** one, and the same point, leading to the possible explanation: that the extracellular matrix in that zone may have been different, causing the cell to be stuck for a while. This phenomenon is yet to be understood due to the lack of data. In the last video,**Video04.SE.Cell1**, the "anchoring" occurs with the nailed point located on the wall. This is a singular yet fundamental instance in which the anchoring in the wall is captured. However, it is impossible to give a general explanation as to why that may arise..

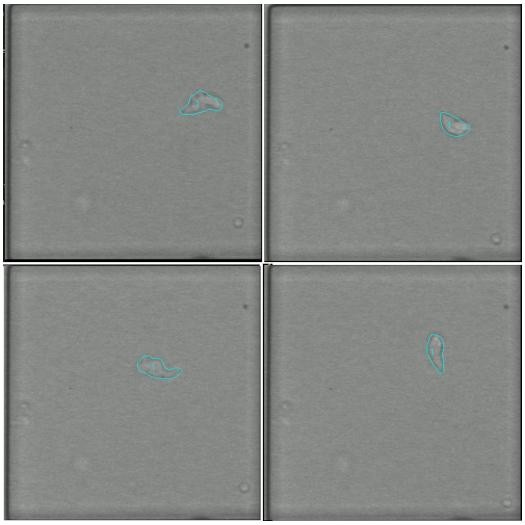

Figure 7.4: Video01.NE.Cell14 different frames. Attempt to show the anchoring. Top left frame 101; top right frame 122; bottom left frame 210; bottom right frame 227

7.4 Angles

As the acquisition of cells' centroids is done, the evaluation[2] for the angles is immediate through the formula 6.6 in the Chapter 6.

By plotting all the angles over the frequency of every data, one can notice a uniform distribution, with a slight preference for going to the right, see figure 7.5.

[2]Even though the formula to get the angles involves the speed, the latter has a simple formula as the displacement of positions. By putting such a definition into the angles', one can get the direct dependence of the angle with respect to the position.

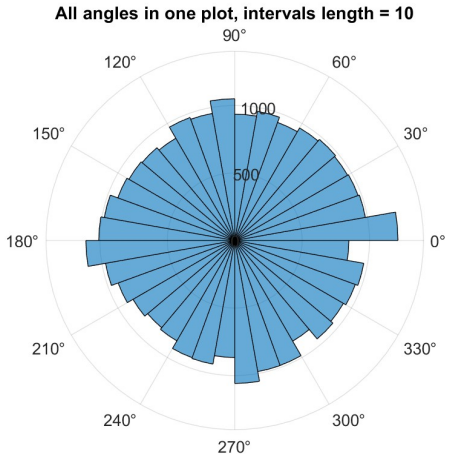

Figure 7.5: Histogram of all directions of **Cell1**

However, the previous result does not reflect on the single **Cell1**'s preference of directions. There are cases in the "good" sample, which has over a thousand frames, where the angles are on average uniformly distributed, and there are other cases, like in **Video02.SW.Cell1** where the cell prefer to go along the main directions (top, bottom, left and right). In the "bad" data, most of the plot has a preferential direction, while there are some cases of even distribution. These observations, from both data types, consolidate the hypothesis that the cell's motility is not a simple random walk, otherwise the distribution of angles should be generally evenly distributed. In the cases of referral directions, the cell is located close to the wall with a physical constraint, limiting its range of motility (See Fig. 7.2, top left).

7.5 Speed

The next statistic of interest is the speed of the cells. By looking at the plot of all speeds in a histogram, a clearly shaped graph appears
leading to the consideration of two candidates to fit such graph: the gamma distribution and the log-normal distribution. However, neither the former nor the latter case seem to fit the re-scaled data by using the mean ($\mu = 20$) and variance ($\sigma^2 = 144$) values estimated from the speed distribution.ù (Fig. 7.8). The log-normal is also often present in biological processes that are asymmetric and we were somewhat surprised by its absence in this case. According to Section 3.3, another possible candidate can be a bimodal motion, where the superposition of two distributions occurs. A

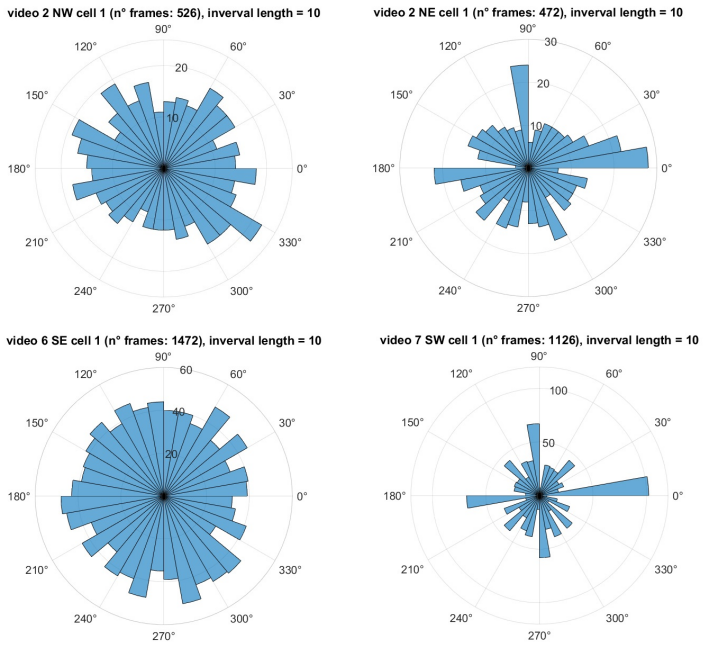

Figure 7.6: Four plots of angles: on the top left **Video02.NW.Cell1**, on the top right **Video02.NE.Cell1**, on the bottom left **Video06.SE.Cell1** and on the bottom right **Video07.SW.Cell1**. Interval length indicates the length of each interval. In this case, since the interval length is 10, there are 36 intervals in the [0 360] degree graph

non-rigorous modus operandi has been considered to find such two distributions, which was trying to find the suitable parameters that would fit the real data. For the slow speed, a normal distribution seems to be fine while for the heavy tail, an attempt to find the power law has been tried as one may see in Fig. 7.9. Even though the parameters were not estimated but somewhat guessed, the graph still showed a promising result that T-cell may present a type of power law distribution, according to the Lévy Flight Paradigm.

In the case of the histograms of the single **Cell1**, all the distribution's shape does seem to resemble 7.7. There are some singular instances of speed exceeding over 30 $\frac{frames}{timestep}$ (see figure 7.7), bringing the hypothesis that cell may have a heterogeneous nature: every single cell is different from the others in term of motility. Such a statement is verified by attempting to calculate the average speed of Cell1s and plotting the resulting numbers over their life's span (figure 7.10).
By looking at the "good" sample there seems to be a correlation between average speed and the life of cells. The resulting hypothesis of cell hetero-

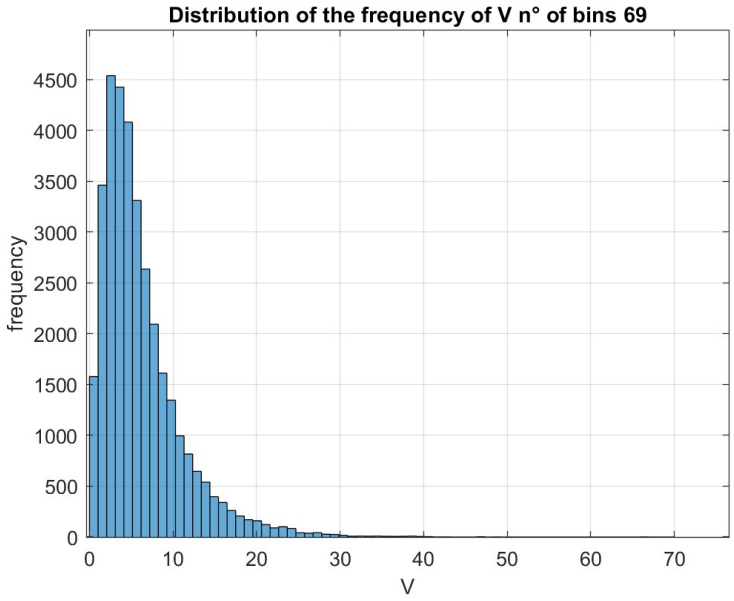

Figure 7.7: Histogram of all speed in one plot

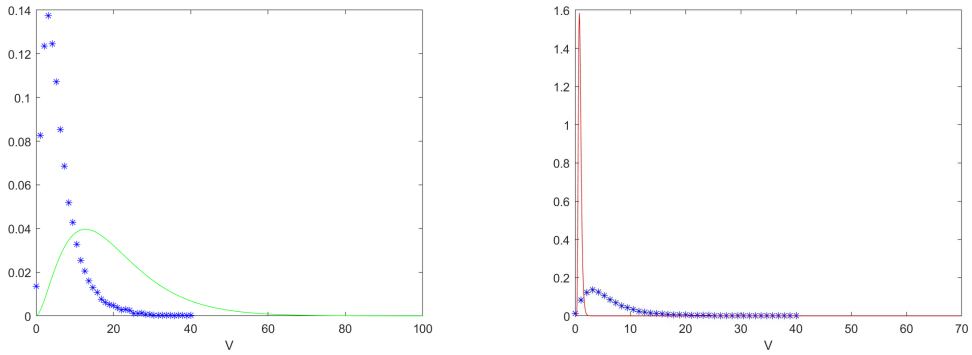

Figure 7.8: Comparison of the real data of all speeds in one plot (blue asterisks) with the corresponding gamma distribution on the left (green line) and lognormal distribution on the right (red line). Neither works well

geneity may be invalid since all the data have similar distribution shapes. Furthermore, the longer cells can live, the more it is likely to pick a faster speed from the pool. On the other hand, the "bad" sample gives a different scenario: similar lengths of recorded cell life have different average speeds. Two important considerations follow:

- Human intervention may interfere with cell's mobility and cells do

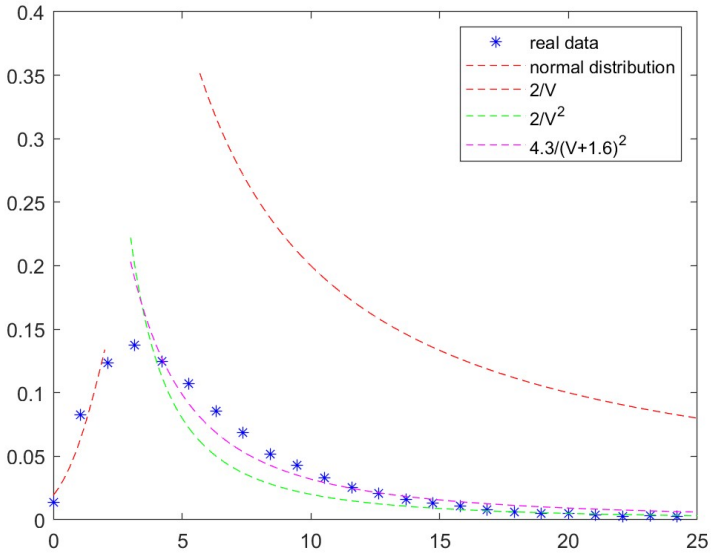

Figure 7.9: Speed distribution of all cells in one plot (blue asterisks). The red dotted line is part of a normal distribution, the green, other red and magenta dotted lines are power-laws.

remember sudden changes. As a consequence, they might change their motility (in the Section 7.5.1, such hypothesis will be discussed);

- After the activation of T-cells, the expansion phase began. The T-cells are induced to divide and eliminate the present pathogens in a certain area. For the immune system to be efficient, after the first division, the lymphocytes have a shorter life-span to guarantee a more prolific division [22] and it should be faster to reach the pathogen or the antigen;

The latter observation, which has a biological foundation, consolidates the heterogeneous hypothesis, and it is also reinforced by the presence of non-proper Cell1 in the **Cell1**'s pool: the lineage of the videos is still unknown and since the re-usage of T-cells from another batch of proliferated cell in a well is more convenient than starting to induce a generation 1 T-Cell1, the non-proper Cell1 is likely the majority of the data and the very few ancestors are in the mix. This makes it impossible for us to validate our ideas.

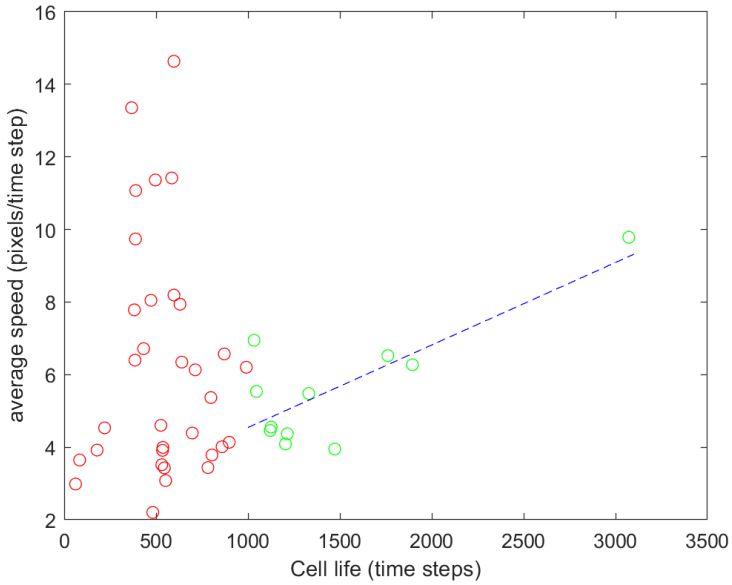

Figure 7.10: Plot of average speed over **Cell1**s' life-span. When looking at the "good" data (green dots), there seems to be a correlation between average speed and Cell's life (blue dashed line). The "bad" data (red dots), does not present the same linear relations.

7.5.1 Auto-correlation speed

In the plot of the graph of the auto-correlation speed coefficient of every single cell, whether they are "good" data or not, there is a common denominator: the decay from lag 0 to lag 1 is quite "rapid", leading to a necessity of a much higher frequency of images acquisition, otherwise the system could be misinterpreted as memoryless, which is high unlikely due to the presence of some persistency of paths, as one can see in the figure 7.11. note that the speed of decay of the autocorrelation function is related to the persistence factor P in eq 3.8 and an "instantaneous" decay would imply $P = 0$.

Furthermore, it seems like there is a type of non-exponential decay in the log-log plot of the auto-correlation and time lag of the single cell very similar to the result of using the fractional Klein-Kramer equation in Ref. [9].

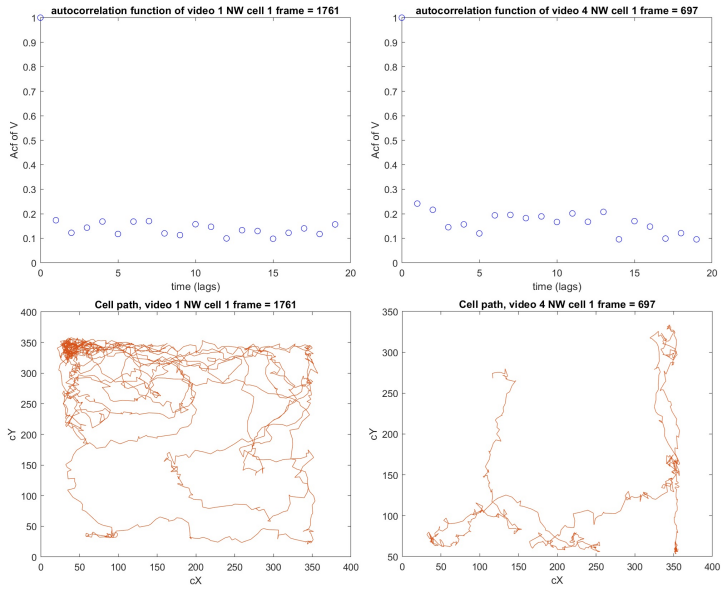

Figure 7.11: Plot of the auto-correlation coefficient of speed in the first 20 time lags. An instance of a "good" sample on the left column (**video01.NW.Cell1**) and an instance of "bad" data on the right column (**video04.NW.Cell1**).

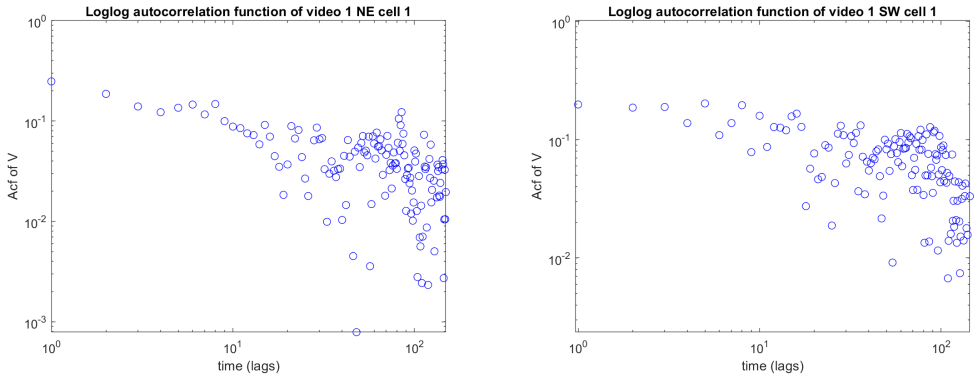

Figure 7.12: Plot of auto-correlation function over time lags. On the left, we have an instance of "good" data. On the right, there is an instance of "bad" data. Both of them seem to have the same type of decay

7.6 Average life span of Cell1

Since information on the cell in the first generation's life is not fully available, a different approach can be considered. With some successful tests of Deepkymotracker, a few videos have been processed until the third generation, in particular:

- **Video02.NE**
- **Video02.NW**
- **Video02.SE**
- **Video07.SW**

With these considered, the average number of frames is 1163. Furthermore, due to the biological need to fight pathogens, it is logical to propose that the expectancy life of T-cells in the first generation is longer than its offspring. However, since the complete picture of the cells lineage is not available (for instance **Video04.SW.Cell1** may be the **Video01.SW.Cell100**'s offspring) and information of the first generation in each microwell is not available, the best approach is to set a lower thresh-hold of **Video01.SW.Cell1**'s life as 1163 average frames and filter all the individual **Cell1**'s frame with such number. As a result, the average first-generation T-cell's life is represented by the average of all the filtered **Cell1**'s frames. The average number of frames for the first generation is estimated to be 1707. From now on, the term Cell1 refers to the true first generation found by the filtering, while **Cell1** refers to all the data collected and labelled as the first generation.

7.7 Average-time mitosis

When analysing the videos of the cells in the well, their shape is very round right before they split into two daughter cells. The unknown biological parameters during the cell's division are phenotypical and seen in the shape of the living cell, which is why there is an attempt to estimate the average time of cells' mitosis through the data of circularity. The method to determine the length of mitosis' duration is to count the number of circularities above a certain threshold $C_{threshold}$ with a finite number of opportunities to be below and not interrupt the count. It $C_{threshold}$ is set to be 0.85. The average mitosis time of every cell is in table 1. As a result, the average time of mitosis based on circularity is the mean value of all cells' mitosis time which is 60.

7.8 MSD - Mean Square Displacement

According to Chapter 6, there are two ways to calculate the Mean Square Displacement, one according to Einstein (Ref. [11]) and one according to

Nordlund(Ref. [30]). For the sake of simplicity, the $\langle MSD \rangle$ will indicate the distributional mean square displacement while the MSD will indicate the single time-averaged mean squared displacement.

After the selection of **Cell1** in section 7.6 and assuming that there is no heterogeneity among cells, the distributional $\langle MSD \rangle$ is found by using the formula 6.26. However, there are two major problems:

- the number of frames is different;
- the starting point of the cell is very likely not to be available due to the procedure of restarting a new experiment;

As a consequence, the possible power-law of the $\langle MSD \rangle$ is calculated by assuming that the starting point of their life is the starting point in the data and the maximum time considered is the minimum of all the involved **Cell1**'s life-span.

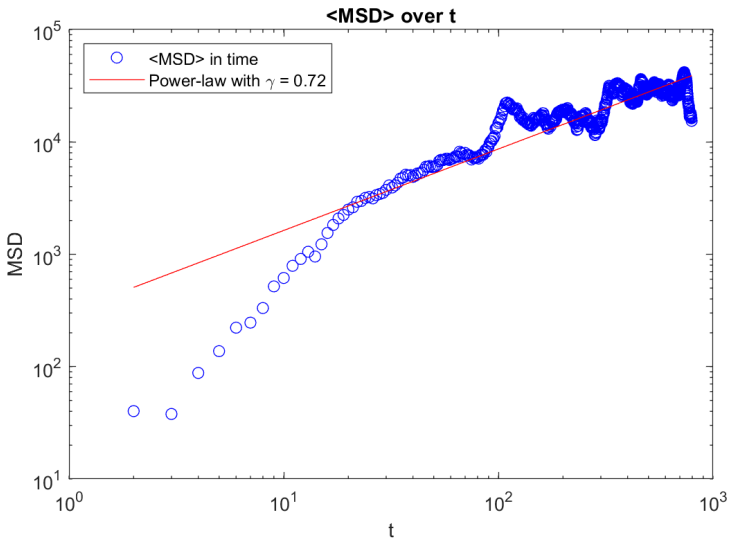

Figure 7.13: $\langle MSD \rangle$ of the chosen **Cell1**s with the eq. 6.26

The power-law obtained is $\gamma = 0.72$ which indicates a subdiffusive behaviour of the cells at the beginning, the formula used to calculate the first region of MSD is 6.26. Unfortunately, the restriction of studying the standard $\langle MSD \rangle$ has to be limited to such time steps, due to the different cells' life duration. As a result, the estimation of the mean square displacement is quantitatively incorrect. By looking at the single MSD of each cell according to Nordlund's formula Eq. 6.24, another way to calculate any

power-laws can be attempted without discarding any frames. After filtering the very "bad" data, which have frames under two hundred, some of the MSD does have a power-law above $\gamma > 1$, which may indicate that some **Cell1** are not activated by a dendritic cell, but are the offspring involved in the expansion phase, requiring faster motility to eradicate the pathogen. The data, unfortunately, is not completely clear.

According to the experiment, we expect more offspring of dendritic-induced active Cell1s in the **Cell1**'s pool. However, until the cell's lineage is reconstructed, one may assume one hypothesis, regardless of the lineage: there are two classes of cells, the "fast" class and the "slow" class. The **Cell1** are classified in such categories based on the power-law exponent *gamma*: if it is less than 1 (γ<1) then it is considered slow and if it is more than 1 (γ>1) then the cell is considered fast. Interestingly, this classification does not lead to a correlation between the class of cells and its average speed, for instance, a fast cell could not have a high average speed. This may be counter-intuitive, but the average speed of a fast cell can be low and some slower cells may have high average speed (with the filtration of very "bad" data). This is because, along the MSD over t in the log-log graph, there are some changes of convexity, modifying the slope of the line obtained through linear regression of the whole MSD. If the MSD tends to go "down" in the log-log plot after a change of convexity, biologically this represents that the cell is going toward the directions of the starting point on average. On the other hand, a local increase in MSD represents the distancing of the cell from the starting point on average. The speed impacts only on the local slope between changes of convexity. A cell might have very high speeds when returning to its starting position, for instance right before the mitosis process (See Fig. 7.14), which leads to a very steep negative slope for the MSD (in the log-log form), contributing to lower the *gamma* in the overall search of the power law.

The stagnation of the cell is guaranteed to happen at the end of its lifespan when the splitting occurs. In the MSD, such phenomena can be seen as a drop or increase of the MSD graph in the log-log plot in the last frames. To appreciate the subdiffusive nature of the phenomena the following is proposed: after determining the time for the mitosis to happen in Section 7.7, the new $\langle MSD \rangle$ of just the splitting phase will be calculated with the new starting point of the cell to be the first time frame of mitosis, the length of the MSD taken is the minimum of the mitosis times among the first-generations Cell1. The gamma found is $\gamma = 0.45$, which hints at the sub-diffusiveness nature of the mitosis phase as one may see in Fig. 7.15.

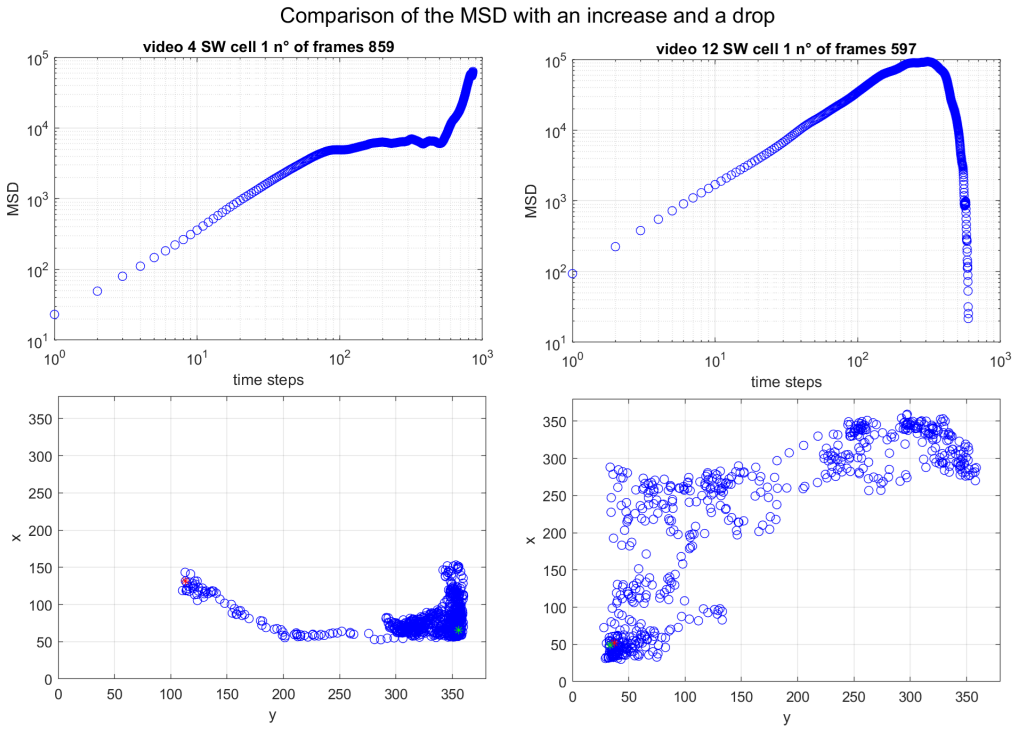

Figure 7.14: An instance of an increasing slope during the mitosis on the top-left. On the bottom-left, the corresponding path is illustrated. An instance of a drop of slope during the mitosis on the top-right. On the bottom right, the corresponding path is illustrated.

Before going to the next chapter, a sum-up of the results follows:

- The cells would "split" themselves very close to the wall;
- There are no preferential directions for the cells if they do not stay too long close to the wall;
- The frequency of frame capturing needs to be higher to understand the type of memory the cells' path has;
- There are two main phases in the **Cell1**s' life: a wandering and the splitting phase. In the former case, the local power law of the MSD (according to Nordlung) may be super- or sub-diffusive, hinting at the classification of the cells into two major classes, the fast and slow cells. On the other hand, during the mitosis, the resulting position-shifted $\langle MSD \rangle$ is sub-diffusive;

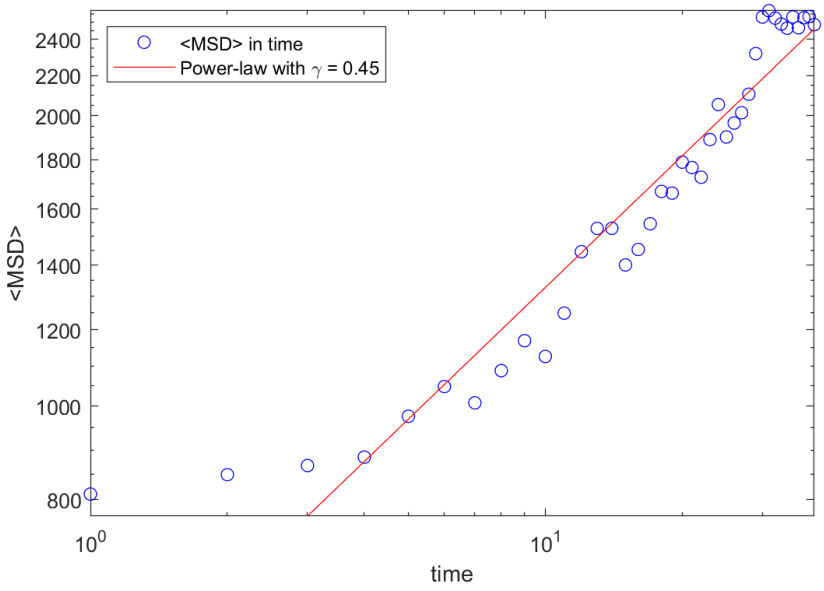

Figure 7.15: $\langle MSD \rangle$ of the mitosis phase. The initial point starts at the position where the mitosis is calculated to start.

Such observations will be the key to simulating some simple models in the next chapter to take a step further in understanding T-cell motility.

Chapter 8

Modelling some behaviours seem in experiments

This Chapter is divided into two main parts:

- The first part is the simulation of a variation of the random walk in a rectangular and hexagonal 2D lattice, where the probability and the step length are chosen from the real data and implemented in Eq. 6.8[1] according to Subsection 6.4.1;
- The second part is the simulation of the standard random walks in a rectangular 2D lattice and the analysis of their statistics with a specific sampling and different conditions along the walls;

All the simulations were coded in the software Matlab 2023a and the specifics are found in the Appendix.

8.1 Random Walk with real data

To verify if the simple models can be suitable candidates to describe the real data, a first simulation of them in the form of a non-standard random walk has been first implemented, the steps are the following:

1. A few suitable **Cell1**s has been selected, specifically we picked the proper Cell1s from Section 7.6 and three **Cell1** that has a super-diffusive MSD;

[1]This is true for the rectangular lattice, while for the hexagonal one, the formula has more terms since the neighbourhood is different, but the logic of the mass conservation is the same.

2. the vectors θ and V have been used to find the probability distribution of directions and step length of each case for a 2D rectangular and hexagonal 2D lattice, according to Subsection 6.4.1. In the case of the square 2D lattice, the direction probability is given by:

$$p_{\text{East}} = \frac{n(]-\frac{\pi}{4},\frac{\pi}{4}])}{n(]0,2\pi])} \qquad p_{\text{North}} = \frac{n(]\frac{\pi}{4},\frac{3\pi}{4}])}{n(]0,2\pi])}$$

$$p_{\text{West}} = \frac{n(]\frac{3\pi}{4},\frac{5\pi}{4}])}{n(]0,2\pi])} \qquad p_{\text{South}} = \frac{n(]\frac{5\pi}{4},\frac{7\pi}{4}])}{n(]0,2\pi])}$$

Where $n(]a,b])$ is the function that counts the number of angles from a to b of the vector θ^*.

The time of each simulation is based on the average lifespan of **Cell1** estimated in Section7.6, which is 1707 for the proper Cell1 and 1163 for the others. For each selected **Cell1**, a simulation of one hundred RWs has been computed.
For each non-standard random walk, statistics like positions, speed and MSD have been analysed and plotted. No physical restriction has been implemented in the simulations since the wall's influence component is already present in the speed of real data.

Between the rectangular and hexagonal configurations, no particular features were found to be different, besides the approximation of directions from four main directions to six.

The $\langle MSD \rangle$ is found for each Video through Eq. 6.26 when each cell track is one of the hundred RWs. Most of the Video's simulation has the $\langle MSD \rangle$ with a power law greater than one, $\gamma < 1$, while the rest has the γ slightly under one. This was expected qualitatively in the case of the last three Videos of fast cells, but it was not a result for the proper Cell1. The major reason may be the limitation of freedom the cell has around the wall. It would be impossible to perform a long step length toward the wall when the cell cannot proceed beyond. This implies that the position and speed may be correlated: whenever the cell is close to the wall it is more probable to have a short step length. This leads to the fact that more realistic simulation models must consider space constraints. Quantitatively, the γs of $\langle MSD \rangle$s found in these simulations are always greater than the time-averaged MSD found in Section 7.8. Another possible reason that may have contributed to aumenting the γ is the approximation of the model in a 2D regular lattice in combination with the wall restriction since each Video has its history of directions, as we can see in Fig. 8.2 and in Fig. 8.1.

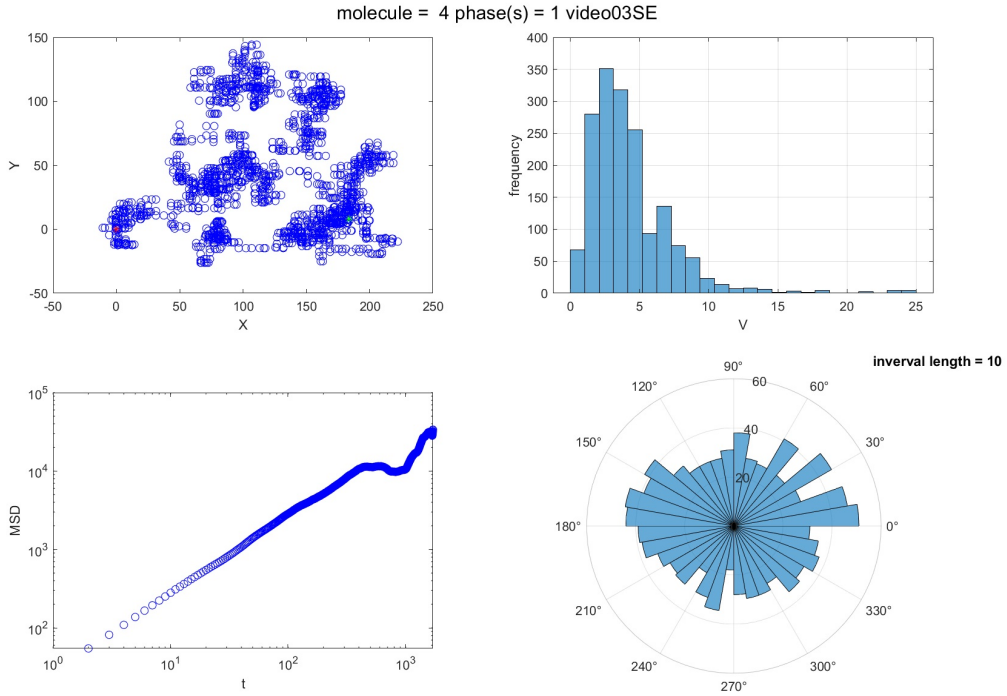

Figure 8.1: 4 plots of the simulation in a 2D rectangular lattice of **video03SE.Cell1** with a sampling = 25. **video03SE.Cell1** is classified as a proper Cell1. Top-left: plot of the path, the red dot is the starting point and the green dot is the ending point. Top-right: speed distribution. Bottom-left: plot of the MSD over t in the double logarithmic plot. Bottom-right: distribution of the angles from the real data.

Lastly, the simulations may have revealed the possible true nature of the cell motility, which should be super-diffusive in most cases, even for the proper Cell1, with a quantitative difference from their offspring.

To see how the numerical results in the $\langle MSD \rangle$ during the mitosis time may still be sub-diffusive, an adjustment has been implemented in the simulation, which is the differentiation of two-speed distributions: the first distribution for the wandering phase and the second distribution for the splitting phase. To achieve that, the estimation of mitosis time has been considered. The resulting $\langle MSD \rangle$ has a power law greater one, an incorrect prediction of what it was found in Section 7.8, for both types of cells, proper Cell1s and fast cells. The possible reasoning is again the absence of space constraint when mitosis occurs. In fact, according to Section 7.3, the cell divides very close to the wall.

By looking at the speed distribution during mitosis in Fig. 8.5 and Fig

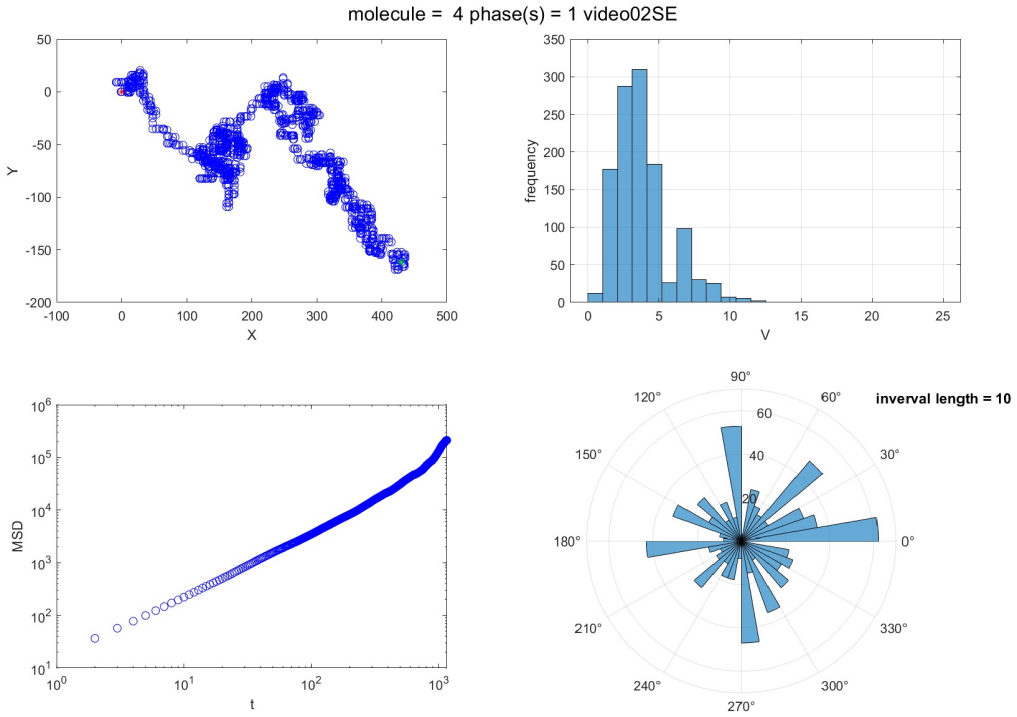

Figure 8.2: 4 plots of the simulation in a 2D rectangular lattice of **video02SE.Cell1** with a sampling = 25. **video02SE.Cell1** is classified as a fast cell. Top-left: plot of the path, the red dot is the starting point and the green dot is the ending point. Top-right: speed distribution. Bottom-left: MSD over t in the double logarithmic plot. Bottom-right: distribution of the angles from the real data.

8.4, most recorded speeds rarely exceed 10, which is expected since the splitting is the main focus of the cell.

8.2 Random Walks - α models

In this section, we will discuss the random walk and verify if it is suitable to describe the cells' movement. The simulation of the random walk is performed in a 2D square lattice.

To mimic the step length found in Section 7.5 with a unitary step-length random walk in a 2D lattice, a sample parameter has been implemented to decide the frequency of picking the elements from the cell track set T. For example, after the simulation of RW, the cell track $\hat{T}$ is produced and, if the sampling frequency is k, the final Cell track T sequence would be the sub-sequence of $\hat{T}$ by taking the i-th elements as multiple of k. The

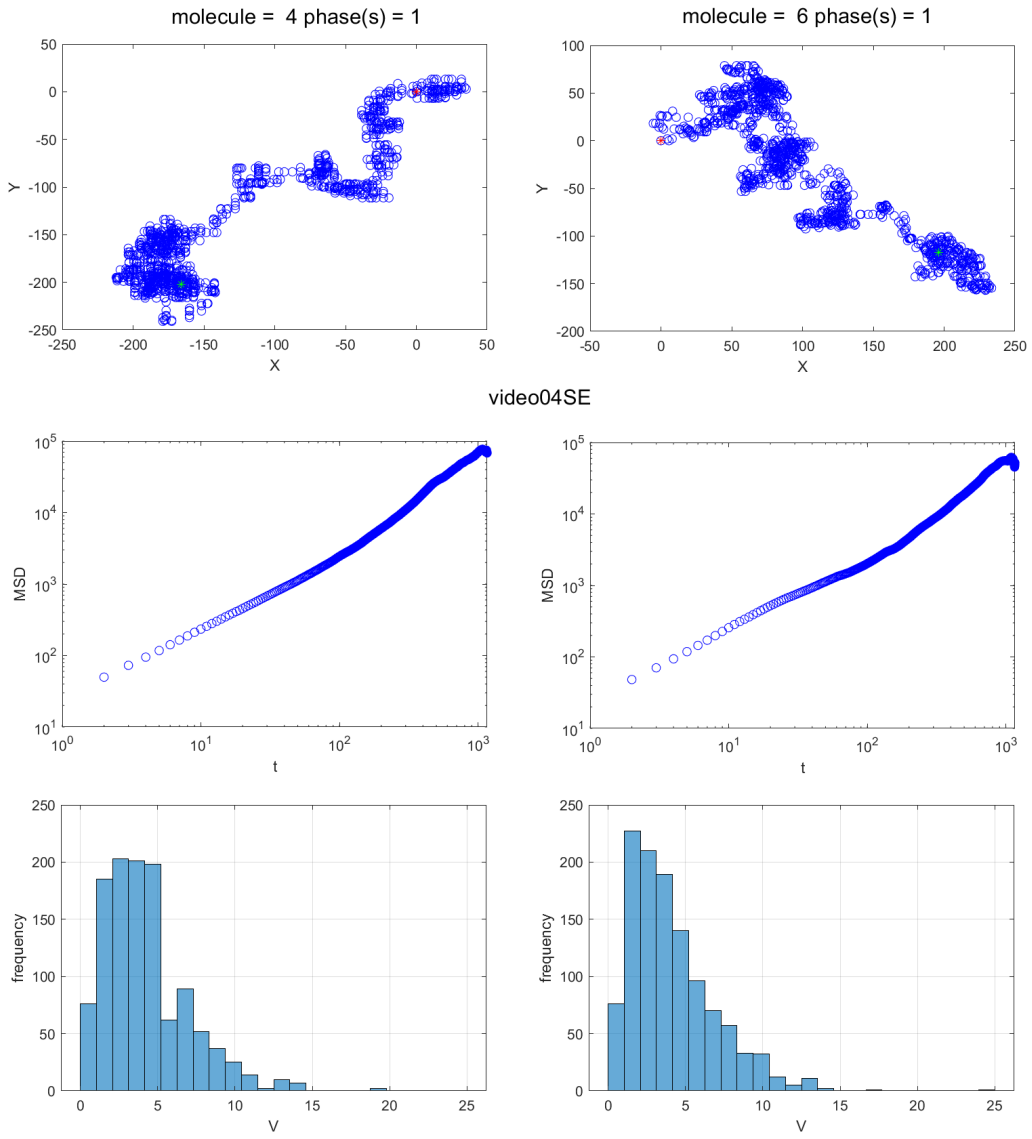

Figure 8.3: Comparison of simulation between the two cells of **video04SE.Cell1**'s directions. On the top are the plots of paths. In the middle is the MSD over t in the double logarithmic plots. At the bottom are the speed distributions.

sampling frequency chosen was twenty-five (sampling = 25) since such a parameter is also an upper limit to the maximum speed that can occur in one step length. Furthermore, since the importance of space constraints in the models according to Section 8.1, we performed two types of simulations: unrestricted and restricted RW. Each unrestricted random walk has no space constraints while the restricted one is simulated in a 380×380

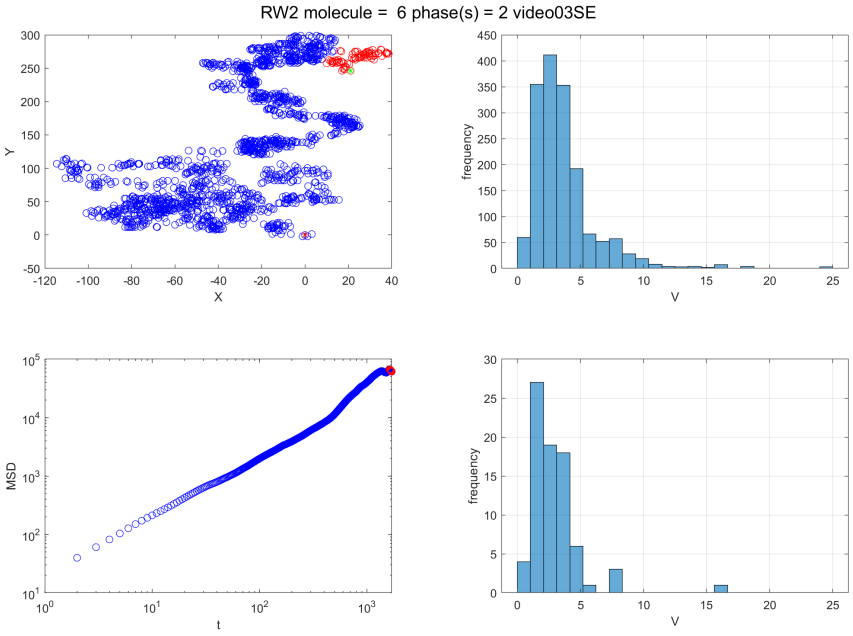

Figure 8.4: 4 plots of the simulation in a 2D hexagonal lattice of **video03SE.Cell1** with a sampling = 25. **video03SE.Cell1** is classified as a proper Cell1. Top-left: plot of the path, the red dot is the starting point, the red circles are the position of mitosis, and the green dot is the ending point. Top-right: speed distribution of the wandering phase. Bottom-left: plot of the MSD over t in the double logarithmic, where the red circles represent the mitosis. Bottom-right: speed distribution during the mitosis phase.

pixels^2 square space.

8.2.1 Unrestricted model

Each random walk has a cell track, in the sense of the model, with a length equal to 1707. A total of one hundred random walks have been simulated. In Figure 8.6, the distribution of speed doesn't show any speed close to the upper limit, since the probability of occurring to a step length equal to twenty-five is very low (in the case of 4 possible actions, the probability would be $p = 4 * \left(\frac{1}{4}\right)^{25} = \left(\frac{1}{4}\right)^{24} \approx 3.55 * 10^{-15}$). The $\langle MSD \rangle$ is also calculated and the power law with respect to the time is very close to one $\gamma = 1$, as expected from the theory.

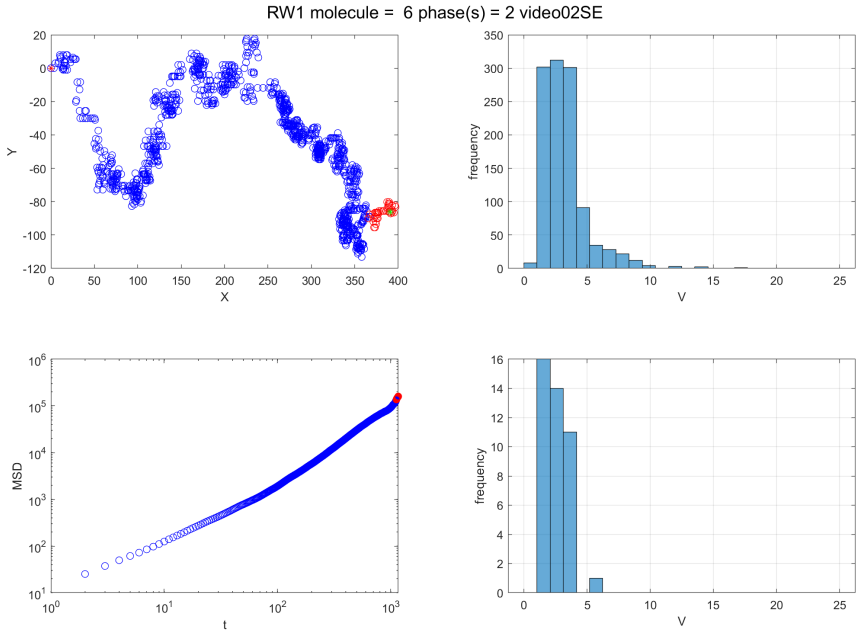

Figure 8.5: Four plots of the simulation in a 2D hexagonal lattice of **video03SE.Cell1** with a sampling = 25. **video02SE.Cell1** is classified as a fast cell. Top-left: c. Top-right: speed distribution of the wandering phase. Bottom-left: plot of the MSD over t in the double logarithmic, where the red circles represent the mitosis. Bottom-right: speed distribution during the mitosis phase.

8.2.2 Restricted models

When simulating a two-dimensional random walk in a finite square, different ideas may come to mind to model the walker when it comes close to the wall. In this thesis, three are the main concepts that will be used:

1. Equiprobability of actions;

2. Stillness on walls;

3. Moving on walls;

In the first case, when the cells are moving along the lattice, every action has an equal probability of occurring:

In the second case, the "stillness" refers to the fact that the probability of an action, if it is not feasible, would merge with the probability of staying still. Such a model mimics the possibility of the biological behaviour of

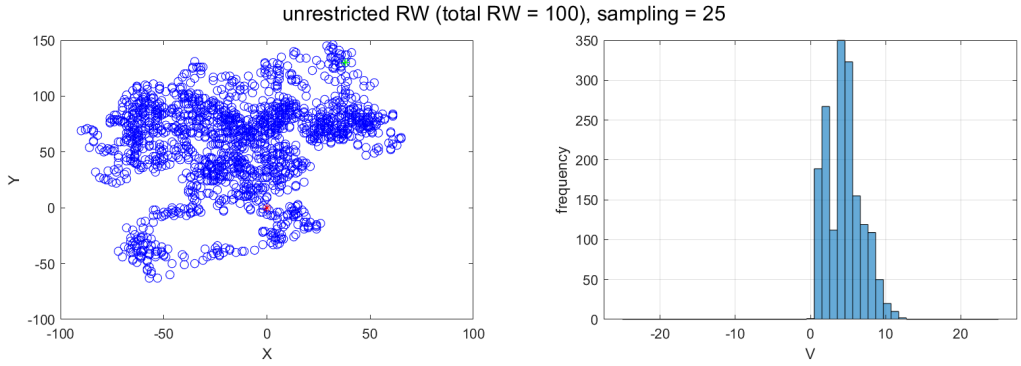

Figure 8.6: Simulation of a classic random walk with no physical constraint with a sampling = 25. On the left: plot of the path, the red dot is the starting point and the green dot is the ending point. On the right: speed distribution.

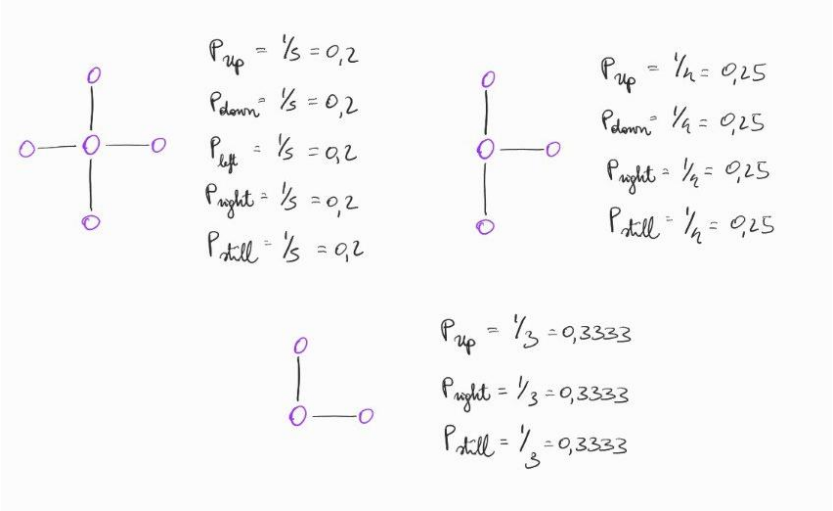

Figure 8.7: Types of movement for the restricted RW model

the cells, whenever it is close to the wall, analysing the physical obstacle before moving.

The last case, the opposite of the second case, takes the probability of unfeasible actions and equally distributes them to the two probabilities of moving along the wall. Such a model mimics the cells' awareness of the physical obstacle and they are more likely to wander around it.

A parameter of stillness α is implemented in the last two models, which controls the probability of staying still. That is why the last two are called α-models. For the simulations, one hundred RWs have been analysed for different values of α. The sampling frequency of choice is 25, the time

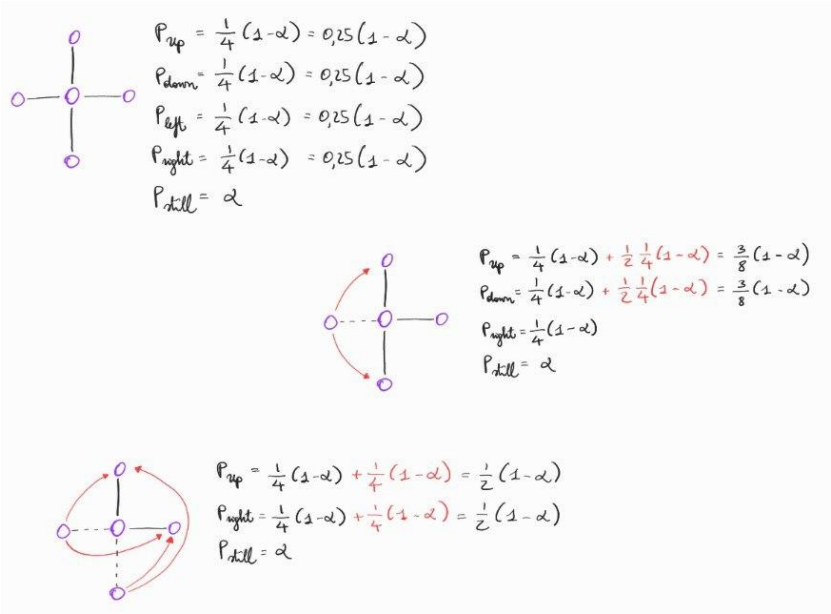

Figure 8.8: Types of movement for the stillness model

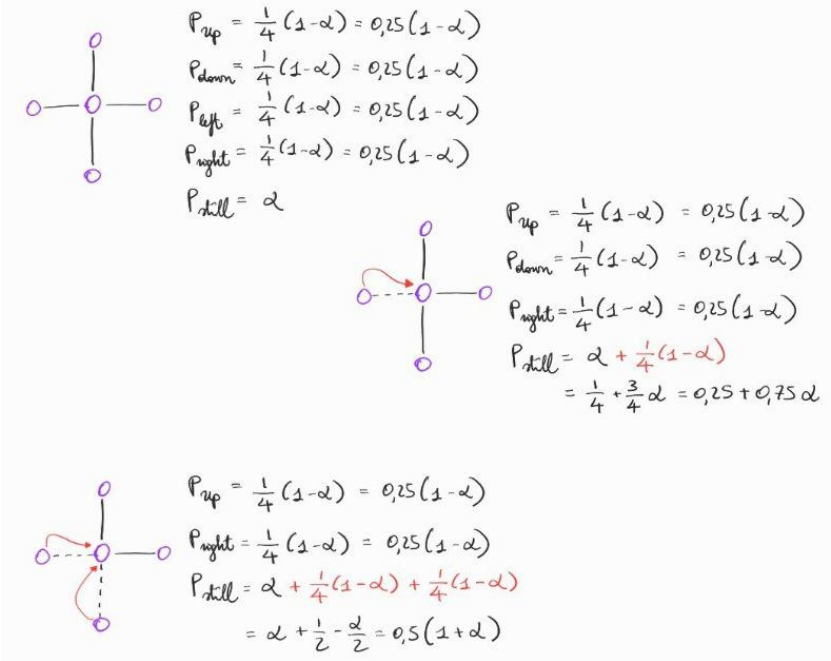

Figure 8.9: Types of movement for the moving model

steps are set to be 1707 and the starting points of the random walks are randomly chosen from the starting point of the real data (see the red dots

in figure 7.2).

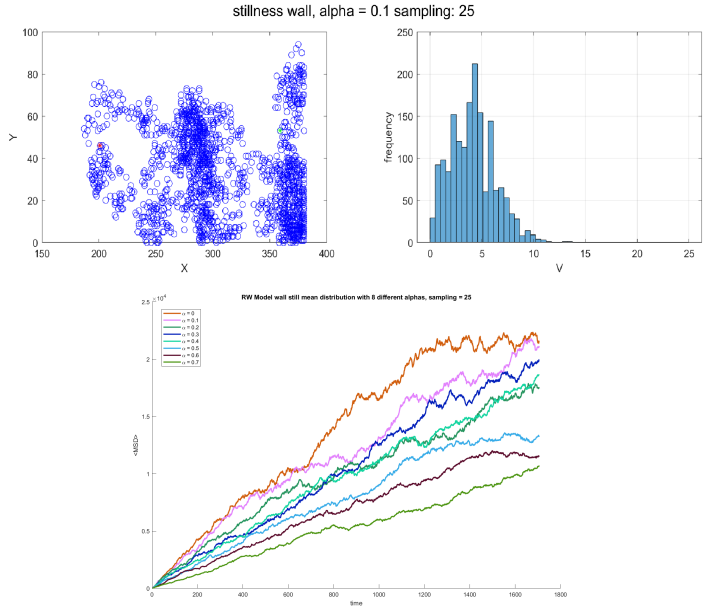

Figure 8.10: Simulation of an RW with the stillness model on the wall. On top: two plots, the first is the path in the confined space, the red dot is the starting point and the green dot is the ending point, and speed distribution. At the bottom, $\langle MSD \rangle$ over t of eight different αs in the double logarithm plots.

As one may predict, the resulting gamma from these models can only be less than one since the physical constraints limit the diffusion if the time passed is long enough. As a consequence, these cannot be fit to represent all the real data inside the wall, since some of them showed a super-diffusive behaviour. Furthermore, these models don't differ much from one other, due to the short life-span set for each simulation, and the walker visiting the wall very few times. As a result, the decreasing γs seems to occur with an increasing α (See table 4). Nonetheless, the $\langle MSD \rangle$ found in Section 7.8 for the proper Cell1 has a subdiffusive nature and can be reproduced qualitatively in the α for the "stillness" model and the "moving" model.

To sum up, we conclude that the wall plays a major role in the simulation since its presence would correlate with the position and the speed of cells, while in Section 8.1 the information was already implicitly in the speed distribution. Still, no walls have been placed in the model. Even though the α models proposed are not suitable candidates to describe cell

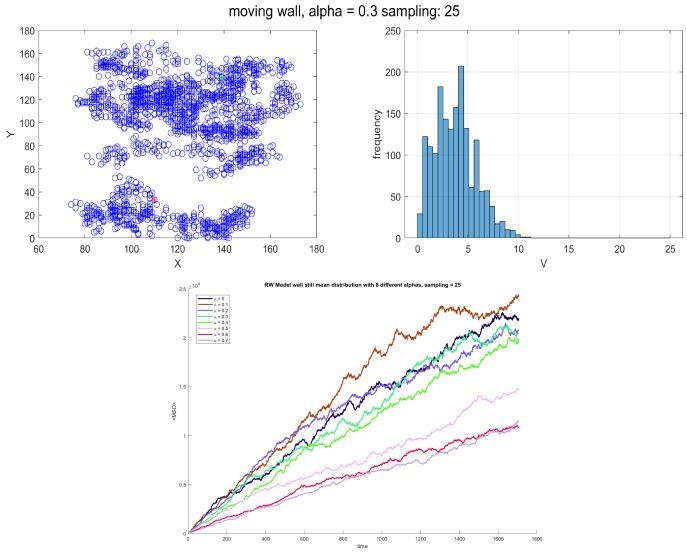

Figure 8.11: Simulation of an RW with the moving model on the wall. On top: two plots, the first is the path in the confined space, the red dot is the starting point and the green dot is the ending point, and speed distribution. At the bottom, $\langle MSD \rangle$ over t of eight different αs in the double logarithm plots.

motility, some statistics are still relevant, hinting at the possibility that a proper modification of such models may lead to the correct one. So far, no single model seems to be able to reproduce all the different features we observed on the experimental Data or motility.

Chapter 9

Conclusion and future ideas

The thesis started with the types of literature review: the first was related to the necessary knowledge for data acquisition, while the second was cell-motility-related. Different complications were encountered during the work, including extended time to test the DeepKymoTracker program. Mathematical tools necessary for the analysis of statistics were presented in one chapter and analysed in the next. Relevant results found as follows:

- Cells split themselves close to walls;
- The speed distribution of all the cells combined seems to be the superposition of two distributions, satisfying the diluted Levy Hypothesis;
- No ergodicity was found, with high variability in MSDs;
- There are two classes of cells, depending on their γs. Those that have a sub-diffusive behaviour ($\gamma < 1$) are the proper Cell1s, while the other class have a super-diffusive behaviour ($\gamma > 1$) which are the Cell1s' offspring, but not enough data is available;
- The cells might all have a higher γ when the wall is non present;
- Subdiffusivity can be achieved with a simple RW in a confined space;

A few aspects might be addressed and optimized for studying cell motility, in the future:

- In the experimental set-up, a larger well is needed to minimize the wall's effect on cells' behaviour. Furthermore, the starting point of

each cell should be the same. Also, there is the necessity of finding a way to record the cells from the very first frame of their life while being completely isolated, whether there is generation 1 or the offspring;

- In the acquisition of data, DeepKymoTracker should be upgraded to be a more user-friendly program in terms of robustness and time optimization. A new approach needs to be studied and implemented in the program to correctly contour the overlapping cells;
- In the modelling, the persistency term could be implemented in the α-models to generate a super-diffusive behaviour in the walker. Having a walker from a Langevin equation like the one used by Jerison and Quake in Ref. [21] is also a good idea;

Much more can be studied and developed from this type of data. Even though I just scratched the surface of what type of motility T-cells have, those analyses are fundamental for possible future work that will continue after the Thesis, which will hopefully answer the question: *what type of motility does the T-cell have?* At this stage, this question is too difficult. Nevertheless, we now possess good insight into which models and which quantities tell the right story.

Appendix

Matlab and Tables

All the analysis has been processed in the software in Matlab ver. 2023a. All the codes that have been implemented can be found in the additional materials linked to this Thesis. There are two main files, **T-cell** and **RW**. In the former, there are all the codes regarding the analysis of the real data and its statistics, while in the latter, there are all the codes for the simulations of different random walks shown in chapter 8.

	position	lifetime	average speed	γ of TAMSD	mitosis time
01	NE	1895	6.27	0.74	74
01	NW	1761	6.52	-0.06	62
01	SE	713	6.13	0.31	94
01	SW	640	6.34	0.70	71
02	NE	472	8.04	-0.18	57
02	NW	526	4.60	0.87	58
02	SE	804	3.78	1.16	41
02	SW	782	3.43	1.21	126
03	NE	84	3.64	1.28	18
03	NW	62	2.98	1.17	61*
03	SE	1214	4.37	0.82	78
03	SW	1121	4.45	0.36	85
04	NE	798	5.36	0.47	82
04	NW	697	4.38	0.98	92
04	SE	898	4.13	1.20	54
04	SW	859	4.01	0.83	17
05	NE	991	6.20	0.26	33
05	NW	871	6.57	0.03	62
05	SE	531	3.51	-0.01	73
05	SW	537	3.99	0.49	49
06	NE	545	3.43	0.32	23
06	NW	535	3.91	0.99	59
06	SE	1472	3.94	0.25	81
06	SW	1204	4.08	0.23	80
07	NE	552	3.07	0.14	20
07	NW	482	2.20	0.68	119
07	SE	1331	5.48	0.16	70
07	SW	1126	4.55	0.39	65
11	NE	496	11.36	0.57	43
11	NW	585	11.41	0.54	31
11	SE	367	13.35	0.29	28
11	SW	597	14.63	0.47	26
12	NE	432	6.71	0.14	41
12	NW	389	11.07	0.14	34
12	SE	388	9.73	0.10	51
12	SW	597	8.19	0.13	60
16	NE	385	6.39	1.50	63
16	NW	3027	9.78	0.52	39
16	SE	382	7.78	1.16	46
16	SW	630	7.93	0.91	35
41	NE	179	3.92	1.64	126
41	NW	220	4.53	1.56	61
41	SE	1046	5.53	1.04	80
41	SW	1033	6.94	0.60	78

Table 1: Table summarising the results from the real data analysis. The asterisk indicates that the mitosis times coincide with the whole recorded lifespan of the cell. The blues are the "good" samples, the blues are the "good" samples.

	$\langle MSD \rangle$ wandering	$\langle MSD \rangle$ mitosis
01.NE.Cell1 01.Nw.Cell1 03.SE.Cell1 06.SE.Cell1 06.SW.Cell1 16.NW.Cell1 16.NW.Cell1	0.72	0.45

Table 2: Table summarising the distributional MSD for the wandering phase and the MSD for the mitosis. The blues are the "good" samples.

Cell1 chosen for simulations	timespan	# RWs	γ $\langle MSD \rangle$	mitosis time	γ mitosis
01.NE.Cell1	1707	100	0.96	74	1.10
01.Nw.Cell1	1707	100	1.00	62	1.07
03.SE.Cell1	1707	100	1.54	78	1.25
06.SE.Cell1	1707	100	1.07	81	1.05
06.SW.Cell1	1707	100	1.37	80	1.15
07.SE.Cell1	1707	100	1.09	70	1.11
16.NW.Cell1	1707	100	0.91	39	1.11
02.SE.Cell1	1163	100	1.54	41	1.25
04.SE.Cell1	1163	100	1.13	78	1.19
41.SE.Cell1	1163	100	1.22	80	1.13

Table 3: Table summarising the results of choosing the 10 real data for the simulation for the first part of Chapter 8. Simulations of RW in a 2D hexagonal lattice, sampling rate = 25. The blues are the "good" samples, the blues are the "good" samples.

	n° RW	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
Classic RW	100	0.98			
Restricted RW	100	0.97			
Stillness wall	100	0.99	0.97	0.94	0.93
Moving wall	100	0.98	0.97	0.96	0.96

$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$
0.90	0.88	0.86	0.84
0.94	0.88	0.87	0.86

Table 4: Table summarising the value of γs for each set of parameter α. Simulations of RW in a 2D square lattice, sampling rate = 25.

Bibliography

[1] Schneider Caroline A et al. "NIH Image to ImageJ: 25 years of image analysis". In: *Nature* (2017). DOI: https://doi.org/10.1038/nmeth.2089.

[2] Nakano Akihiko. "Spinning-disk Confocal Microscopy — A Cutting-Edge Tool for Imaging of Membrane Traffic". In: *Cell Structure and Function,* (2002). DOI: https://doi.org/10.1247/csf.27.349.

[3] Alberts B et al. *Molecular Biology of the Cell. 4th edition.* Garland Science, 2002.

[4] J. M. Benitez, J. L. Castro, and I. Requena. "Are artificial neural networks black boxes?" In: *IEEE* (1997). DOI: https://doi.org/10.1109/72.623216.

[5] Sally R. M. Bennett et al. "Help for cytotoxic-T-cell responses is mediated by CD40 signalling". In: *Nature* (1998). DOI: https://doi.org/10.1038/30996.

[6] Martin Chalfie et al. "Green Fluorescent Protein as a Marker for Gene Expression". In: *Science* (1994). DOI: https://doi.org/10.1126/science.8303295.

[7] José-Angel Conchello and Jeff W Lichtman. "Optical sectioning microscopy". In: *Nature Methods* (2005). DOI: https://doi.org/10.1038/nmeth815.

[8] P. Davidovits and M. D. Egger. "Scanning Laser Microscope for Biological Investigations". In: *Applied Optics* (1971). DOI: https://doi.org/10.1364/AO.10.001615.

[9] Peter Dieterich et al. "Anomalous dynamics of cell migration". In: *PNAS* (2008). DOI: https://doi.org/10.1073/pnas.0707603105.

[10] Andrew M. Edwards et al. "Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer". In: *Nature* (2007). DOI: https://doi.org/10.1038/nature06199.

[11] A. Einstein and R. furth. "Investigations on the theory of the Brownian movement". In: *Dover* (1956). DOI: https://doi.org/10.2307/3609228.

[13] Khelina Fedorchuk. "Machine Learning in Dynamic Microscopy". Ph.D. Thesis. Swinburne University of Technology, 2022. URL: http://hdl.handle.net/1959.3/469072oai:researchbank.swinburne.edu.au:1d11cb51-5dff-4e2f-a2cc-f10ae00831e7/1.

[14] Megan K. Gautier and Stephen D. Ginsberg. "A method for quantification of vesicular compartments within cells using 3D reconstructed confocal z-stacks: Comparison of ImageJ and Imaris to count early endosomes within basal forebrain cholinergic neurons". In: *Journal of Neuroscience Methods* (2021). DOI: https://doi.org/10.1016/j.jneumeth.2020.109038.

[15] Ben N. G. Giepmans et al. "The Fluorescent Toolbox for Assessing Protein Location and Function". In: *Science* (2006). URL: https://www.jstor.org/stable/3846032.

[16] Dale I Godfrey et al. "The burgeoning family of unconventional T cells". In: *Nature* (2015). DOI: https://doi.org/10.1038/ni.3298.

[18] G. E. HINTON and R. R. SALAKHUTDINOV. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* (2006). DOI: https://doi.org/10.1126/science.1127647.

[19] Nicolas E Humphries et al. "Environmental context explains Lévy and Brownian movement patterns of marine predators". In: *Nature* (2010). DOI: https://doi.org/10.1038/nature09116.

[20] Thomas C Ings and Lars Chittka. "Speed-accuracy tradeoffs and false alarms in bee responses to cryptic predators". In: *Current Biology* (2008). DOI: https://doi.org/10.1016/j.cub.2008.07.074.

[21] Elizabeth R Jerison and Stephen R Quake. "Heterogeneous T cell motility behaviors emerge from a coupling between speed and turning in vivo". In: *Elife* (2020). DOI: https://doi.org/10.7554/elife.53933.

[22] Susan M. Kaech, E. John Wherry, and Rafi Ahmed. "Effector and memory T-cell differentiation: implications for vaccine development". In: *Nature Review Immunology* (2002). DOI: https://doi.org/10.1038/nri778.

[23] Rainer Klages. "Search for Food of Birds, Fish and Insects". In: *DIffusive Spreding in Nature, Technology and Society* (2018). DOI: http://dx.doi.org/10.1007/978-3-319-67798-9_4.

[24] Friedrich Lenz et al. “Spatio-temporal dynamics of bumblebees foraging under predation risk”. In: *Ithaca* (2012). DOI: https://doi.org/10.1103/PhysRevLett.108.098103.

[25] Nicolas Levernier et al. “Inverse Square Lévy Walks are not Optimal Search Strategies for d$\leq$2”. In: *Phys. Rev. Lett.* (2020). DOI: https://doi.org/10.1103/PhysRevLett.124.080601.

[26] Masud Mansuripur. *Classical Optics and Its Applications*. Cambridge University Press, 2011.

[27] Makoto Miyara et al. “Human FoxP3+ regulatory T cells in systemic autoimmune diseases”. In: *Autoimmunity Reviews* (2011). DOI: https://doi.org/10.1016/j.autrev.2011.05.004.

[28] Mehryan Mohri, Afshin Rostamizadeh, and Ammet Talwalkar. *Foundation of Machine Learning second edition*. The MIT press, 2018.

[29] Isha Monga, Karambir Kaur, and Sandeep Kumar Dhanda. “Revisiting hematopoiesis: applications of the bulk and single-cell transcriptomics dissecting transcriptional heterogeneity in hematopoietic stem cells”. In: *Briefings in Functional Genomics* (2022). DOI: https://doi.org/10.1093/bfgp/elac002.

[30] I. Nordlund. “Eine neue Bestimmung der Avogadroschen Konstante aus der Brownschen Bewegung kleiner, in Wasser suspendierten Quecksilberkügelchen”. In: *Zeitschrift für Physikalische Chemie* (1914). DOI: https://doi.org/10.1515/zpch-1914-8703.

[31] Jordan Orchard. “Anomalous diffusion in Polygonal Dilliard channels”. PhD thesis. Politecnico di Torino and Swinburne University of Technology, 2022.

[32] Elvina Parlindungan. “Dissecting fate determination in CD8+ T cells using in vitro population studies and single-cell time lapse microscopy”. Bachelor’s Thesis. The University of Melbourne, 2014.

[33] K. Pearson. “Mathematical Theory of Random Migration”. In: *London, Dulau and co.* (1906). URL: https://archive.org/details/cu31924003051285/mode/2up?q=diffusion.

[34] Kiryl D. Piatkevich et al. “Extended Stokes Shift in Fluorescent Proteins: Chromophore–Protein Interactions in a Near-Infrared TagRFP675 Variant”. In: *Scientific Reports* (2013). DOI: https://doi.org/10.1038/srep01847.

[35] Evdokiia Potolitsyna et al. “Cytoskeletal rearrangement precedes nucleolar remodeling during adipogenesis”. In: *Communication Biology* (2024). DOI: https://doi.org/10.1038/s42003-024-06153-1.

[36] Dr. habil. Rainer Klages, Prof. Dr. Günter Radons, and Prof. Dr. Igor M. Sokolov. *Anomalous Transport: Foundations and Applications*. Wiley-VCH, 2008. DOI: https://doi.org/10.1002/9783527622979.

[37] Borlinghaus Rolf and Kuschel Lioba. “Spectral fluorescence lifetime imaging microscopy: new dimensions with Leica TCS SP5”. In: *Nature* (2006). DOI: http://dx.doi.org/10.1038/nmeth941.

[38] Sarah M Russell. “Determination of T-cell fate by dendritic cells: a new role for asymmetric cell division?” In: *Immunology & Cell Biology* (2008). DOI: https://doi.org/10.1038/icb.2008.24.

[40] Raz Shimoni. “TACTICS for bioimaging informatics and analysis of T Cells”. Ph.D. Thesis. Swinburne University of Technology, 2014. URL: http://hdl.handle.net/1959.3/391517.

[41] Haina Shin and Akiko Iwasaki. “Tissue-resident memory T cells”. In: *Immunological Review: volume 255, Issue 1* (2013). DOI: https://doi.org/10.1111/imr.12087.

[42] Thimothy K. Starr, Stephen C. Jameson, and Kristin A. Hogquist. “Positive and Negative Selection of T Cells”. In: *Annual Review of Immunology* (2003). DOI: https://doi.org/10.1146/annurev.immunol.21.120601.141107.

[43] G. M. Viswanathan et al. “Lévy flight search patterns of wandering albatrosses”. In: *Nature* (1996). DOI: https://doi.org/10.1038/381413a0.

[44] M. Yassin and S. M Russell. “Polarity and asymmetric cell division in the control of lymphocyte fate decisions and function”. In: *current-opinion-in-immunology* (2016). DOI: https://doi.org/10.1016/j.coi.2016.02.004.

Webography

[12] Khelina Fedorchuk. *DeepKymoTracker: A tool for the accurate construction of cell lineage trees for highly motile cells*. 2024. URL: https://github.com/khelina/T-cell-lineages-tracking.

[17] U.S. Department of Health & Human Rights. *Hematopoietic Stem Cells*. 2001. URL: https://web.archive.org/web/20161029020128/https://stemcells.nih.gov/info/2001report/chapter5.htm.

[39] Google Scholar. URL: (https://scholar.google.com/scholar?q=animal+movement+ecology&hl=it&as_sdt=0,5&as_ylo=2023&as_rr=1).