TECHNISCHE UNIVERSITÄT DRESDEN

Bereich Mathematik und Naturwissenschaften Fakultät Mathematik

Mathematical modelling of collective cell decision-making in complex environments

Arnab Barua

Born on: 26th May 1994 in Chinsurah, India

Dissertation

to achieve the academic degree

Doctor of Philosophy (PhD)

First referee
Axel Voigt

Second referee
Michael Meyer-Hermann

Supervisors
Axel Voigt
Haralampos Hatzikirou

Submitted on: 22nd September 2021

Dedicated to my sweetheart *Nature*...

Acknowledgments

While writing this thesis, I admire the support and assistance from my supervisors and friends.

I would like to first thank my best supervisor, friend and teacher of my life, Professor Haralampos Hatzikirou, whose support both from scientific parts and mental health issues is non-describable. Maybe there are no words in English dictionary which I can elucidate. I really enjoyed the scientific discussions over the whiteboard and insightful feedback, which helped me to reorganize and serrated my critical thoughts.

I would also like to thank Professor Axel Voigt and Professor Andreas Deutsch for their kind support at difficult times in TU Dresden, which I can't forget. Their useful discussions for my research and thesis are extremely helpful.

I would like to thank my colleagues both from Braunschweig and Dresden. I would also like to thank Helmholtz Centre for Infection Research, Braunschweig and TU Dresden for their financial support throughout my PhD studies.

Finally, I would not have finished and completed this dissertation without the assistance from my friends and roommates, who furnished beautiful scientific discussions as well as cheerful diversions to unconfined my mind from studies.

Contents

Contents

Statement of authorship

I hereby certify that I have authored this Dissertation entitled *Mathematical modelling of collective cell decision-making in complex environments* independently and without undue assistance from third parties. No other than the resources and references indicated in this thesis have been used. I have marked both literal and accordingly adopted quotations as such. There were no additional persons involved in the intellectual preparation of the present thesis. I am aware that violations of this declaration may lead to subsequent withdrawal of the degree.

Dresden, 22nd September 2021

Arnab Barua
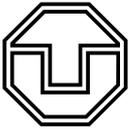
Abstract

Cellular decision-making help cells to infer functionally different phenotypes in response to microenvironmental cues and noise present in the system and the environment, with or without genetic change.

In Cellular Biology, there exists a list of open questions such as, how individual cell decisions influence the dynamics at the population level (an organization of indistinguishable cells) and at the tissue level (a group of nearly identical cells and their corresponding extracellular matrix which simultaneously accomplish a set of biological operations)? As collective cell migration originates from local cellular orientation decisions, can one generate a mathematical model for collective cell migration phenomena without elusive undiscovered biophysical/biochemical mechanisms and further predict the pattern formations which originates inside the collective cell migration? how optimal microenvironmental sensing is related to differentiated tissue at the spatial scale ? How cell sensing radius and total entropy production (which precisely helps us to understand the operating regimes where cells can take decisions about their future fate) is correlated, and how can one understand the limits of sensing radius at robust tissue development ? To partially tackle these sets of questions, the LEUP (Least microEnvironmental Uncertainty Principle) hypothesis has been applied to different biological scenarios.

At first, the LEUP has been enforced to understand the spatio-temporal behavior of a tissue exhibiting phenotypic plasticity (it is a prototype of cell decision-making). Here, two cases have been rigorously studied i.e., migration/resting and migration/proliferation plasticity which underlie the epithelial-mesenchymal transition (EMT) and the Go-or-Grow dichotomy. On the one hand, for the Go-or-Rest plasticity, a bistable switching mechanism between a diffusive (fluid) and an epithelial (solid) tissue phase has been observed from an analogous mean-field approximation which further depends on the sensitivity of the phenotypes to the microenvironment. However, on the other hand, for the Go-or-Grow plasticity, the possibility of Turing pattern formation is inspected for the "solid" tissue phase and its relation to the parameters of the LEUP-driven cell decisions.

Later, LEUP hypothesis has been suggested in the area of collective cell migration such that it can provide a tool for a generative mathematical model of collective migration without precise knowledge about the mechanistic details, where the famous Vicsek model is a special case. In this generative model of collective cell migration, the origin of pattern formation inside collective cell migration has been investigated. Moreover, this hypothesis helps to construct a mathematical model for the collective behavior of spherical *Serratia marcescens* bacteria, where the basic
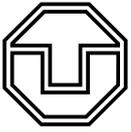
understanding of migration mechanisms remain unknown.

Furthermore, LEUP has been applied to understand tissue robustness, which in turn shows the way how progenitor cell fate decisions are associated with environmental sensing. The regulation of environmental sensing drives the robustness of the spatial and temporal order in which cells are generated towards a fully differentiating tissue, which are verified later with the experimental data. LEUP driven stochastic thermodynamic formalism also shows that the thermodynamic robustness of differentiated tissues depends on cell metabolism, cell sensing properties and the limits of the cell sensing radius, which further ensures the robustness of differentiated tissue spatial order.

Finally, all important results of the thesis have been encapsulated and the extension of the LEUP has been discussed.
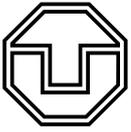
Part I.

Introduction to cell decision-making

1. What is cell decision-making ?

1.1. Introduction

Decision-making is a way of finding optimal alternatives based on some certain objectives. The objectives can depend on multiple factors, and it can be relative as well. In this same way, one can define the decision-making at the cellular level. Cells not only follow their genetic wiring, but also sense their microenvironment, which helps them to make important decisions and change their behavior. So, the information between environmental factors and intrinsic factors (e.g., metabolites concentration, genetic network etc.) are entangled or coupled to each other in a complicated fashion. Cell decision-making is a recipe of changing phenotypes according to the intrinsic programming, noise (extrinsic and intrinsic) and give response to the micro environmental cues [66]. The environmental factors (e.g., ligands, density of cells etc.) attach to the receptors of the cell and triggers a chain of biochemical reactions to give a response to the environment. The phenomena of cell decision-making is a broad topic. The three key factors for cell decision-making are (**I**) internal factors, (**II**) external/microenvironmental factors and (**III**) the noise present in the environment as well as inside the cell. If someone thinks about examples, the processes such as cell differentiation, cell migration, apoptosis, metastasis and more can be thought as a decision-making processes for each individual cell. The cell decision-making phenomena are well studied at the level of single cells. The unknown part is how cells make a decision in a complex, multicellular microenvironment. In the next section the important aspect of cell decision-making in the Phenotypic plasticity, collective cell migration and cell differentiation phenomena at the multicellular level will be discussed.

1.2. Examples of cell decision-making

Here, I am going to present paradigms of cell decision-making that are investigated in this thesis.

1.2.1. Phenotypic Plasticity

Cell differentiation is a way of development from one cell type to the other cell type. As time prolongs, immature or less specialized cells (i.e., stem cell) turned to more mature or specialized cells. So, on one hand cells gained irreversibility in new fates/phenotype by a successive hierarchical descent, where the pluripotent stem cells can be seen in a proper microenvironment (stem cell niche) are turning into e.g., bone, muscle, epithelial and further more specialized cells. But, on the other hand, one can observe a non-genetic plasticity in the phenotype which allows a reversible adaptation to different sets of microenvironmental signals [128]. The plasticity in the phenotype or phenotypic plasticity is a particular process where the particular form of life (e.g., cells, tissues, animals, plants, fungus etc.) switch their actions on the basis of the environment [83]. It plays a huge role in adaptation and evolution, which can be seen from individual cells regulatory networks to cell–cell and cell-microenvironment communication. It can be demonstrated from individual molecules to cells [69]. Cell decision-making plays a key role in the maintenance of phenotypic plasticity. Although, the subject phenotypic plasticity is a verse field, in this thesis, the phenotypic plasticity or switching at the level of cells are mostly concerned. The idea or the concept behind the phenotypic plasticity at cellular scale was first originated when biologists observed non-genetic variations in cellular phenotypes. It precisely tells us how same genetic pattern can influence multiple phenotype based on the set of microenvironment and the noise (i.e., internal and external). Two key examples of phenotypic plasticity are (**a**) Epithelial-Mesenchymal transition or the migration/resting plasticity and (**b**) Go-or-Grow or migration/proliferation plasticity phenomena.

Epithelial-Mesenchymal transition

Epithelial cells are immotile, attached by strong adhesion bonds, display regular shapes and exhibits a strong apico-basal cell polarity, on the other hand Mesenchymal cells can be identified morphologically as a small cell body with a few cell processes that are long and thin, which accommodates large, spherical nucleus with a prominent nucleolus. Epithelial and mesenchymal phenotypes can be illustrated by the representation of two migration modes, i.e., resting and motile, respectively. Epithelial-Mesenchymal transition is a phenomenon where Epithelial cells lose their epithelial phenotype (i.e., cell–cell adhesion and apico-basal polarity) and obtain Mesenchymal properties (i.e., migratory and invasive traits) [157]. One can see this kind of transitions (EMT and MET) mainly in embryonic develop-

ment, cellular reprogramming and cancer. In particular in this biological process also play a key role in Cancer invasions. From the view of biochemistry there exists multiple signaling biochemical pathways that control of EMT-HGF, TGF-β, p53, HIF-1α, EGF, FGF, Hedgehog, Wnt, and Notch signaling [119, 169]. Interestingly, experiments (in vivo and in vitro) and mathematical modelling suggests that there exists another state between Epithelial and Mesenchymal phenotype which is known as hybrid EMT state [83]. It has been shown that at a single cell level, EMT dynamics can show rich nongenetic heterogeneity. Single-cell immunofluorescence analysis helps us to understand the emergence of hybrid EMT state at the level of population average[138]. To characterize the dynamics of EMT, the information is needed at different level from individual cells regulatory networks to cell–cell and cell–microenvironment communication. The migration/resting type of plasticity has been detected during the epithelial-mesenchymal transition (EMT/MET)[86, 131, 156].

Migration/ proliferation plasticity or Go-or-Grow

Tumor is an abnormal development of somatic cells found in organisms. In tumors one can observe two important tumor cell features i.e., **(i)** irregular cell migration and **(ii)** unconstrained growth. During the cancer development, phenotypic plasticity (i.e., the transition from benign neoplasmas to malignant intrusive tumors) can be observed. Typically, tumor cell alterations have been attributed to somatic mutations. Recently, it is discovered that the phenotypic plasticity largely account to seek phenotypic changes. Go-or-Grow or Migration/ proliferation plasticity [57, 58] is a paradigm of the latter. This dichotomy is an out-turn of usual biological signalling pathways, present inside the tumor cells, which has been observed experimentally [56]. Another experimental result [59] also recognized a particular protein which participates in the regulation of migration and proliferation phenomena as per metabolic stresses. Interestingly, from mathematical models [65] one can show that how the lack of oxygen or hypoxia circumstances become a key player to understand the transition between highly proliferative phenotype and an invasive phenotype inside a growing tumor. The migration/proliferation plasticity can be seen in different types of cancers, such as glioblastoma or breast cancer, but also during normal development [65, 79, 71, 91]. The dichotomy also indicates the mutual exclusive events of cell migration and proliferation, which can lead to the survivability of proliferative/immotile and nonproliferative/mobile cell phenotypes.

1.2.2. Cellular migration: orientation decisions

Cell migration is an important process in the maintenance and development of multi-cellular organisms. This phenomenon occurs with respect to two kinds of signals i.e., **(A)** Mechanical signals and **(B)** chemical signals which often comes from an external microenvironmental source [101]. Membrane based integrin receptors attach to the extra-cellular matrix (ECM) ligand which further forms local adhesion

points or bond clusters to create force [66]. The process of cell migration at single cell level starts when the force in the front is greater than the force generates in the back. This disbalance in force later causes the fracture in the cell bonds at the back [139]. On the other hand, collective cell migration characterizes groups of migrating cells, such as swarms or cell aggregates. In biological cells, cellular migration involves a diverse set of biophysical mechanisms such as actin polymerization, receptor recruitment, or in bacteria, flagella motor reversal mechanisms, to name a few [143]. On a coarse-grained level, collective cell migration can be thought as a collective motion of self-propelled particles, which can be defined as the casual emergence of a many body particles/agents inside a system. Usually, individual agents drive themselves in an independent fashion and interact with other agents inside the neighborhood. These agents sometimes move in an organized or synchronized (like similar individuals are brought together) way, and sometimes they move in a chaotic or untidy manner. In nature, one can also monitor this kind of collective motion quite often. Examples can be seen across the scales of Biology, i.e., from massive migration of mammals ([13]) to cells during embryogenesis ([136]). Despite the diversity of collective motion examples, many of these processes are important in human activity. Due to the ubiquitous and universal nature of collective motion, there is always a need for quantitative understanding of collective migration. In this thesis, the collective motion of self-propelled particles will be discussed specifically at the cellular scale level. Here one can think as the collective cell migration is emergent result of cellular orientation decisions. The biological mechanism behind cell migration at single cell level has been well studied experimentally, but the set of biological mechanisms about the collective cell migration world is partially known. One of the central question in collective cell migration is that how biological mechanisms are correlated with the pattern formation. Sometimes we see universality in the pattern formation, though the biological mechanisms behind the pattern is way different. If one can make a coarse-grained model based on a principle from the partial biological knowledge, few questions can be answered about the origin of pattern formation. From theoretical modelling aspect, if one review literature of collective cell migrations, he can find that *Mechanistic* models have been seen to be very effective in analyzing collective cell migration. Such models rely on either explicitly granted phenomenological forces, which can be involved in the interaction and movement of individuals, or define some artificial rules or laws such that the system will mimic or act like the experimental behavior of real cells. Collective migration mechanisms are diverse and typically depend on the cell type. Additionally, the precise and complete regulation of many important chemo-mechanical factors influencing cell movement (from signalling pathways to substrate sensing) are typically either too complex or largely unknown. To address this situation, several mathematical models moderately introduce a *phenomenological* short-range bias that every individual feels. One of the most significant work on collective migration models ([162]) the so-called Viscek model, say that the direction of migration of particles shifts towards the mean velocity of individuals in a local neighborhood, and later at a temporal scale these rule promote a long-range swarming behavior at the level of population. Such models can be further clarified into mechanistic models, where individual particle dynamics are obeying a set of Langevin's equations. Specifically, the Langevin equation models, the reorientation of individual particle velocities is characterized by the

presence of a local interaction or coupled potential, which is resolved by the effects of neighboring particles. Collective migration has been realized through the glasses of a ferromagnetic-like interaction potential, which locally aligns particle velocities in a polar way, or a liquid-crystal-like interaction potential, which aligns particle velocities in a nematic way([124]).

1.2.3. Cell differentiation

Cell differentiation can be thought as a decision-making process, which is an event where one cell type evolve to another cell type [147]. In the continuation of time, one can find that stem cells kick off their odyssey and differentiate to other cell types. This kind of biological phenomena involves the substantial changes of geometry, shape, gene expression inside the cell, etc. [2]. Though we have a large set of multi-omics data for cell differentiation process, the difficulty is to understand the robustness of tissue differentiation from the mechanism point of view. The theoretical understanding of cell differentiation has been understood a long before in a metaphorical path due to C. H. Waddington landscape [60, 3, 29] (by C. H. Waddington) which has been used to study cell differentiation, in particularly the mathematical framework of dynamical systems, concerning single-cell fate decisions [48, 121, 82]. From this school of thoughts provoke an understanding about the mean-field dynamics of cell differentiation processes. It describes a situation where a cell differentiates in a well-mixed environment, but it does not explain the pivotal role of microenvironmental components. On the con side this school of thought can be problematic because it needs detailed information of the underlying genetic networks and other internal components (e.g., metabolites, cell signaling molecules, transcription factors, etc.) network which is hard to find experimentally. To handle these complications, statistical mechanics structure and data-driven perspectives [100, 165, 61, 68, 41, 123] have been recently suggested.

Finally, cell sensing is the central of cell decision-making. The ability of cells to perceive their microenvironment is known as cell sensing. It is a theme of basic research in biophysics since a long time, see for example [164] and references therein. Cells can realize frequent changes in their vicinity, by mainly utilizing two ways to decline their sensing errors: (**1**) to escalate the number of receptors and (**2**) to increase the number of measurements per receptors over a definite span, which are commanded by the downstream signaling pathways [14]. There exists an extensive literature on fundamental limits of cell sensing in both static and dynamic environments [164, 17, 112].

1.3. Challenges and open questions

If someone dives deep inside into Biology will find that phenotypic switching, cellular orientation biochemical decisions and cell differentiation involve numerous biophysical processes such as signalling pathways, genetic network regulation, recep-

tor dynamics, actin polymerization, receptor recruitment, or in bacteria, flagellar motor reversal mechanisms, to name a few [143]. However, in aforementioned biological events the precise knowledge of all coupled interacting biophysical/chemical mechanisms are mostly unknown. If the biophysical mechanisms are still unknown, the broad question is how to predict biological experimental results through the eye of mathematical modelling and biophysical principles. The key challenges of cell decision-making (in a broad picture) are:

- *Calculation of cell's states need a proper knowledge of intrinsic dynamics*

- *As the information in cell's intrinsic states are entangled with extrinsic states, the relative contribution coming from both factors is partially known.*

- *In cell decision-making, particularly in cellular differentiation problem, on one hand the directionality and robustness can be observed in an irreversible manner but on the other hand in phenotypic plasticity and cellular reprogramming phenomena directionality and robustness can be observed in a reversible manner which mainly depends on the microenvironmental factors. The question is, how can we unify both observations through a single theory.*

and for more specific cases one can write the questions in the following way:

- *Cellular decision-making has been precisely studied at the single cell level, but how individual cell decisions (e.g., phenotypic switching, cell differentiation etc.) has been guided by the dynamics at the level of tissue ?*

- *Can one predict the origin of pattern formations in collective cell migration case if one don't know the precise knowledge about all corresponding biophysical/biochemical mechanisms ?*

- *How optimal microenvironmental sensing is related with differentiated tissue at the spatial scale ? Can one understand the limits of sensing radius at robust tissue development ?*

In this thesis, my approach is to tackle the first (i.e., understanding cell's states require the proper knowledge of intrinsic dynamics) and third challenge (i.e., the unifying theory behind the directionality of phenotypes) through LEUP framework to answer the subsequent questions mentioned above.

1.4. Solution strategy

Here, my strategy is to understand biological challenges the set of open questions in a mathematical framework based on premises of assumed biological principles. The mathematical framework used is based on the theory of Bayesian inference [24], stochastic thermodynamics [141, 140] and the Maximum Entropy Principle [76, 77]. It uses the statistical inference tool to bypass the details of intra-cellular regulation, while keeping the dynamic aspect. The logic behind visu-

alizing cells as Bayesian decision-makers below the tree of microenvironmental entropic/energetic constraints and the partial knowledge about internal states as a constraint can be formulated into a free-energy like principle. This free-energy like principle later can be translated to a statistical mechanics formulation for cell decision-making known as LEUP [66]. Following this formulation, cells will try to enhance their internal priors (e.g., genes, RNA molecules, translational proteins, metabolites, membrane receptors, etc.) such that there will be a maximization in their joint entropy (i.e., internal states and the microenvironmental states around the cellular vicinity). From this point of view, one can understand cell decision-making in a following manner :

- *From LEUP, biological mechanisms can be understood through a coarse-grained manner, which offers a low-dimensional mathematical description to circumvent the uncertainty about the underlying mechanisms. Further, it helps to study the complex system in a simplified way to predict experimental results.*

- *The origin of biological pattern formation can be figured out in terms of local information and adaptation. Specifically, one can observe the correlation among the magnitude of sensitivity, sensing radius and the emergence of global pattern.*

- *The switching between Epithelial and Mesenchymal cell can be well explained from microscopic and mesoscopic picture with the help of LEUP without knowing the biological details. It also shows how "fluid"-like to "solid"-like structures can be formed during the EMT.*

- *Using this principle, the spatio-temporal robustness of the tissue can be demonstrated. One can also unfold the cause of irreversibility in tissue differentiation through LEUP driven fluctuation theorem.*

This principle has been applied in different biological systems single cell migration [66], collective cell migration [11], Epithelial-Mesenchymal transition [12] and in cell differentiation [10]. In this thesis, the LEUP theory is further discussed in chapter 2 and later continued in chapter 3, 4 and 5.

1.5. Structure of thesis

In chapter 2, the theory behind LEUP have been presented where the motivations and hypothesis are presented. In turn, it helps us to derive the mathematical formulation.

In chapter 3, the mathematical model based on LEUP is applied on the migration/resting plasticity. At first, a stochastic individual-based model (IBM) has been established for moving and resting cells where cells can switch their phenotypes correspondingly according to the LEUP framework and precisely tells the formation of aggregates of resting cells in corresponding simulations. Then, a macroscopic mean-field approximation is derived of the corresponding microscopic model, which helps to understand the formation of aggregates in the microscopic model. Later,

the stability of the steady states and the pattern formation potential of the macroscopic Go-or-Grow model is calulated. Finally, the biological conclusions of the results in terms of multicellular growth and pattern control is explained.

In chapter 4, the simplest LEUP-driven Langevin model of swarming has been presented, where individuals can sense the velocity orientations of other individuals in their neighbthehood. All the individuals present inside the system act as Bayesian inferred and switch their own orientation to optimize their prior, according to microenvironmental orientation information. Under these set of assumptions, individuals reorient themselves in agreement with the microenvironmental entropy gradient. The sensitivity parameter, controls the strength and directionality of the reorientation with respect to the local gradient. On one hand the system adopts a steady and polar-ordered state for negative values of the sensitivity and on the other hand, the system remains out of equilibrium, but partially nematic-ordered when the sensitivity is positive. After that the qualitative behavior of the model depends on various factors i.e., the values of the particle density, noise strength, sensitivity, and size of the interaction neighborhood. At last, the LEUP principle can mimic the collective behavior of spherical *S. marcescens* bacteria.

In chapter 5, the basic concepts of the LEUP has been used to understand cell differentiation phenomena. After that, the answer to the question of how optimal microenvironmental sensing is related with differentiated tissue at the spatial scale ? has been answered. In particular example, the LEUP framework is studied using the experimental data achieved from the avian photoreceptor mosaics. Then, the fluctuation theorem of cell differentiation has been derived, which in turn exhibit the thermodynamic robustness of this biological process in the case of the avian retina development. Later, the answer to the question of how cell sensing radius and total entropy production are correlated to understand the limits of sensing radius at robust tissue development ? has been explicitly discussed.

In chapter 6, the major results of the thesis are summed up. Extensions in the LEUP theory has been introduced in the thesis are explained.
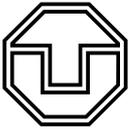
Part II.

Least microEnvironmental Uncertainty Principle (LEUP)

2. Least microEnvironmental Uncertainty Principle (LEUP)

In this chapter, the concept of Least microEnvironmental Uncertainty Principle (LEUP) will be introduced. Additional supplement to the LEUP framework will be presented in chapters 3, 4 and 5.

LEUP-based coarse-grained low-dimensional mathematical model can help to analyze and understand the synergistic effects of cell decision-making, phenotypic switching and the robustness of cell differentiation phenomena which is the main aim/objective of the thesis. Please note that, the LEUP theory is still a hypothesis which currently undergoes experimental validation. LEUP suggested recently to understand the cell decision-making phenomena in multicellular systems and particularly applied in cell migration force distribution [66], collective cell migration [11], phenotypic plasticity [12] and tissue differentiation [10]. In this thesis, the basic theory of LEUP and its corresponding applications will be explained.

2.1. Hypothesis behind LEUP

The hypothesis behind the LEUP is hidden inside the Bayesian inference and learning, which helps to take a decision about their internal states while circumventing the details of the intra-cellular knowledge. Further, one can show that the attenuation of microenvironmental entropy over time happens for a specific sensitivity regime. So, the main idea of cell decision-making is based on the entropy maximization approach. Fabricating an accurate specific microenvironmental sensing distribution is energetically expensive, if one think from cell's prospect, cells gather knowledge from their vicinity implementing different biological processes, such as polymerizing pseudopodia, translocating receptor molecules or modifying its cytoskeleton according to bio-mechanical signals [149, 159]. However, when cells construct informative priors about their microenvironment, then energetically cost

can be minimized. For example, in the advancement of cell polarization along, the employment of actin related proteins [132] during cell migration allows cells to spare the energy from building extra sensory processes. At the same time, in many biological processes such as, cell polarization, development and wound healing, cell phenotypic decisions, tissue differentiation effectively promote the co-evolution of the cellular microenvironment over time into more organized states, leading to a decrease of the microenviromental entropy from high to low over time. In turn, it helps cells to adapt and learn from their microenvironment. Later, the simultaneous resolution of the cellular state and its corresponding microenvironment terminates toward the reduction of the microenvironmental uncertainty tends to an optimum/target value correlated to a given tissue, phenotype or bacteria (e.g., biofilm). The basic challenges are the (i) uncertainty of high-dimensional subcellular regulatory cell decision-making mechanisms, and the (ii) the lack of knowledge in the relative contribution of intrinsic and extrinsic cell decision-making factors to multicellular spatio-temporal dynamics. In Fig. 2.1, how LEUP is positioned in terms of model interpretability and knowledge of biophysical details in comparison the afore-mentioned modelling/computational approaches has been illustrated. Interestingly, LEUP proposes a balanced solution for problems of low mechanistic knowledge and satisfactory interpretability. Akin notions (using information theoretic measures) can be observed, which have been suggested in crucial works by W.Bialek [16].

2.2. Mathematical formulation

2.2.1. Cell as Bayesian decision maker

Let's presume that the cell reacts to the environmental information, Y_i, by adapting its own states, X_i. The cell then reacts as a Bayesian decision-maker, such that

$$P\left(X_i \mid Y_i\right) = \frac{P\left(Y_i \mid X_i\right) P\left(X_i\right)}{P\left(Y_i\right)}. \tag{2.1}$$

Here internal variables are defined as $X_i \in \{[0,1], \theta_i, v_i, x_s, x_d\}$ and external variables are defined as $Y_i \in \{[N_i^1 \mid N_T], \Theta_i, V_i, y_s, y_d\}$. Where $P\left(Y_i \mid X_i\right)$ can be understood as the perfection with which a cell can precept other environmental factors in their circumference and respond accordingly, and $P\left(X_i\right)$ is the probability distribution of the cell's intrinsic states (or prior). Although, detecting neighborhood cells and calculating $P\left(Y_i \mid X_i\right)$ implies a cost for energy. It is valid to take this assumption that the cell will try to improve its prior $P\left(X_i\right)$ due to the economic cost of energy. Respecting cells as Bayesian decision-makers under entropic/energetic constraints, one can come up that cell decisions are dominated by a 'Least microEnvironmental Uncertainty Principle (LEUP)'.
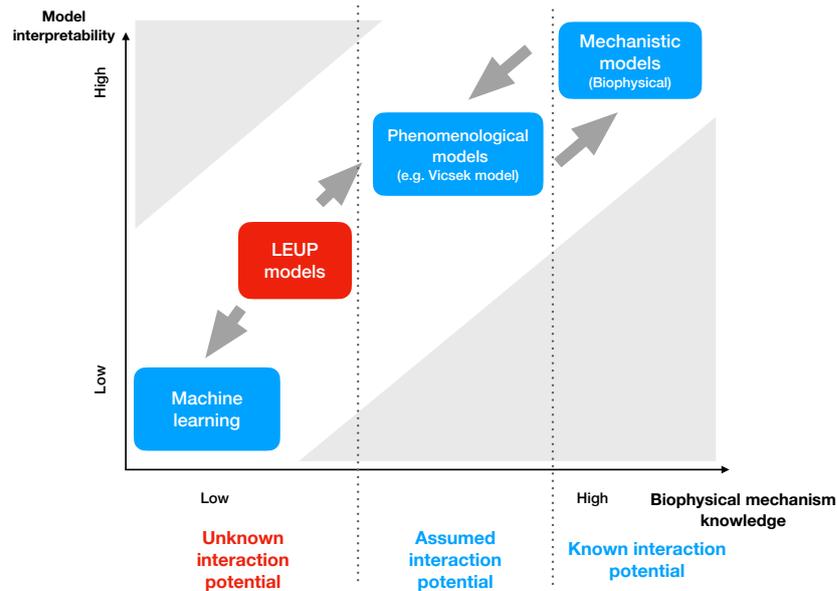
Figure 2.1.. In the figure, the applicability of different modelling/computational approaches are described according to their biological interpretability and the corresponding knowledge of biophysical migration mechanisms. The latter is typically encoded as the degree of interaction potential knowledge in Langevin equations (see text). In the case of extended biophysical mechanism knowledge, mechanistic models are the natural choice. When the effects of cell-cell interaction on cell migration are only partially understood, phenomenological models can be typically used. Finally, when data do not suffice to formulate an interaction potential, machine learning allows for the quantitative reproduction of experimental data. However, this has a toll in the interpretability of the resulting model, since machine learning methods are typically "black boxes". LEUP models offer a compromise that allows for quantitative predictions under lack of mechanism knowledge and satisfactory biological interpretability.

2.2.2. Variational principle for LEUP

To construct a variational principle for cell decision-making, the prior probabilities $P(X_i)$ should accomplish certain premises of LEUP. (**a**) Prior probabilities should be normalized, i.e., $\int P(X_i) \, dX_i = 1$, integrating over all possible values of internal cellular states. (**b**) A biological cell acts as an immature sensor. Therefore, the unpredictability arises during sensing accuracy $S(Y_i \mid X_i)$, should grasp a definite level in average, which is species-dependent or cell type-dependent. This set of assumptions are the heart of the LEUP formalism. An essential ingredient part of the LEUP formalism is the entropy maximization principle which, in general, has its applicability when every other mechanistic detail of the system is unknown or failed to explain the experimental data. This kind of formalism helps to carry out theoretical predictions when very few things from experiments are known. Entropy is a measure of uncertainty or unpredictability and inversely proportional to the information gain. According to information theory, thus entropy should be maximized, in order to follow the lack of mechanistic knowledge of the phenomenon, and to circumvent any kind of artificial prejudice in the model which can be appeared from a set of particular choices of $P(X_i)$[127]. From the definition of the internal Shannon entropy is given by $S(X_i) = -\int P(X_i) \ln P(X_i) \, dX_i$, entropy maximization subjected to probability normalization and the objective mean sensing accuracy deciphers into an optimization problem which reads as

$$\frac{\delta}{\delta P(X_i)} \left\{ -\int P(X_i) \ln P(X_i) \, dX_i \pm \beta \left[\int P(X_i) S(Y_i \mid X_i) \, dX_i - \bar{S}(Y_i \mid X_i) \right] \right.$$
$$\left. -\lambda \left[\int P(X_i) \, dX_i - 1 \right] \right\} = 0, \tag{2.2}$$

where $\frac{\delta}{\delta P(X_i)}$ is the functional derivative, $\bar{S}(Y_i \mid X_i)$ is the target sensing accuracy, and λ and β are Lagrange multipliers. Taking into account the relations among entropy and probability, Eq. (2.2) can be written as

$$P(X_i) = \frac{e^{\pm \beta S(Y_i \mid X_i)}}{Z}, \tag{2.3}$$

where $Z = \int e^{\pm \beta S(Y_i \mid X_i)} dX_i$ is a normalization constant or like a partition function and β is the responsiveness or the sensitivity about the environment of the cell. Using Eq. (2.3), the internal entropy of the cell, defined as $S(X_i) = -\int P(X_i) \ln P(X_i) \, dX_i$, is given by

$$S(X_i) = \mp \beta \overline{S(Y_i \mid X_i)} + \ln Z. \tag{2.4}$$

Using the relation between thermodynamic-like potentials, it is evident that the mean internal energy is given by

$$U(Y_i, X_i) = \overline{S(Y_i \mid X_i)} = \langle S(\mid X_i \neq \alpha) \rangle \tag{2.5}$$

and for a particular realization of X_i can write the internal energy as

$$U_i(Y_i, X_i) = S(Y_i \mid X_i) \tag{2.6}$$

Now, one can also write the Helmholtz-like free energy as

$$F = -\frac{1}{\pm \beta} \ln Z. \tag{2.7}$$
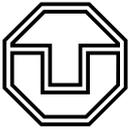
Part III.

LEUP in biological problems

3. Phenotypic plasticity : dynamics at the level of tissue from individual cell decisions

In this chapter,[1] phenotypic plasticity has been studied from the eye of LEUP. It is still an open question, how individual cell decisions influence the dynamics at the tissue level. Here, the implications of phenotypic plasticity on spatio-temporal pattern formation in the cell population have been analysed and the migration/resting and migration/proliferation plasticity have been examined. A biological example i.e., the EMT/MET mechanism has been used to study the role of LEUP in phenotypic plasticity. Interestingly, it has been shown that there exists a tight interaction of EMT decisions and the Notch-Jagged-Delta pathway [19]. It is known that EMT/MET is extremely important in development. For instance in the context of neural crest cells during embryonic development [107], the activation of EMT in pluripotent neural crest cells, migrating away from the embryonic ectoderm, is tightly connected with the binary cell decisions towards neurons, glia, melanocytes and other cell types [148]. These binary cell fate decisions are typically regulated by signalling pathways such as the infamous Notch-Delta pathway [5]. Therefore, the collective dynamics of these individual cell decision to activate the EMT program and migrate promote the organization and maturation of a multitude of vertebrate tissues, such as the brain. Thus, the EMT/MET decision-making leads multicellular ensembles to low entropy mircoenvironmental states, e.g., differentiated tissues. In order to understand the EMT/MET decision-making, a preliminary assumption has been taken i.e., cells change their phenotype in order to decrease their microenvironmental entropy over time following the LEUP (Least microEnvironmental Uncertainty Principle) hypothesis which also help to study the impact

---

[1]This chapter includes text and figures from the publication and corresponding supplementary information: Arnab Barua, Simon Syga, Pietro Mascheroni, Nikos Kavallaris, Michael Meyer-Hermann, Andreas Deutsch and Haralampos Hatzikirou, *New Jthenal of Physics* (2020) 22: 123034. Author contribution: Arnab Barua, Simon Syga and Nikos Kavallaris have performed the analysis, Arnab Barua and Pietro Mascheroni did the simulations of the model, and Arnab Barua and Simon Syga wrote the manuscript. All authors interpreted the results. Haralampos Hatzikirou , Michael Meyer-Hermann and Andreas Deutsch supervised the study.

of the LEUP-driven migration/resting andmigration/proliferation plasticity on the corresponding multicellular spatio-temporal dynamics with a stochastic cell-based mathematical model for the spatio-temporal dynamics of the cell phenotypes. In the case of the Go-or-Rest plasticity, a corresponding mean-field approximation allows to identify a bistable switching mechanism between a diffusive (fluid) and an epithelial (solid) tissue phase which depends on the sensitivity of the phenotypes to the environment. At last, Go-or-Grow plasticity has been studied, which shows the possibility of Turing pattern formation for the "solid" tissue phase and its relation with the parameters of the LEUP-driven cell decisions.

3.1. Mathematical framework

A mathematical model is defined for LEUP-based switching between moving and resting phenotypes. An assumption has been taken such that cells can sense their microenvironment and can change their own phenotype X_i on a domain $\mathcal{L} \subset \mathbb{R}^2$ accordingly, where $X_i = 0, 1$ corresponds to the resting and migrating state, respectively. The microenvironment of the cell i is denoted as the number N_i^0 of cells having phenotype $X_i = 0$ and by the number N_i^1 of cells having phenotype $X_i = 1$. By assuming a maximum cell capacity N of the microenvironment , one can define

$$N = N_i^1 + N_i^0 + N_i^\phi, \tag{3.1}$$

where N_i^ϕ are the free spaces/slots. The total number of cells N_T is defined as the sum of cells having phenotype $(X_i = 1)$ and cells having phenotype $(X_i = 0)$. So,

$$N_T = N_i^1 + N_i^0. \tag{3.2}$$

Cells will select their phenotype in a Bayesian fashion according to Eq. (2.1). It is reasonable to assume that the cell will try to optimize its prior $P(X_i)$ for the sake of energetic frugality. Here Y_i is defined as the extrinsic random variable of the i-th cell, where $Y_i = N_i^{X_i} \mid N_T, X_i = 0, 1$ represents the number of cells of phenotype X_i given the total local number of cells N_T. The conditional probability of having N_i^1 the number of cells present in the microenvironment follows a binomial distribution ($\boldsymbol{B}$) (for the derivation details, see the Supplementary Material 7):

$$P\left(N_i^1 \mid N_T\right) = P\left(Y_i = N_i^1 \mid N_T\right) = \boldsymbol{B}\left(N_T, p_1\right), \tag{3.3}$$

where p_1 is the probability of N_i^1 number of cells having phenotype X_i = 1 out of N_T cells: eqn.

$$p_1 = \frac{N_i^1}{N_T}. \tag{3.4}$$

Taking into account the relations between entropies, Eq. (2.2) (using +ve sign of β) yields

$$P\left(X_i\right) = \frac{e^{\beta S(Y_i \mid X_i)}}{Z} = \frac{1}{1 + e^{\beta \Delta S}}, \tag{3.5}$$

where Z is the normalization factor

$$Z = \sum_{\alpha=0,1} e^{\beta S(Y_i \mid X_i = \alpha)} = e^{\beta S(Y_i \mid X_i = 0)} + e^{\beta S(Y_i \mid X_i = 1)}. \tag{3.6}$$

Please note that the parameter β has a biological interpretation, since it quantifies the intensity with which a cell senses and complies to the microenvironment. To facilitate the evaluation of the entropy distribution of the microenvironment, have used a Gaussian distribution, which approximates the binomial distribution (for $N_T > 5$), i.e.,

$$P(N_i^1, p_1) \rightarrow \boldsymbol{N}(N_T p_1, N_T p_1 (1 - p_1)), \tag{3.7}$$

for the microenvironment [24]. Now one can evaluate the equilibrium distribution $P(X_i = 0)$ as

$$
P(X_i = 0) = \frac{e^{\beta S\left(Y_i = [N_i^1 | N_T] | X_i = 0\right)}}{Z} = \frac{1}{1 + e^{\beta \Delta S}} =
$$
$$
= \frac{1}{1 + \left[\frac{N_i^0 \left(N_i^1 - 1\right)}{N_i^1 \left(N_i^0 - 1\right)}\right]^{\frac{\beta}{2}}} = \frac{1}{1 + \left[\frac{\rho_i^0 \left(\rho_i^1 - \frac{1}{V}\right)}{\rho_i^1 \left(\rho_i^0 - \frac{1}{V}\right)}\right]^{\frac{\beta}{2}}}, \tag{3.8}
$$

with the cell densities $\rho_i^{0,1} := \frac{N_i^{0,1}}{V}$ (in the Supplementary Material 7 Fig. (7.4), it can be shown that the dependency of $P(X_i = 0)$ on resting density ρ_0 and β). Accordingly, the probability of a i-th cell's phenotype $P(X_i = 1)$ reads as

$$
P(X_i = 1) = \frac{e^{\beta S\left(Y_i = [N_i^1 | N_T] | X_i = 1\right)}}{Z} = 1 - \frac{1}{1 + e^{\beta \Delta S}} =
$$
$$
= 1 - \frac{1}{1 + \left[\frac{N_i^0 \left(N_i^1 - 1\right)}{N_i^1 \left(N_i^0 - 1\right)}\right]^{\frac{\beta}{2}}} = \frac{\left[\frac{\rho_i^0 \left(\rho_i^1 - \frac{1}{V}\right)}{\rho_i^1 \left(\rho_i^0 - \frac{1}{V}\right)}\right]^{\frac{\beta}{2}}}{1 + \left[\frac{\rho_i^0 \left(\rho_i^1 - \frac{1}{V}\right)}{\rho_i^1 \left(\rho_i^0 - \frac{1}{V}\right)}\right]^{\frac{\beta}{2}}}. \tag{3.9}
$$

Please note that, the difference between the microenvironmental entropy of the i-th cell is

$$
\Delta S = S(Y_i = [N_i^1 \mid N_T] | X_i = 1) - S(Y_i = [N_i^1 \mid N_T] | X_i = 0) = \frac{1}{2} \ln \left[\frac{N_i^0 \left(N_i^1 - 1\right)}{N_i^1 \left(N_i^0 - 1\right)}\right], \tag{3.10}
$$

N_i^0 and N_i^1 are the number of cell of phenotype (X_i = 0) and phenotype (X_i = 1) in the microenvironment of i-th cell, respectively (See Supplementary Material 7 for details).

Fig. (3.1i), shows a sketch of the model. In short, make the following assumptions
$(A1)$ The microenvironment is defined by the numbers of two phenotypes only.
$(A2)$ A binomial distribution is assumed for the occurrence of the two phenotypes.
$(A3)$ Cells are making decisions at a fast timescale. This justifies to assume an equilibrium distribution for the different phenotypes.

In this section, one can define a minimal LEUP-driven cell decision-making model and derived the corresponding phenotypic steady states. Based on this, subsequently develop a cell-based model to understand the resulting multicelluar spatio-temporal dynamics.
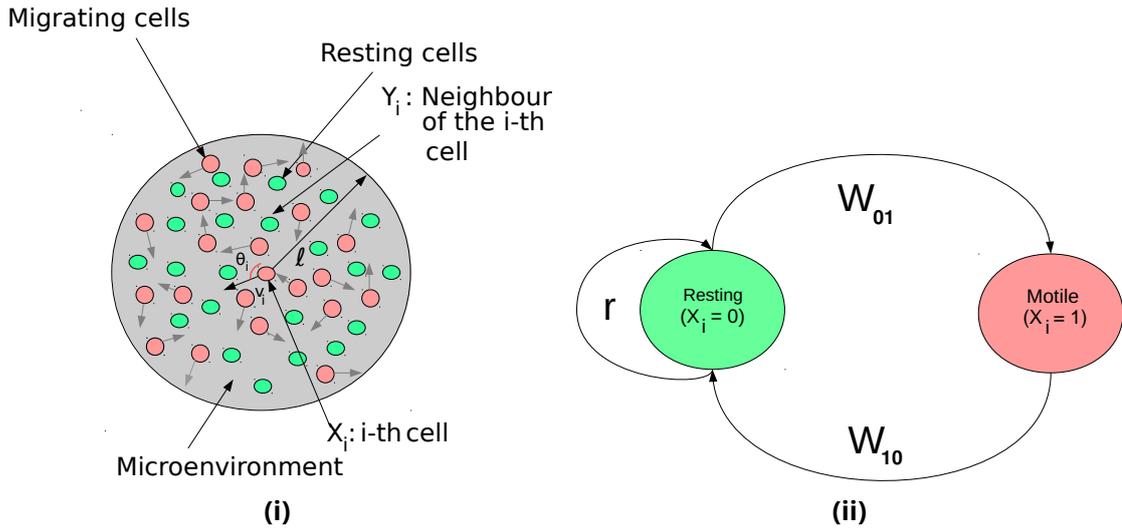
Figure 3.1.. Schematic diagram of (**a**) the microenvironment of a motile cell, where ℓ is the sensing radius. The sensed number of cells is proportional to the volume of the sensed microenvironment, i.e., $N \propto \ell^d$, where d is the dimension. (**b**) Transitions in the "Go-or-Grow" model [65]. The switching between motile $(X_i = 1)$ and resting $(X_i = 0)$ phenotype is shown, where the transition probabilities are defined as W_{10} and W_{01}. The proliferation rate is defined by r.

3.2. Individual based model (IBM)

A discrete stochastic, spatio-temporal IBM has been presented such that it can incorporate the phenotypic switch dynamics according to LEUP. To this end, the movement of single cells have been modelled with Langevin equations. Langevin equations are well-suited to model cell-cell interactions and cell migration [63, 137].

The Langevin's equation is defined on a domain $\mathcal{L} \subset \mathbb{R}^2$ with periodic boundary conditions. Define an interaction radius $\ell \in \mathbb{R}$ around the i-th cell at position $\mathbf{x_i} \in \mathcal{L}$. The expected interaction volume is $V \propto \ell^2$. The time evolution of the model is defined by the following rules:

- (R1) Cells change their phenotype by sensing their microenvironment within the interaction radius ℓ according to LEUP.

- (R2) Moving cells change their orientations randomly (random walk).

- (R3) Once cells become migratory, they move with a constant speed $\bar{v}$.

The above are translated to the following Langevin's equations,

$$\frac{\mathrm{d}\mathbf{x_i}}{\mathrm{d}t} = \bar{v}\mathbf{v_i}(\theta_i),$$
$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = \frac{1}{\bar{v}}\xi_i^\theta(t),$$
$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{1}{\tau}(p_i - p_i^{eq}), \tag{3.11}$$
$$v_i = \begin{cases} \bar{v}, & \text{if} \quad X_i = 1, \\ 0, & \text{if} \quad X_i = 0, \end{cases}$$

where $\mathbf{v_i} = (\cos\theta_i, \sin\theta_i)^T$ is the direction of movement of the cell i. For the temporal evolution of the probability p_i of the motile state, the BGK (Bhatnagar–Gross–Krook) operator technique [15] has been used. The assumption is that the probability p_i evolves weakly out of its equilibrium probability p_i^{eq}, which is the LEUP steady state probability $P(X_i = 1)$ (see Eq. (3.9)). In turn, the parameter τ is the relaxation time towards the corresponding probability distribution p_i^{eq}. Here the noise is assumed to have a zero-mean, white noise term, which has the statistical properties $\langle\xi_i^\theta(t)\rangle = 0$ and $\langle\xi_i^\theta(t_1)\xi_j^\theta(t_2)\rangle = 2D_\theta\delta(t_1 - t_2)\delta_{ij}$, where t_1 and t_2 are two time points, $D_\theta \in \mathbb{R}_+$ is the angular diffusion coefficient, $\delta(t)$ is the Dirac delta, and δ_{ij} is the Kronecker delta. Parameter $\bar{v}$ is the constant speed of every motile cell.

Simulate the Langevin model, on a two-dimensional domain, with a varying interaction radius ℓ, mean cell density and sensitivity β. The assumption is that an initial state that is approximately homogeneous in space and that each of the initial 1000 cells is in the resting/migratory state with equal probability. In Figures (3.2i) and (3.2ii), showcase simulations that exhibit clustering and random spatial configurations, respectively. One can consider a density regime which is compatible with typical situations during cancer invasion. In particular, the densities in Fig. 2 correspond to 1000 glioma cancer cells in a $2.5 \times 10^4 \mu m^2$ domain, which agrees with the typical glioma cell density $4.5 \times 10^{-3}\,\mu m^{-2}$ [81].

To quantify cell clustering, the radial distribution function has been calculated for different sensitivities and different interaction radii, as shown in Figures (3.3i) and (3.3ii). The key observations from the simulation study are:

1. There is a critical threshold for the parameter $\beta > 0$, where clustering of resting cells occurs (Fig. 3.3i). This is rather expected, since a sufficient sensing of the microenvironment is required.

2. For an intermediate interaction radius ℓ, observe cell clusters as shown in Fig. 3.3ii. Very high ℓ corresponds to a large number of sensed cells. This leads to also equal steady state probabilities, as the entropy difference in the microenvironment, associated with the single cell states, becomes negligible. On the other hand, the lower bound of ℓ is expected, since enough sampling size of cellular microenvironment is required to induce aggregation patterns.

3. In Fig. 3.3iv, show the phase diagram of the system, when parameter β and the average density are varied. In particular, one can observe that there is a parametric regime where clustering behavior is emerging. Interestingly, high
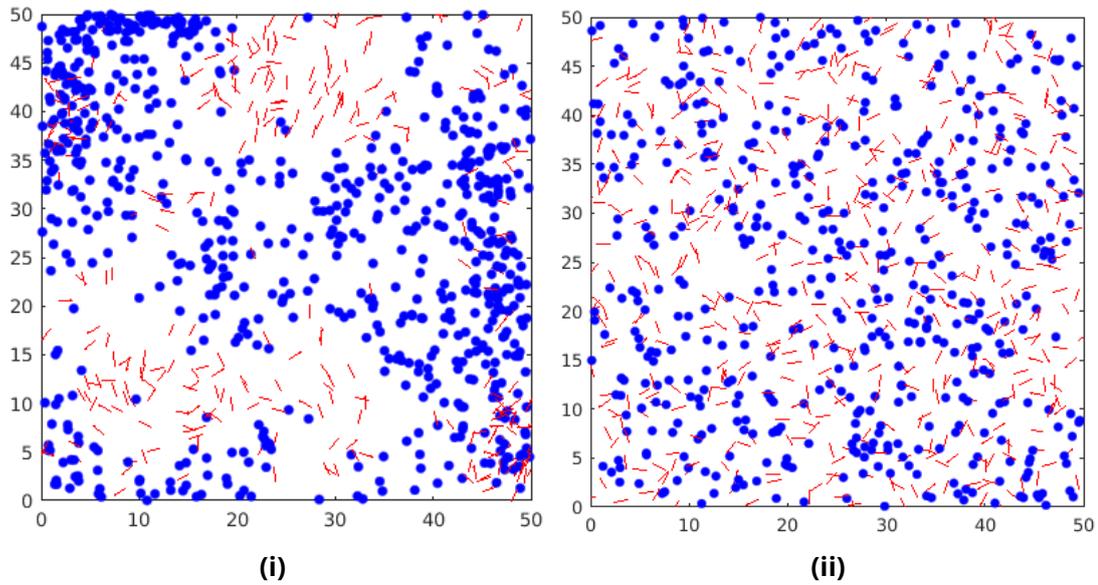
Figure 3.2.. Clustering of resting cells in the IBM. (**a**) Snapshot of an example realization at $t = 100$, where the total number of cells is fixed at 1000 and $\beta = 40$. The spatial configuration shows the clustering of resting cells and motile cells moving in the free space. (**b**) Snapshot of the realization at $t = 100$, where the total number of cells is fixed at 1000 and $\beta = 5$. Here, one observe a completely random spatial configuration. Resting cells are shown in blue and migrating cells are marked by red arrows indicating their direction of movement.

densities do not always imply patterning and require higher values of β to support cluster emergence (see Fig. 3.3iii). Finally, as expected, low densities also reduce the area of the patterning regime for low β.

Interestingly, existing radial distribution function data from patient-derived glioma tumors show also a "flat" behavior, i.e., they do not exhibit any clustering [81]. In particular, the configuration in Figure (3.2ii) corresponds to such a "flat" radial distribution function, for $\beta = 5$, at densities relevant to real glioma tumors. However, there are some extra cell-cell interactions mechanisms acting in glioma cells, i.e., a repulsive force conferring a large volume exclusion effect. Studying effects of these additional mechanisms is the topic of current research.

3.3. Mean-field approximation

In this section, A continuous approximation is derived of the aforementioned discrete model using a mean-field approach. The goal is to shed light on the pattern formation mechanisms, i.e., cell clustering, as observed in the microscopic simulations. The main idea of the mean-field approximation is to replace the description of many-particle interactions by a single particle description based on an average or effective interaction. Thereby, any multi-particle problem can be replaced by an effective description, that can be stated in the form of an ordinary (ODE) or partial differential equation (PDE). In order to proceed, one will first treat the switch dynamics and the migration process separately.

24

(i)

(ii)

(iii)

(iv)
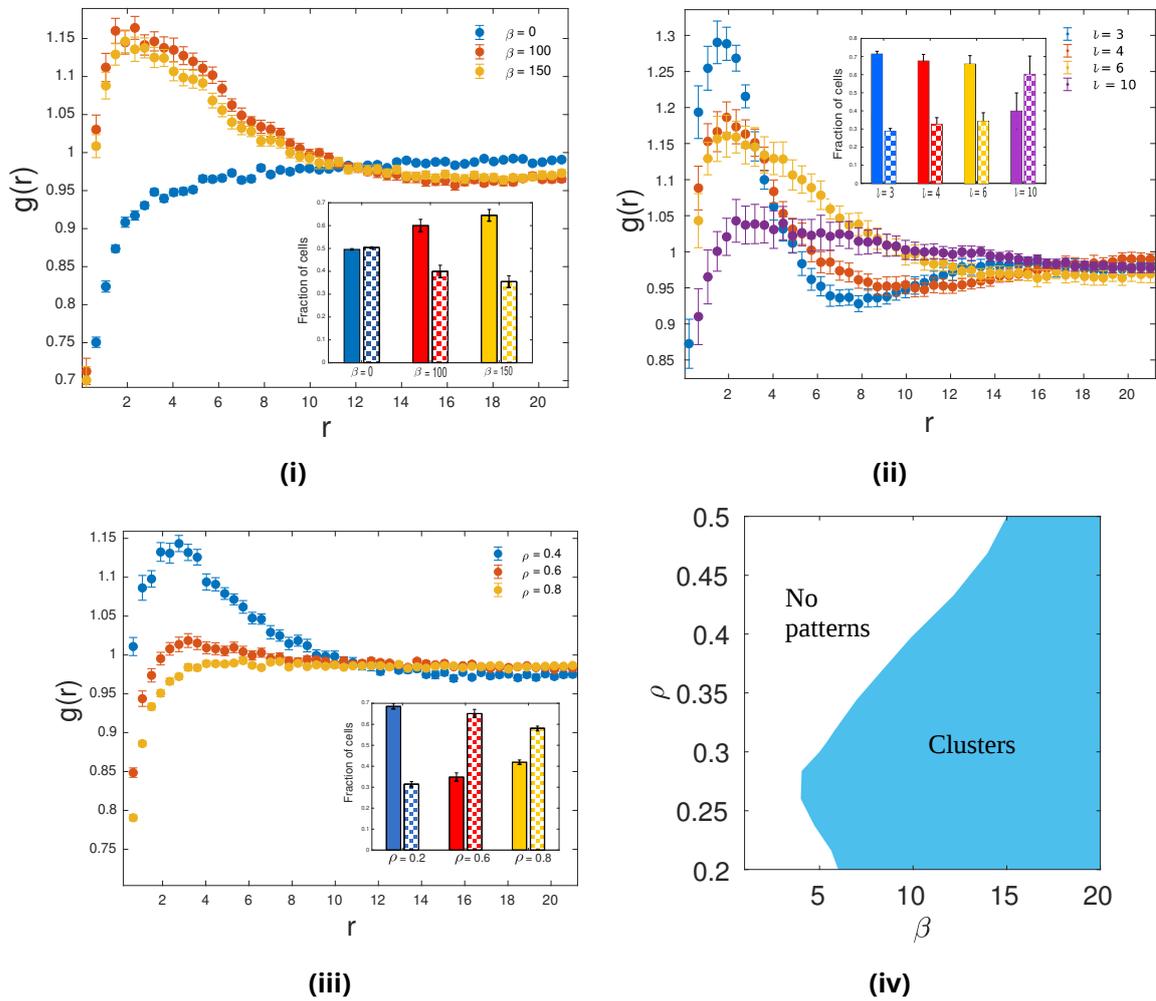
Figure 3.3.. Quantification of cell clustering. (**a**) Radial distribution function with standard deviation for different values of β, for a fixed interaction radius $\ell = 6$. The number of cells was fixed at 1000, corresponding to a mean cell density of 0.4. (**b**) Radial distribution function with standard deviation for varying interaction radius ℓ at fixed sensitivity $\beta = 100$. There is an optimal sensing radius for a given mean cell density, so that interaction radii that are too large or too low do not lead to aggregation, indicated by an almost flat radial distribution, see purple and blue lines. (**c**) Radial distribution function with standard deviation for different mean densities, at $\beta = 20$ and $\ell = 4$. When the mean density becomes too large, no clustering can be observed. The inlets in (**a - c**) show the corresponding fractions of motile (checkerboard bars) and resting (filled bars) cells. (**d**) Phase space diagram for cluster pattern formation. Here define clustering when $\max_r (g(r) > 1.09)$ (see text for explanation). The interaction radius is fixed at $\ell = 4$. All data points in (**a - d**) correspond to the mean of 20 independent simulations at $t = 100$.

3.3.1. Phenotypic switching dynamics

Now consider cell i at position $\mathbf{x_i}$ with states $X_i = 0, 1$. The cell changes its state in dependence on the local microenvironment according to the LEUP with rates W_{01}, W_{10} and the probabilities to be in either state are denoted by $p_{0,1}$. One assume that the system is always close to the steady state, so the master equation reads

$$\frac{\mathrm{d}}{\mathrm{d}t}p_0 = W_{10}p_1 - W_{01}p_0 \approx 0, \tag{3.12}$$

and $p_{0,1}$ are given by eq. (3.5). By rearranging the terms, obtain for the switching rates

$$\frac{W_{10}}{W_{01}} = \frac{p_0}{p_1}. \tag{3.13}$$

For simplicity, one can set $W_{01} = p_1$ where the transition probability towards motile phenotypes equals to the moving steady state probability p_1. This coincides with the "detailed balance condition". The steady state probabilities $p_{0,1}$ depend on the number of cells $N_i^{0,1}$ in the respective phenotypes in the microenvironment. Cell i at position $\mathbf{x_i}$ senses

$$N_i^X(\mathbf{x_1}, \ldots, \mathbf{x_i}, \ldots, \mathbf{x_i}, t) := \sum_{j=1}^{N} \xi(\mathbf{x_j}(t) - \mathbf{x_i}(t), t)\delta_{X_j, X}, \quad X = 0, 1. \tag{3.14}$$

Here, sum over all cells j, and $\xi(\mathbf{x_j}(t) - \mathbf{x_i}(t), t)$ is a Boolean stochastic variable that serves as the sensing function of a cell at $\mathbf{x_i}$. It depends on the distance between cells and time, and $\xi = 1$ indicates that the cell j at $\mathbf{x_j}$ is sensed by cell i at position $\mathbf{x_i}$. To match the IBM, assume that

$$\xi(\mathbf{x_j}(t) - \mathbf{x_i}(t), t) = \Theta(\ell - |\mathbf{x}_j(t) - \mathbf{x_i}(t)|), \quad \ell > 0, \tag{3.15}$$

with the Heaviside step function $\Theta(x)$, so that all cells in a ball of radius ℓ around $\mathbf{x_i}$ are sensed.

To proceed, a mean-field approximation is applied to calculate the expected switching rate $\langle W_{10}(N_i^0, N_i^1) \rangle \approx W_{01}(\langle N_i^0 \rangle, \langle N_i^1 \rangle)$. Note that the dependence on space and time is dropped for better readability. Let $\phi_i(X, \mathbf{x}, t)$ denote the probability of finding a cell i in a small volume $\mathrm{d}V$ around $\mathbf{x}$ at time t with the phenotype $X = 0, 1$. Formally have

$$\phi_i(X, \mathbf{x}, t) = \langle \delta(\mathbf{x} - \mathbf{x_i}(t))\delta_{X, X_i} \rangle, \tag{3.16}$$

where the average is the ensemble average, and the total densities of resting/migrating cells are

$$\Phi(\mathbf{x}, t) := \sum_i \phi_i(0, \mathbf{x}, t), \tag{3.17}$$

$$\Psi(\mathbf{x}, t) := \sum_i \phi_i(1, \mathbf{x}, t). \tag{3.18}$$

For the expected density of sensed cells in the microenvironment around $\mathbf{x}$ obtain

$$\rho_0(\mathbf{x}, t) := \frac{\langle N_i^0(\mathbf{x}, t) \rangle}{V} = \frac{1}{V} \int_{\mathcal{B}_\ell(\mathbf{x})} \Phi(\mathbf{y}, t) \, dV, \tag{3.19}$$

$$\rho_1(\mathbf{x}, t) := \frac{\langle N_i^1(\mathbf{x}, t) \rangle}{V} = \frac{1}{V} \int_{\mathcal{B}_\ell(\mathbf{x})} \Psi(\mathbf{y}, t) \, dV, \tag{3.20}$$

where $V \propto \ell^d$, and where have already used eq. (3.15). Consequently, to obtain the approximate switching rate

$$\bar{W}_{10} = p_0(\rho_0, \rho_1), \tag{3.21}$$

and the switch dynamics is (dropping dependencies on space and time for simplicity)

$$\frac{\partial}{\partial t} \phi_i(0) = p_0 \phi_i(1) - p_1 \phi_i(0). \tag{3.22}$$

Summing over all cells, obtain for the total density of resting cells

$$\frac{\partial}{\partial t} \Phi = p_0 \Psi - p_1 \Phi. \tag{3.23}$$

Note that this is a non-local, non-linear set of PDEs, which are difficult to treat analytically. However, one can further simplify the analysis by making another approximation, assuming that the sensing radius ℓ is much smaller than the total domain size. Then, one can replace the non-local sensing function $\xi(\mathbf{x_j}(t) - \mathbf{x_i}(t), t)$ by a local delta distribution

$$\xi(\mathbf{x_j}(t) - \mathbf{x_i}(t), t) \to \delta(\mathbf{x_j}(t) - \mathbf{x_i}(t)). \tag{3.24}$$

In this case, the expected number of sensed cells in the microenvironment around $\mathbf{x}$ simply becomes the local density of the respective phenotype

$$\rho_0(\mathbf{x}, t) := \langle N_i^0(\mathbf{x}, t) \rangle = \frac{\Phi(\mathbf{x}, t)}{\ell^d}, \tag{3.25}$$

$$\rho_1(\mathbf{x}, t) := \langle N_i^1(\mathbf{x}, t) \rangle = \frac{\Psi(\mathbf{x}, t)}{\ell^d}. \tag{3.26}$$

Finally, the rate $\bar{W}_{10}$ reduces to

$$\bar{W}_{10} = p_0(\rho_0, \rho_1) = \frac{1}{1 + e^{\beta \Delta S}} = \frac{1}{1 + \left[\frac{\rho_0 \left(\rho_1 - \frac{1}{V} \right)}{\rho_1 \left(\rho_0 - \frac{1}{V} \right)} \right]^{\frac{\beta}{2}}}, \tag{3.27}$$

and the reversed transition probability $\bar{W}_{01}$ is

$$\bar{W}_{01} = p_1(\rho_0, \rho_1) = 1 - \frac{1}{1 + e^{\beta \Delta S}} = 1 - \frac{1}{1 + \left[\frac{\rho_0 \left(\rho_1 - \frac{1}{V} \right)}{\rho_1 \left(\rho_0 - \frac{1}{V} \right)} \right]^{\frac{\beta}{2}}}, \tag{3.28}$$

where the volume is defined as $V = \ell^d$.

3.3.2. Cell migration dynamics

In this section, the macroscopic equation has been derived in two spatial dimensions for the motile cell population. As before it can be written in the corresponding stochastic Langevin's equations as

$$
\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x_i} = \bar{v}\mathbf{v_i}(\theta_i),
$$
$$
\frac{\mathrm{d}}{\mathrm{d}t}\theta_i = \frac{1}{\bar{v}}\xi_i^\theta(t).
$$

(3.29)

This process can be considered as a special kind of active Brownian motion. In this kind of Langevin's equation, the stochastic force creates variations of orientation. According to [108] one can derive the corresponding Fokker-Planck equation for migrating cells using adiabatic elimination and averaging it over different noise realizations obtaining the following diffusion equation

$$
\frac{\partial \rho_1(x,t)}{\partial t} = \frac{\bar{v}^4}{2D_\theta}\nabla^2\rho_1(x,t).
$$

(3.30)

Here D_1 is denoted as diffusion coefficient which is $\frac{\bar{v}^4}{2D_\theta}$.

3.3.3. Superposition of phenotypic switching dynamics and cell migration

Combining the results in the previous sections, someone can easily formulate a system of PDEs as

$$
\partial_t\rho_0(x,t) = \nu E(\rho_0,\rho_1) \tag{3.31}
$$
$$
\partial_t\rho_1(x,t) = D_1\nabla^2\rho_1 - \nu E(\rho_0,\rho_1) \tag{3.32}
$$
$$
E(\rho_0,\rho_1) = \bar{W}_{10}(\rho_0,\rho_1)\rho_1 - \bar{W}_{01}(\rho_0,\rho_1)\rho_0 \tag{3.33}
$$

where $E(\rho_0,\rho_1)$ is the phenotypic exchange term and ν is the corresponding timescale ratio of the switching and diffusion processes, i.e., $\nu = \frac{\tau_S}{\tau_D}$. To ensure the numerical consistency of the above system, resting cells diffuse in a very slow manner i.e., $D_0 \ll D_1$ has been assumed, which results in the following reaction-diffusion system of equations

$$
\partial_t\rho_0(\mathbf{x},t) = D_0\nabla^2\rho_0 + \nu E(\rho_0,\rho_1), \tag{3.34}
$$
$$
\partial_t\rho_1(\mathbf{x},t) = D_1\nabla^2\rho_1 - \nu E(\rho_0,\rho_1). \tag{3.35}
$$

3.4. Spatio-temporal dynamics of cell migration/proliferation plasticity

In this section, the mean-field approximation is studied of the aforementioned stochastic plasticity dynamics for different regimes of the sensitivity β and the inter-

action radius ℓ. In particular, when cells have a very large interaction radius ($\ell \gg 1$) and then a finite one, one can study the system dynamics. Finally, the special case $\beta = 0$ has examined, i.e., cells decide independently of their microenvironment.

3.4.1. Case I : Large interaction radius

Here, the very large interaction radius system ($\ell \gg 1 \implies V \gg 1$) is focused. Although, this parameter regime is not biologically relevant, it is very instructive since it allows us to derive analytical estimates for the system's dynamics. Here, the macroscopic system is generalized by adding proliferation dynamics (logistic growth). the phenotypic switch dynamics are recapitulated by setting the proliferation rate to zero, i.e.,$r = 0$. The full system reads

$$
\begin{aligned}
\frac{\partial \rho_0}{\partial t} &= D_0 \nabla^2 \rho_0 + \nu E_0 \left(\rho_0, \rho_1\right) + r \rho_0 \left(1 - \rho_1 - \rho_0\right), \\
\frac{\partial \rho_1}{\partial t} &= D_1 \nabla^2 \rho_1 - \nu E_0 \left(\rho_0, \rho_1\right).
\end{aligned}
\tag{3.36}
$$

In turn, a non-dimensionalization of eqns. (3.36) is conducted to identify the variables. Moreover, this helps us to gain a knowledge about the relationships between the different model parameters. By assuming that the system size is fixed at L the non-dimensional quantities read

$$
\begin{aligned}
x^* &= \frac{x}{L} \implies \frac{\partial^2}{\partial x^{*2}} = L^2 \frac{\partial^2}{\partial x^2}, \\
t^* &= \frac{D_0 t}{L^2} \implies \frac{\partial}{\partial t^*} = \frac{L^2}{D_0} \frac{\partial}{\partial t}, \\
\gamma &= \frac{L^2 \nu}{D_0}, \\
D &= \frac{D_1}{D_0}, \\
r' &= \frac{r}{\nu}.
\end{aligned}
\tag{3.37}
$$

In the limit $V \gg 1$, the Eq. (3.36) can also be written as

$$
\begin{aligned}
\frac{\partial \rho_0}{\partial t^*} &= \nabla^{*2} \rho_0 + \gamma \left(\left(\rho_1 - \rho_0\right) \left(\frac{1}{2} - \tilde{\beta} \frac{\left(\rho_1 + \rho_0\right)}{\rho_1 \rho_0} \right) + r' \rho_0 \left(1 - \rho_1 - \rho_0\right) \right), \\
\frac{\partial \rho_1}{\partial t^*} &= D \nabla^{*2} \rho_1 - \gamma \left(\left(\rho_1 - \rho_0\right) \left(\frac{1}{2} - \tilde{\beta} \frac{\left(\rho_1 + \rho_0\right)}{\rho_1 \rho_0} \right) \right).
\end{aligned}
\tag{3.38}
$$

To understand the behavior of the system at long times, a fixed point analysis has conducted. Initially, one assume a well-stirred system, i.e., no spatial interactions. Then, Eqs. (3.38) can be written as coupled non-linear ODEs which have three fixed

points

$$\left(\rho_0^{A*}, \rho_1^{A*}\right) = \left(\frac{1}{2}, \frac{1}{2}\right),$$

$$\left(\rho_0^{B*}, \rho_1^{B*}\right) = \left(\frac{1}{2}\left(1 + \sqrt{1 - 8\tilde{\beta}}\right), \frac{1}{2}\left(1 - \sqrt{1 - 8\tilde{\beta}}\right)\right), \tag{3.39}$$

$$\left(\rho_0^{C*}, \rho_1^{C*}\right) = \left(\frac{1}{2}\left(1 - \sqrt{1 - 8\tilde{\beta}}\right), \frac{1}{2}\left(1 + \sqrt{1 - 8\tilde{\beta}}\right)\right).$$

The above imply a pitchfork bifurcation for the $\tilde{\beta} = \frac{\beta}{8V}$ parameter, i.e., there exists a critical value $\tilde{\beta}_c$ that introduces the bistable state. From Eqs.(3.39), one can easily deduce that the critical sensitivity value

$$1 - 8\tilde{\beta}_c = 0 \iff \beta_c = V. \tag{3.40}$$

This is an acceptable approximation even for the finite interaction radius system (see next section). In the following, the system is analyzed if it is able to produce spatial patterns. Applying linear stability analysis for the spatially resolved system Eqs. (3.38), one can deduce that no pattern formation is possible (for details check next section). Any perturbations to the homogeneous state lead always to a spatially homogeneous steady state. When $r = 0$ this result is consistent with the findings in discrete IBM simulations where very large interaction radii do not confer any clustering, as shown in Fig. 3.3ii.

3.4.2. Case II : Finite interaction radius

Now let's turn to the full system for intermediate interaction radius (also assuming proliferation). This implies a finite interaction volume, V and for analytical feasibility are interested in the Gaussian approximation of the switching probabilities and their corresponding mean-field terms. The full system of PDEs assuming also proliferation reads

$$\frac{\partial \rho_0}{\partial t^*} = \nabla^{*2} \rho_0 + \gamma \left(\frac{\rho_1 - \rho_0 \left(\frac{\rho_0 \left(\rho_1 - \frac{1}{V}\right)}{\rho_1 \left(\rho_0 - \frac{1}{V}\right)} \right)^{\frac{\beta}{2}}}{1 + \left(\frac{\rho_0 \left(\rho_1 - \frac{1}{V}\right)}{\rho_1 \left(\rho_0 - \frac{1}{V}\right)} \right)^{\frac{\beta}{2}}} + r' \rho_0 \left(1 - \rho_1 - \rho_0\right) \right),$$

$$\frac{\partial \rho_1}{\partial t^*} = D \nabla^{*2} \rho_1 - \gamma \frac{\rho_1 - \rho_0 \left(\frac{\rho_0 \left(\rho_1 - \frac{1}{V}\right)}{\rho_1 \left(\rho_0 - \frac{1}{V}\right)} \right)^{\frac{\beta}{2}}}{1 + \left(\frac{\rho_0 \left(\rho_1 - \frac{1}{V}\right)}{\rho_1 \left(\rho_0 - \frac{1}{V}\right)} \right)^{\frac{\beta}{2}}}. \tag{3.41}$$

Since closed expression of steady states are not analytically feasible, one obtain the bifurcation diagram numerically, see Fig (3.4i). The existence of a supercritical pitchfork bifurcation and the existence of a critical β_c can be observed. For $\beta \geq \beta_c$,

(i)   **(ii)**
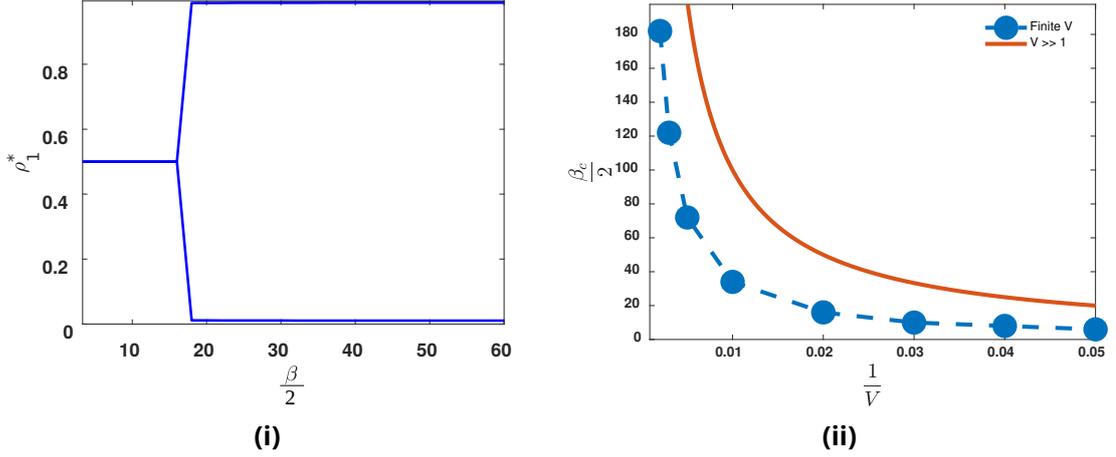
Figure 3.4.. (**a**) Bifurcation diagram for moving cells in the finite interaction radius limit with respect to β, where proliferation rate r is fixed to 0 and V^{-1} at 0.02. The bifurcation diagram is symmetric for resting cells, i.e.,ρ_0^*. (**b**) The critical point value of the bifurcation diagram (where the solution splits in two branches) is plotted against the inverse volume $\frac{1}{V}$.

the systems depart from balanced state ($\rho_0 = \rho_1 = \frac{1}{2}$) to the coexistence of a "fluid" (most cells migrate) and a "solid" (most cells are resting) phase. The switch is controlled by the perturbation on the ratio of migratory and resting cells. Interestingly, someone can compare the analytic estimate of $\beta_c = V$, from the $V \gg 1$ case, with the one calculated for the finite V system. Fig. (3.4ii) shows that the infinite system approximation provides an upper bound for β_c, which is not too far from the real value of finite systems. The existence of a critical β_c with respect to the two phases is evident in the IBM simulations as well. In particular, if the ratio of stationary over motile cells is quantified, a similar behavior of the critical sensitivity for increasing V is observed, since it decreases (see Fig.7.5 in SI). However, one cannot conduct a strict quantitative comparison since the bifurcation analysis does not involve any diffusion, as opposed to the IBM simulation.

In turn, linear stability analysis is applied to identify parameter regimes that promote diffusion driven pattern formation or Turing instability. To analyze the Turing instability [115] have to find the system's steady state, (i.e., when diffusion is not present in the systems of Eq.(3.36)). It has been shown that $d = \frac{D_1}{D_0} \gg 1$ is a necessary condition for the emergence of spatially heterogeneous solution, i.e., patterns. Now, the system of PDEs (i.e., Eq.(3.36)) can be written in a generalized matrix form

$$\frac{\partial \boldsymbol{\rho}}{\partial t} = \boldsymbol{D}\nabla^2 \boldsymbol{\rho} + \gamma \boldsymbol{R}\boldsymbol{\rho}, \quad \boldsymbol{D} = \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix},$$

$$\boldsymbol{R} = \begin{pmatrix} \frac{\partial E_0(\rho_0,\rho_1)}{\partial \rho_0} + r'(1 - \rho_1 - \rho_0) - r'\rho_0 & \frac{\partial E_0(\rho_0,\rho_1)}{\partial \rho_1} - r'\rho_0 \\ -\frac{\partial E_0(\rho_0,\rho_1)}{\partial \rho_0} & -\frac{\partial E_0(\rho_0,\rho_1)}{\partial \rho_1} \end{pmatrix}_{(\rho_0^*,\rho_1^*)}, \tag{3.42}$$

$$= \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}_{(\rho_0^*,\rho_1^*)},$$
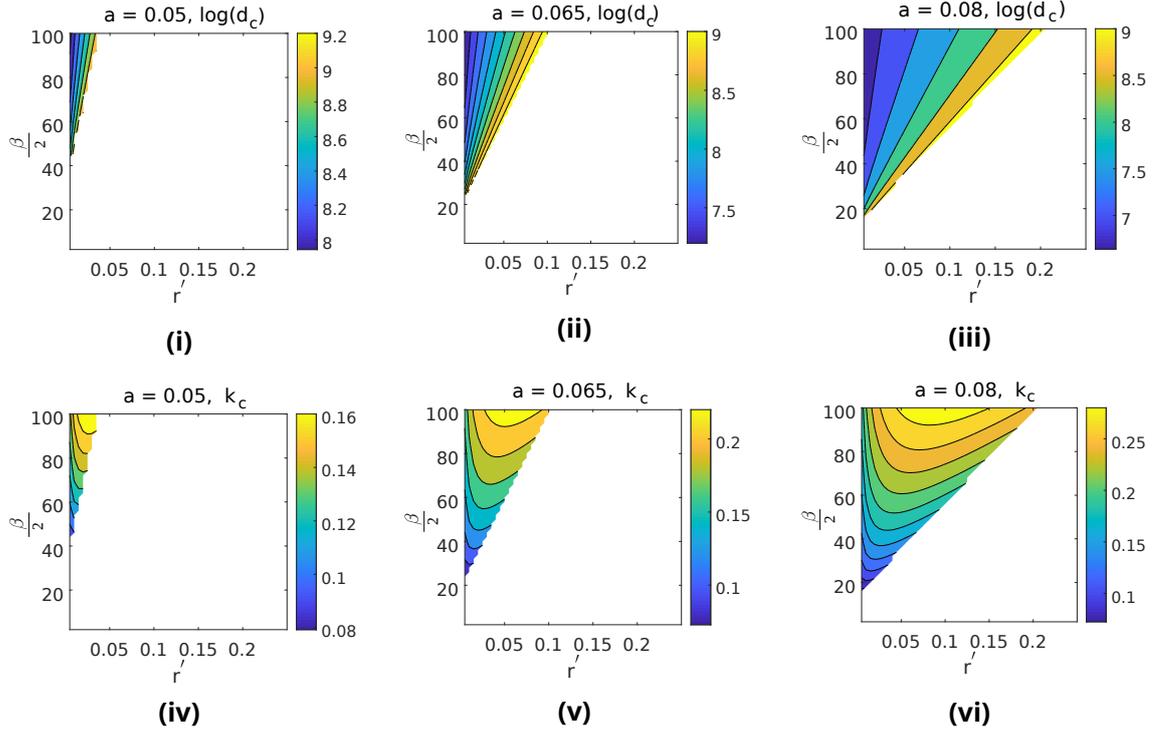
Figure 3.5.. Spatio-temporal pattern formation in dependence of sensitivity $\beta/2$, proliferation rate r' and inverse sensing volume $a = \frac{1}{V}$. **(a - c)** The critical diffusion coefficient $d_c = D_1/D_0$ decreases for increasing sensitivity or increasing proliferation rate. **(d - f)** The critical wave number k_c of the observed pattern increases for an increasing sensitivity.

where ρ is defined as $(\rho_0 - \rho_0^*, \rho_1 - \rho_1^*)^T$ and R is the Jacobian at $\rho^* = (\rho_0^*, \rho_1^*)$. Using the Turing conditions of instability [115], patterns can be found when N is finite in zero-flux boundary conditions. All the Turing instability conditions have been checked using the given relations:

$$f_u + g_v < 0, \qquad f_u g_v - f_v g_u > 0,$$
$$D f_u + g_v > 0, \qquad (D f_u + g_v)^2 - 4D(f_u g_v - f_v g_u) > 0. \tag{3.43}$$

Interestingly, only when the density of resting cells is larger than that of the moving cells, patterns are formed under Turing instability conditions. In order to investigate the system's potential to exhibit pattern formation, the range of validity of the Turing instability criteria (3.43) is checked. Diffusion-driven instability conditions are satisfied for a large portion of the parameter space. In turn, for these parameters, calculate the critical diffusion coefficient d_c. For values $d > d_c$ are able to observe patterns. This condition is generally fulfilled as the resting cells are only moving passively due to external noise and cell-cell interactions like cell-cell adhesion, while the migratory cells also move actively. The existence of d_c is associated with the existence of a critical wave number k_c [115].

$$d_c = \frac{-2(2f_v g_u - f_u g_v) \pm \sqrt{(2(2f_v g_u - f_u g_v))^2 - 4f_u^2 g_v^2}}{2f_u^2},$$
$$k_c = \gamma \left[\frac{Det(A)}{d_c} \right]^{\frac{1}{2}}. \tag{3.44}$$
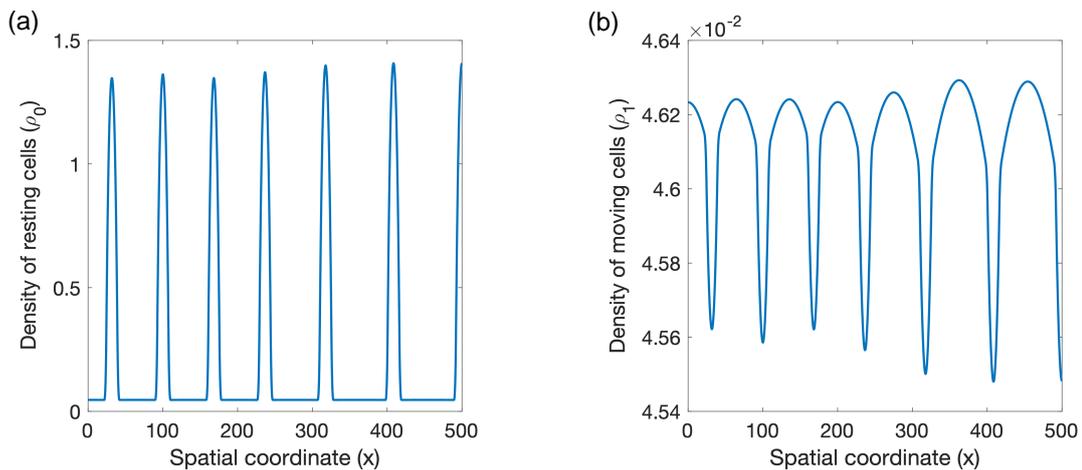
Figure 3.6.. Pattern formation in the mean-field model. Densities of proliferating (**a**) and migrating (**b**) cells at $t = 50000$. Starting from an almost homogeneous initial state with a sinusoidal perturbation, the resting cells form a periodic pattern of high-density spikes, while the density of moving cells has minima at the position of the high-density spikes of the resting cells.

In Fig. (3.5) the parameter regime is identified such that it allows observing patterns and then corresponding d_c and k_c.

By fixing the initial conditions to a cosine wave, one can observe in one dimension the existence of regular spikes as shown in Fig (3.6). Please note that the exact pattern is sensitive to the initial conditions. In turn, the system in 2D is simulated and for the same parameters and patterning in the form of dots is observed. Finally, one can investigate if the type of patterns changes for variations in parameters β and r' (see Fig. (3.7)). If $d > d_c$ then need a large domain or size of the system to observe the patterns. In the 2D case, the discoidal patterns of resting phenotype have observed, which resemble the 1D case. So, the radius of the circles of the patterns are increased if someone fixes the domain size. Moreover, one observes in both dimensions (i.e.,1D and 2D) the critical spatial frequency increases with decreasing r'.

In comparison with the discrete IBM simulations, the simulation clusters can be identified by the discoidal mean-field patterns. Under this statement, one find that:

1. There exists a critical β that allows for the emergence of patterns, as in the IBM simulations.

2. For very low and very high interaction radius ℓ, observe no patterns, which is consistent with the IBM results (see Fig.3.3ii).

33
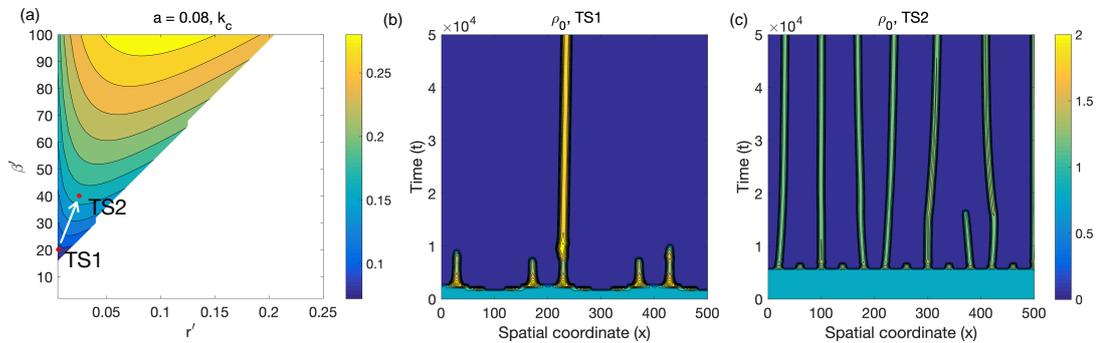
Figure 3.7.. Spatio-temporal pattern formation of the mean-field model. **(a)** Turing space and critical wave number. **(b - c)** Density of resting cells over time. **(b)** A single high-density spike forms in agreement with the prediction of a small critical wave length for the given parameters (red dot TS1 in **(a)**). **(c)** For a higher sensitivity and proliferation rate, a periodic pattern with smaller wave length emerges, corresponding to the prediction of a larger critical wave number (red dot TS2 in **(a)**). Here $\beta' = \frac{\beta}{2}$ and $a = \frac{1}{V}$.
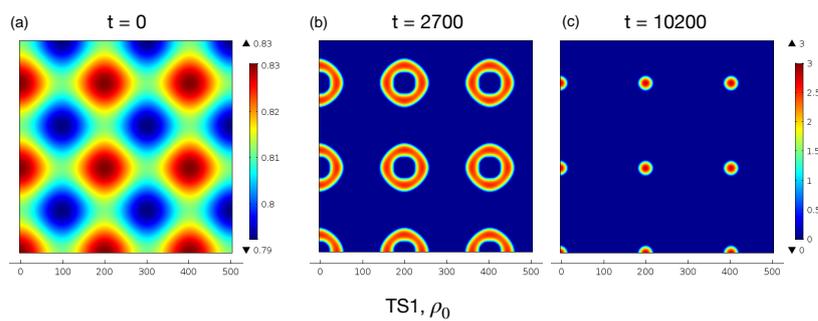

Figure 3.8.. Two-dimensional simulation of the mean-field approximation with finite interaction radius. (a) Initial state, (b) $t = 2700$, (c) $t = 10200$ corresponding to a stable steady state. Parameters refer to point (TS1) in the Turing space, as displayed in Fig. (3.7).

3.4.3. Phenotypic switching dynamics in the absence of microenvironmental sensing

Finally, the case $\beta = 0$ has been investigated where cells do not sense their microenvironment. So, they end up with the following system,

$$
\begin{aligned}
\frac{\partial \rho_0}{\partial t^*} &= \nabla^{*2} \rho_0 + \gamma \left(\frac{(\rho_1 - \rho_0)}{2} + r' \rho_0 \left(1 - \rho_1 - \rho_0 \right) \right), \\
\frac{\partial \rho_1}{\partial t^*} &= D \nabla^{*2} \rho_1 - \gamma \frac{(\rho_1 - \rho_0)}{2}.
\end{aligned}
\tag{3.45}
$$

If one do a fixed point analysis similar to equation (3.38). Assuming no spatial dynamics, two fixed points can be found i.e.,

$$
\begin{aligned}
\left(\rho_0^{P*}, \rho_1^{P*} \right) &= \left(\frac{1}{2}, \frac{1}{2} \right), \\
\left(\rho_0^{Q*}, \rho_1^{Q*} \right) &= (0, 0).
\end{aligned}
\tag{3.46}
$$

For further details see Fig. 7.1 in Supplementary Material 7 It can be clearly shown that the fixed point $(0,0)$ is unstable and $\left(\frac{1}{2}, \frac{1}{2} \right)$ is a saddle point. One can also write the coupled PDE equations in a single PDE as ll, if $\rho_0 = \rho_1 = \tilde{\rho}$. Then the equation can be obtained as

$$
\frac{\partial \tilde{\rho}}{\partial t^*} = \frac{(1+D)}{2} \nabla^{*2} \tilde{\rho} + \frac{r}{2} \tilde{\rho} \left(1 - 2\tilde{\rho} \right).
\tag{3.47}
$$

Now, clearly it can be seen that the Eq.(3.47) is similar to the Fisher-Kolmogorov equation [115]. It is known that the Fisher-Kolmogorov equation does not exhibit any pattern formation instabilities. The latter observation is consistent with the discrete IBM simulations, where no clustering is observed.

3.5. Summary and outlook

Recently, the Least microEnvironmental Uncertainty Principle (LEUP) has been proposed as an organization principle for cell decision-making in multicellular systems. In this chapter, this principle has been applied to shed light on the effects of phenotypic plasticity for tissue dynamics with a mathematical model. Two types of plasticity has been focused: a Go-or-Rest, and a Go-or-Grow phenotypic dichotomy, which play key roles in important processes in biological development and pathological situations as cancer invasion. Someone can take an assumption such that in any given spatial mesoscopic sample, the presence of cell phenotypes follows a binomial distribution. Using this assumption, one can calculate the microenvironmental entropy and the LEUP-driven probability distribution of each phenotype. On the basis of this distribution, an appropriate microscopic stochastic model has been defined for the spatio-temporal dynamics of both phenotypes, and in turn derived the corresponding mean-field description resulting in a system of coupled reaction-diffusion equations. The main results of the study are: (i) in the case of

3. Phenotypic plasticity : dynamics at the level of tissue from individual cell decisions

Go-or-Rest plasticity, there exists a supercritical pitchfork bifurcation that defines a switch between a "fluid" and a "solid" phase, and (ii) in the case of Go-or-Grow plasticity, for the "solid" phase, one can derive conditions for the emergence of Turing patterns. Interestingly, the existence of Turing patterns requires a critical LEUP sensitivity to the microenvironment and a minimum interaction range. Please note, that in this work a clustered, immobile configuration is defined as "solid" and the homogeneous mobile configuration is defined as "fluid". In physics, solid states are characterized by deep "valleys" in their radial distribution function, separated by characteristic distances controlled by the underlying repulsive and adhesive forces. Here, the radial distribution function of the "solid" state does not like physical analogs because cell-cell adhesions and volume exclusions are not included in the cell dynamics. Adding such interaction forces to the corresponding IBM can be included which would result in classical solid and fluid radial distribution functions, and subject to future work.

The model assumes binary transitions between two discrete phenotypes, which is an extreme simplification of biological reality. Certainly, one could include a continuous spectrum of motile/proliferative phenotypes that is expected to imply even richer spatio-temporal dynamics. Indeed, assuming a continuous state (velocity) space would potentially lead to further interesting bifurcations, such as the metastable EMT state as found in [84, 19], or complex spatio-temporal patterns as indicated in [11].

Interestingly, the LEUP-driven IBM is an extension of a Vicsek-type model [162], which was formulated in the context of self-propelled particles [62]. This model exhibits a novel collective behavior when compared to the past published results from Vicsek-type of models. In particular, to the knowledge it is the first time to produce with such models Turing patterns, i.e.,dynamics clusters of non-motile cells of specific characteristic wavelength. Typically, in Viscek-type models may observe moving clusters of swirling cells (e.g. the milling Viscek model) but never static ones.

At this point, one can focus on the biological assumptions and implications of the study. The molecular regulatory mechanisms involved in EMT or GoG remain largely unknown, where the latter can be viewed as an EMT with proliferation constrained to the epithelial/resting phase. Here, the phenotypic regulation of both mechanisms is based on the decreasing of microenvironmental entropy in physiological tissues is assumed precisely, which helps to predict the multicellular spatio-temporal dynamics. This assumption is supported by the fact that healthy physiological processes, biological development or processes like wound healing, where EMT/MET or GoG are present, typically lead to an ordered (low entropy) tissue from a disordered initial condition. On the other hand, deregulation of EMT and GoG have been already identified as pivotal elements in invading cancers [6, 65], where genetic and phenotypic heterogeneity, characterized by high entropy, is a key characteristic. Assuming a LEUP-driven migration/proliferation, phenotypic regulation helps to understand how cells control multicellular dynamics in terms of growth and patterning.

Implications in multicellular growth control: The central finding of this study is associated with the bifurcation diagram of the LEUP sensitivity parameter β in Fig. 3.4i.

As stated before, the parameter β quantifies how prone cells are in sensing and responding to their microenvironmental stimuli. When $\beta = 0$ cells migrate and proliferate independent of their microenvironmental cues, which corresponds to one of the cancer hallmarks [64]. In this case, the systems grows uncontrollably, resembling a cancerous tissue [64]. The resulting Fisher-Kolmogorov macroscopic behavior has been prototypically used to model invading tumors [39]. Moreover, for low β LEUP theory predicts no spatial order, in terms of clustering implying "flat" radial distribution functions (see Figure (3.2ii)). This finding is confirmed by the spatial analysis of patient glioma biopsy samples in Jiao et al. [81]. On the other hand, by adding a death process for any motility state in the GoG model, one can recapitulate the Allee effect (bistability between extinction and growth) as found in Boettger et al. [22]. A cell population in the "fluid" state will go extinct (motile cells do not proliferate but still have a probability to die) whereas systems in the "solid" state will always grow until carrying capacity. Therefore, in the bistability regime, cell sensing properties lead to multicellular growth control.

Implications in multicellular pattern control: Increased cell sensing β represents the physiological tissue dynamics, since it allows the system to control its behavior. By tuning β and the ratio of motile/resting cells, the system exhibits a bistable behavior between a "fluid"/mesenchymal-type and a "solid"/epithelial-type tissue phase. This kind of tissue level switch is of utmost importance in physiological processes such as wound healing or embryogenesis [8]. After a tissue injury, the healing is characterized by a "fluid" diffusive expansion of fibroblast cells, that adopt a migratory phenotype via EMT [167]. After covering the wound, the "solid" phase emerges as cells stop the migration program and proliferate to finalize tissue repair. In the absence of proliferation, the bistable switch from the "solid" to the "fluid" state could potentially explain the jamming phase transition observed in epithelial colonies, under EGF modulation and Rab5a knock-out [102]. Typically, the "solid" multicellular phase is prone to the emergence of pattern formation, which frequently occurs in physiological epithelial tissues [167, 67]. When EMT is combined with a Notch-Delta cell-cell communication, then epithelial/immotile cell clusters emerge [19], as observed in the IBM and mean-field simulations. Adding proliferation in the GoG model, the type and the size of such emerging patterns require a tight regulation of microenvironmental sensing β and proliferation rate r, as indicated by Fig. 7.3i, since the critical wavelength k_c depends on the ratio $\frac{\beta}{r}$. Such Turing patterns are in agreement with previous GoG studies [125], where Turing patterns are emerging. At this point, one would like to outline some ideas on the experimental validation of LEUP. An ideal experimental system would involve a cell population exhibiting either EMT/MET or migration/proliferation plasticity. As an example of EMT/MET plasticity, an epithelial cell colony can be choosen. The main premise of LEUP is that cells equip Bayesian inference to decide over their phenotypes, expressed as a combination of a sensed microenvironmental distribution of external variables and a phenotypic prior (cell intrinsic variable). As internal variables can choose the cell motility and proliferation rates. The former can be measured by particle image velocimetry (PIV) and the latter by live imaging or Ki67 (or BrdU) staining. Also, one could pharmacologically control cell proliferation. As extrinsic variables, one can assume the local cell density and in particular the corresponding fluctuations. The second moment of the local cell density depends on

the average local cell density (first moment) and can be manipulated by localized cell ablation or by mixing with non-cellular material, such as collagen, thus creating different spatial cell distributions. After splitting the experiments in training and test data, one needs to fit the LEUP parameters, i.e., the sensitivity to the microenvironment and the interaction radius, according to training data in order to predict the experimentally measured test data. A good basis for the approach are experiments by Puliafito [129], which show that increasing density in a colony of epithelial kidney cells may induce transitions from the mesenchymal to the epithelial phenotype (MET). In the future, one could also investigate the inverse transition and control the changes in the local cell density distribution. Such an experimental setting would allow us to support the LEUP hypothesis.

The focus of this chapter is the application of the LEUP idea to study collective effects of different types of phenotypic plasticity allowing reversible adaptations to change environmental conditions. However, the LEUP idea can also be applied to analyze collective effects of irreversible cell fate determination playing a key role in cell differentiation and development. In particular, early embryonic tissue consists of a mix of proliferating pluripotent progenitor cells [94].]. In this case, the cellular microenvironment possesses a high phenotypic entropy, since all possible phenotypes are still equiprobable. Subsequently, spatial mechanochemical perturbations produce appropriate microenvironmental changes leading to so-called stem cell niches that allow for cell differentiation [85]. Pluripotent cells sense a broad spectrum of microenvironmental cues aiming to find an appropriate differentiation niche. As pluripotent cells further specialize, they irreversibly fine-tune their sensing machinery to adapt to the corresponding microenvironment, which further differentiates to a specific tissue (implying decreasing of microenvironmental entropy over time). A particularly important cell fate specification mechanism is the Notch-Delta interaction [5], where neighboring cells use lateral inhibition to adapt distinct fates (high Delta low Notch and vice versa). The collective effect of this mechanism is the evolution of the system to low entropy organized patterns [21].

In conclusion, this study shows how individual LEUP-driven cell decisions the dynamics at the tissue level and how knowledge of collective cell decision-making can be used to control of growth and pattern.

4. Cellular orientation decisions: origin of pattern formations in collective cell migrations

In this chapter, [1] the LEUP based theory on collective cell migration has been represented. One can think of migration as an active decision-making process. Often, neither biophysical nor phenomenological models are able to provide a plausible explanation or quantitative reproduction of collective migration patterns, due to the lack of complete mechanistic knowledge. Such an example is the spatio-temporal dynamics of spherical *S. marcescens* bacteria. Interestingly, prior modeling works ([4]) were able to partially reproduce the experimental results, since the underlying biophysical mechanisms are still unclear. In such cases, one could rely on machine/statistical learning methods that circumvent the biophysical details ([117];[170];[120]). However, such methods are typically of high accuracy but low interpretability, i.e., they are "black boxes" that do not offer mechanistic insights, and prone to overfitting. Applying the LEUP to collective cell migration, one can aspire (i) to provide a low-dimensional statistical mechanics description, (ii) circumvent the uncertainty about the underlying biophysical mechanisms and (iii) provide a relationship to phenomenological models (e.g., the Vicsek model).

4.1. Mathematical framework

Moving and interacting cells are modeled by a two-dimensional self-propelled particle model (SPP). In this model, $N \in \mathbb{N}$ cells move on a two-dimensional area. The i-th cell is characterized by its position, $\vec{r}_i \in \mathbb{R}^2$, speed, $v_i \in [0, \infty) \subset \mathbb{R}$, and an orientation $\theta_i \in [0, 2\pi) \subset \mathbb{R}$. Due to the small size of cells, it is assumed that viscous forces dominate. Changes in speed and orientation result from local potentials, $U_\theta(\vec{r}_m, \theta_m, v_m), U_v(\vec{r}_m, \theta_m, v_m) : \mathbb{R}^2 \times [0, 2\pi) \times [0, \infty) \mapsto \mathbb{R}$ which depend on the positions and polar velocity components of cells within a radius $R \in \mathbb{R}_+$. The bias of the cell to follow the potential gradients is regulated by the parameters $\beta_\theta, \beta_v \in \mathbb{R}$, called angular and radial sensitivities, respectively. Additionally, velocity fluctuations occur due to stochastic noise terms $\xi_i^\alpha(t) \in [0, 2\pi)$, $\alpha \in \{\theta, v\}$ where $t \in \mathbb{R}_+$ denotes time. The noise will be assumed to be a zero-mean, white noise term, which has the statistical properties $\langle \xi_i^\alpha(t) \rangle = 0$ and $\langle \xi_i^\alpha(t_1) \xi_m^\alpha(t_2) \rangle = 2D_\alpha \delta(t_1 - t_2) \delta_{im}$, where t_1 and t_2 are two time points, $D_\alpha \in \mathbb{R}_+$ is either the angular ($\alpha = \theta$) or radial ($\alpha = v$) diffusion coefficient, $\delta(t)$ is the Dirac delta, and δ_{im} is the Kronecker delta. Finally, the radial acceleration will be assumed to be damped by a density dependent friction, $\psi(\rho_i)$. In the following, it will be assumed that the density-dependent friction is given by $\psi(\rho_i) = \rho_i - \bar{\rho}$, where ρ_i is the local cell density within the i-th cell's interaction radius, and $\bar{\rho}$ is the global average cell density. Taking everything into account, the stochastic equations of motion of the i-th cell reads as [135]

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{r}_i = v_i \vec{v}(\theta_i) \tag{4.1a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_i = -\beta_\theta \frac{\partial}{\partial \theta_i} U_\theta(\vec{r}_m, \theta_m, \vec{v}_m) + g(\vec{v}_i)\xi_i^\theta(t) \tag{4.1b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}v_i = -\beta_v \frac{\partial}{\partial v_i} U_v(\vec{r}_m, \theta_m, \vec{v}_m) - \epsilon\psi(\rho_i)v_i + \xi_i^v(t). \tag{4.1c}$$

where $\vec{v}(\theta_i)$ is the normalized velocity of the cell and ϵ is a parameter. A representation of the SPP model is shown in Fig. 4.1. The function $g(\vec{v}_i)$ modulates the noise variance and allows us to model certain distributions (such as the Rayleigh distribution in Section 5).

The interaction potentials $U_i(\vec{r}_m, \theta_m, \vec{v}_m)$, which dictate the velocity dynamics of cells, need to be specified. Biophysically, the potentials should encompass steric effects, hydrodynamic interactions, chemotactic effects, and terms arising from internal cellular processes, for example, flagella motor dynamics, actin polymerization, receptor dynamics, etc. Finding such potentials is a formidable task, since not all of the mechanisms and interactions involved are known. To circumvent this problem, a variational principle of cell decision-making related to entropy maximization [16], known as the least microenvironmental uncertainty principle (LEUP), will be used [66]. In the next section, such a case will be discussed.
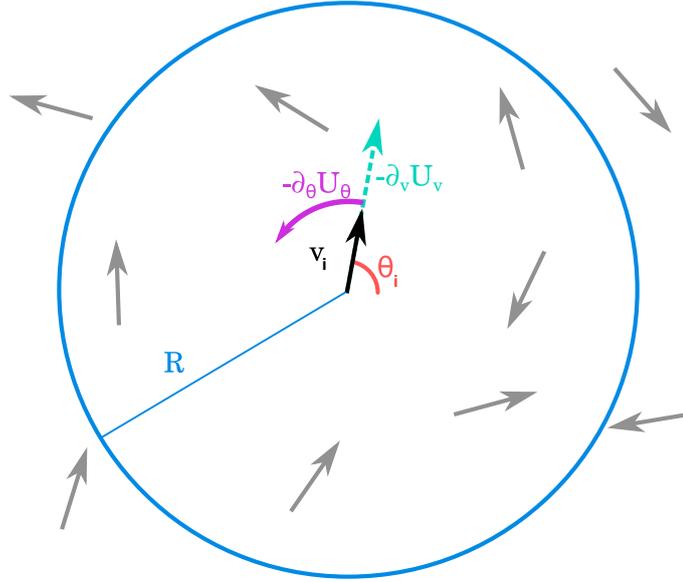
Figure 4.1.. Graphic representation of the dynamics of the SPP model. The i-th cell is represented by a point particle with speed v_i and orientation θ_i. Depending on the form of the interaction potential, the cell may feel a reorientation force $-\partial_\theta U_\theta$ and a radial force $-\partial_v U_v$ due to interaction with other cells inside the interaction neighborhood defined by the radius R.

4.1.1. Self-propelled particle model with leup based decision-making

The internal energy depends on the internal states of the cell, as well as the internal states of other cells in the surroundings. Such internal states can be a vector of physical quantities (e.g., velocity, acceleration) and/or chemical variables such as intracellular proteins, genes and so on. An interaction potential that models the equations of motion has been defined.

By doing so, it is evident that the responsiveness of the cells to LEUP can be quantified by the sensitivity $\beta_\alpha = -\beta$. Analogically, the equations of motion of the model can be written using Eq. (2.6) and Eq. (4.1c) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{r}_i = v_i \vec{v}\left(\theta_i\right) \tag{4.2a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_i = \beta_\theta \frac{\partial}{\partial \theta_i} S\left(\Theta_i \mid \theta_i\right) + g(\vec{v}_i)\xi_i^\theta(t) \tag{4.2b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}v_i = \beta_v \frac{\partial}{\partial v_i} S\left(V_i \mid v_i\right) - \epsilon \psi\left(\rho_i\right) v_i + \xi_i^v(t). \tag{4.2c}$$

To illustrate entropy calculation, it will be assumed that the orientations of cells within the interaction neighborhood are distributed according to

$$P\left(\vartheta \in \Theta_i \mid \theta_i\right) = \frac{\sinh \gamma}{2\pi\left[\cosh\left(\gamma\right) - \cos\left(\vartheta - \mu\right)\right]}, \tag{4.3}$$

where μ is the mean of the distribution and γ is a parameter related to the variance. This is a wrapped Cauchy distribution, periodic over the interval $[0, 2\pi]$. Similarly,

cell velocities will be assumed to be distributed half-normally,

$$P\left(v \in v_i \mid v_i\right) = \sqrt{\frac{2}{\sigma^2 \pi}} \exp\left(-\frac{v^2}{2\sigma^2}\right),$$
(4.4)

where σ^2 is proportional to the variance of the distribution. Accordingly, the angular entropy is

$$S\left(\Theta_i \mid \theta_i\right) = \ln\left(2\pi\right) + \ln\left(1 - e^{-2\gamma}\right),$$
(4.5)

while the velocity entropy is

$$S\left(V_i \mid v_i\right) = \frac{1}{2}\ln\left(\frac{\pi\sigma^2}{2}\right) + \frac{1}{2}.$$
(4.6)

The parameter σ can be determined from the local speed variance, while the parameter γ depends on the local polar order (i.e., the degree of parallel alignment) of cell velocities in the neighborhood. It should be noted that the qualitative behavior of the model is independent of the particular choice of distributions, and the distributions considered here are suggested only for ease of calculation. Before defining γ, first the observables will be defined to characterize the order of the velocity field.

4.1.2. Order parameters and observables

The normalized complex velocity of the i-th cell is denoted by $z_i \in \mathbb{C}$ as $z_i = e^{i\theta_i}$, where i is the imaginary unit. The k-th moment of the velocity over an area A is given by $\left\langle z^k \right\rangle_A = \frac{1}{N_A} \sum_{m \in A} z_m^k$, where the sum is over all cells in area A, and N_A is the total number of cells in A. The polar order parameter in the area A is given by

$$S_A^1 = \left|\langle z \rangle_A\right|,$$
(4.7)

which is the modulus of the first moment of the complex velocity in A, while the nematic order parameter in the area A is given by

$$S_A^2 = \left|\langle z^2 \rangle_A\right|,$$
(4.8)

which is the modulus of the second moment of the complex velocity in A. The order parameters are bounded, i.e.,

$$0 \leq S_A^1, S_A^2 \leq 1,$$
(4.9)

due to the complex velocities z_i being normalized. The parameter γ for the distribution of orientations in the neighborhood of the i-th cell is given by,

$$\gamma = -\ln\left(S_{C_{R,i}}^1\right),$$
(4.10)

where the subindices $C_{R,i}$ indicate a circular area of radius R centered at $\vec{r}_i$. The latter directly stems from the properties of the wrapped Cauchy distribution.

   While global polar and/or nematic order are characteristic of steady flows, rotating flow fields are commonly observed in out-of-equilibrium systems. The vorticity is an observable which is equal to twice the local angular velocity, and is thus a measure of the local strength and direction of rotation of the field. The vorticity ω is defined as

$$\omega\left(\vec{r}\right) = [\nabla \times \vec{v}_{\mathrm{mean}}\left(\vec{r}\right)] \cdot \vec{k}, \tag{4.11}$$

where $\vec{v}_{\mathrm{mean}}\left(\vec{r}\right)$ is the mean velocity field at a point $\vec{r}$, and $\vec{k}$ is the vector normal to the plane where cells move.

4.1.3. Statistical test

For the statistical evaluation of the results, the χ^2-test has been used. The testing hypothesis is that the experimental data are explained by the model predictions. To test it, one can construct

$$\chi_j^2 = \sum_{i=1}^{N}\left(\frac{\hat{O}_i^{(j)} - O_i^{(j)}}{\hat{\sigma}_i^{(j)}}\right)^2, \tag{4.12}$$

where $O_i^{(j)}$ is the experimental values of a certain observable j, being either speed or vorticity, and the $\hat{O}_i^{(j)}$ and $\hat{\sigma}_i^{(j)}$ is the corresponding mean value of the stochastic model predictions, based on an ensemble of 50 simulations for each density point $i = 1...N$, $N = 11$. The quantity $\frac{\hat{O}_i^{(j)} - O_i^{(j)}}{\hat{\sigma}_i^{(j)}}$ can be viewed as a z-score for each, $\hat{O}_i^{(j)}$ and for a large enough simulation ensemble should converge to a normal distribution. The total degrees of freedom for both observables is $2N = 22$. The calculate the reduced χ^2-statistic, or χ^2 per degree of freedom, which is defined as $\chi_{2N}^2 = (2N)^{-1}\sum_j \chi_j^2 = 1.97$ being close enough to 1. This suggests that the fitting is satisfactory, since values $\chi_{2N}^2 \gg 1$ indicate a bad fit to the experimental data.

4.2. Comparison with Vicsek model

By using the LEUP, interaction as a change in velocity dictated by the local entropy gradient has modeled. The modulation of β_α parameters modulates the response of cells to the local entropy gradient and gives rise to relationships with known phenomenological models, such as the Vicsek model. The absolute value $|\beta_\alpha|$ is proportional to the likelihood of the cell to change its velocity according to a given entropy gradient. For $\beta_\alpha < 0$, cells tend to go against the local entropy gradient towards the entropy minimum. In the specific case of $\alpha = \theta$, a negative sensitivity would restrict the distribution of angles to a narrow selection. Conversely, $\beta_\alpha > 0$ forces cells to follow the entropy gradient towards the entropy maximum, broadening the distribution. From here on, the effect of cell interactions is assumed, and it will be averaging the radial component, therefore $\beta_v < 0$.

4. Cellular orientation decisions: origin of pattern formations in collective cell migrations

To evaluate the effect of these two opposite migration strategies, the angular steady states are analyzed in the two parameter regimes. Without loss of generality, assume that, in the steady state, the mean velocity is $\bar{v} = 1$. By expanding $S^1_{C_{R,i}}$ using Eq. (4.7), defining the components of the mean neighborhood velocity as $\bar{v}_{y,i} = \sum_{C_{R,i} \ni m \neq i} \sin\theta_m$ and $\bar{v}_{x,i} = \sum_{C_{R,i} \ni m \neq i} \cos\theta_m$, and differentiating Eq. (4.5), find that the orientation of θ_i at the entropy extrema must be such that (see Supplementary Material 7)

$$\tan\theta_i = \frac{\bar{v}_{y,i}}{\bar{v}_{x,i}},$$

but $\frac{\bar{v}_{y,i}}{\bar{v}_{x,i}} = \tan\bar{\theta}$, the tangent of the mean orientation of the neighbors, excluding the i-th cell. This results in two extremum points $\theta_i = \bar{\theta}$ and $\theta_i = \bar{\theta} + \pi$, one where the velocity of the i-th cell is parallel to the average velocity of its neighbors, and one when it is antiparallel. In the first case,

$$\sin\theta_i \propto \bar{v}_{y,i} \text{ and} \tag{4.13a}$$
$$\cos\theta_i \propto \bar{v}_{x,i}, \tag{4.13b}$$

while in the second case

$$\sin\theta_i \propto -\bar{v}_{y,i} \text{ and} \tag{4.14a}$$
$$\cos\theta_i \propto -\bar{v}_{x,i}. \tag{4.14b}$$

It can be shown (see Supporting information) that $\theta_i = \bar{\theta}$ corresponds to an entropy minimum, while $\theta_i = \bar{\theta} + \pi$ corresponds to an entropy maximum. Consequently, the behavior of **the regime $\beta_\theta < 0$ is analogous to that of the Vicsek model** [162]. Conversely, the **regime $\beta_\theta > 0$ corresponds to an anti-ferromagnetic analog of the Vicsek model**.

Next, it has been assumed that the model has a steady state, where the Helmholtz free energy per cell is given by Eq. (2.7). Due to its extensivity, the Helmholtz free energy of complete, non-interacting, steady state system is

$$F_T \approx -\frac{1}{\beta_\theta} \sum_{i=1}^{N} \ln Z_i = -\frac{1}{\beta_\theta} \ln\left(\prod_{i=1}^{N} Z_i\right),$$

where Z_i is the normalization constant of Eq. (2.3) for the i-th cell. For a weakly interacting system, the mean-field effective normalization constant $Z_T := \prod_{i=1}^{N} Z_i$ is given by,

$$Z_T = \int e^{-\beta_\theta \sum_{i=1}^{N}\left[\ln(2\pi) + \ln\left(1 - e^{-2\gamma_i}\right)\right]} d\vartheta_i. \tag{4.15}$$

Note that this is only valid in the limit $\beta_\theta \to 0$. Integrating and substituting the resulting Z_T into Eq. (2.7) (see Supporting information), yields the Helmoltz-like free energy

$$F = N\left[\left(1 - \frac{1}{\beta_\theta}\right)\ln(\gamma_i) + \ln(4\pi) + \frac{\ln(1 - \beta_\theta)}{\beta_\theta}\right]. \tag{4.16}$$

Eq. (4.16) is well-defined only for $\beta_\theta < 1$. This indicates that no steady state exists for $\beta_\theta \geq 1$, hinting at an out-of-equilibrium regime [133]. The present model belongs to the class of models with logarithmic potentials (see Eqs. (2.6) and (4.5)). The existence of a non-normalizable state in certain parameter regimes is a staple of systems with logarithmic potentials [87].

(i) $\beta_\theta > 0$ **(ii)** $\beta_\theta < 0$
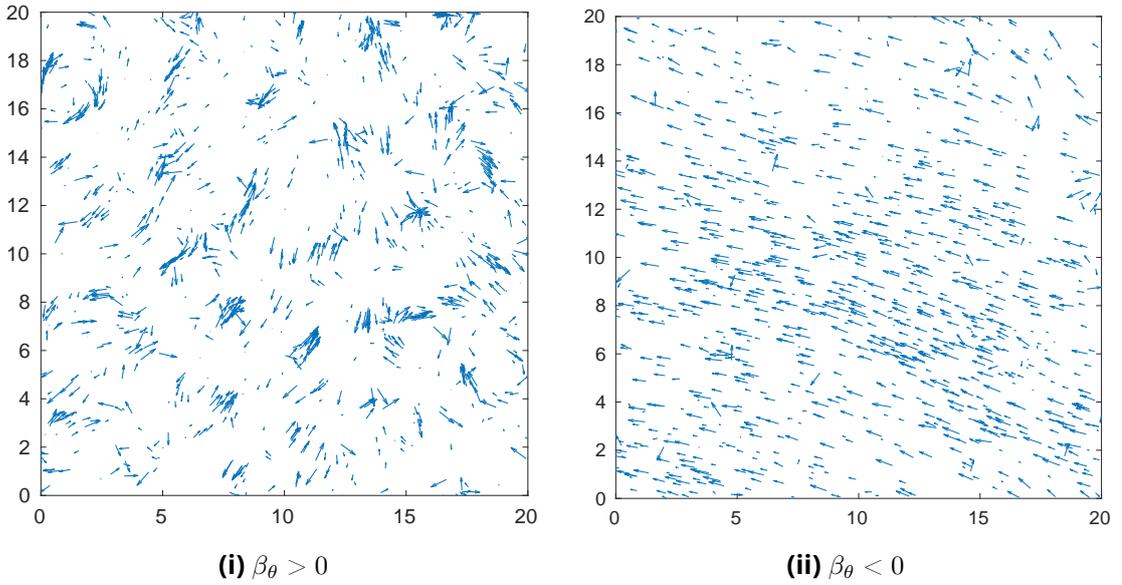
Figure 4.2.. Simulation snapshots of the velocity field at long times. Arrows show the direction and magnitude of the velocity field. The snapshots re taken after 1000 time steps. 1000 particles re simulated, with an interaction radius of 3, and noise standard deviation of angles and speeds equal to 0.01. Here $g = 1, \beta_v = -5$ and $\epsilon = 0$. In (i) the value of the angular sensitivity was equal to, 18 while in (ii) the angular sensitivity was equal to -0.25. Periodic boundaries re employed.

4.2.1. Patterns in different parameter regimes

The model was implemented computationally to characterize the model and the effects of the different parameters on the resulting macroscopic behavior. The general qualitative behavior of the model can be observed in Fig. 4.2. In the regime $\beta_\theta < 0$, cells tend to travel in a single direction after some time has elapsed, similar to the Vicsek model. Conversely, in the $\beta_\theta > 0$ regime, cells are seen to move collectively in transient vortex-like structures, even after long times have elapsed. Qualitatively, the patterns resulting from different parameter combinations are summarized in Table 4.1. Analyzing simulations, two important phenomena are observed. First, there is a critical parametric regime $\Omega_C := \{(\beta_\theta, R) : S_A^1, S_A^2 > 0\}$ where patterns emerge. Specifically, for low values of interaction radius, R no structures can be formed. This indicates that medium-to-long range spread of information is necessary for ordering. On the other hand, for $\beta_\theta > 0$ for large values of R, outside of Ω_C, again no patterns occur. This implies that for large interaction radii, there is a destructive interference of the travelling information. A second important observation is that patterns do not depend on the choice of β_v when this is different than zero. For $\beta_v \neq 0$, LEUP dynamics divide the population into fast and slow cells. While fast cells are useful for spreading information (and therefore, increasing the effective interaction range), slow cells are necessary for maintaining local ordering. On the other hand, if one fix the initial speed distribution and assume $\beta_v = 0$, then different patterns can be found shown in Table 1 (also see in Supplementary Material 7). Furthermore, global ordering at long times has been characterized interestingly. The global polar order parameter, given by Eq. (4.7), for the complete simulation domain, measures the global degree of polar alignment, or polarization.

Table 4.1.. Qualitative description of the observed patterns for different angular sensitivity and interaction radius regimes, as well as radial sensitivity. The patterning regime Ω_C is the blue area in Fig. 4.3i,4.3ii.

| $\beta_\theta < 0$	Radial Sensitivity (β_v)	$R \notin \Omega_C$	$R \in \Omega_C$
	$\beta_v \neq 0$	Polar aligned streets of cells	Scattered polar aligned cells
	$\beta_v = 0$ (for uniform distribution)	Compact polar aligned cluster	Compact polar aligned cluster
$\beta_\theta > 0$	Radial Sensitivity (β_v)	$R \notin \Omega_C$	$R \in \Omega_C$
	$\beta_v \neq 0$	No order or patterns	Vortices
	$\beta_v = 0$ (for uniform distribution)	No order or patterns	Nematic streaming and vorticules
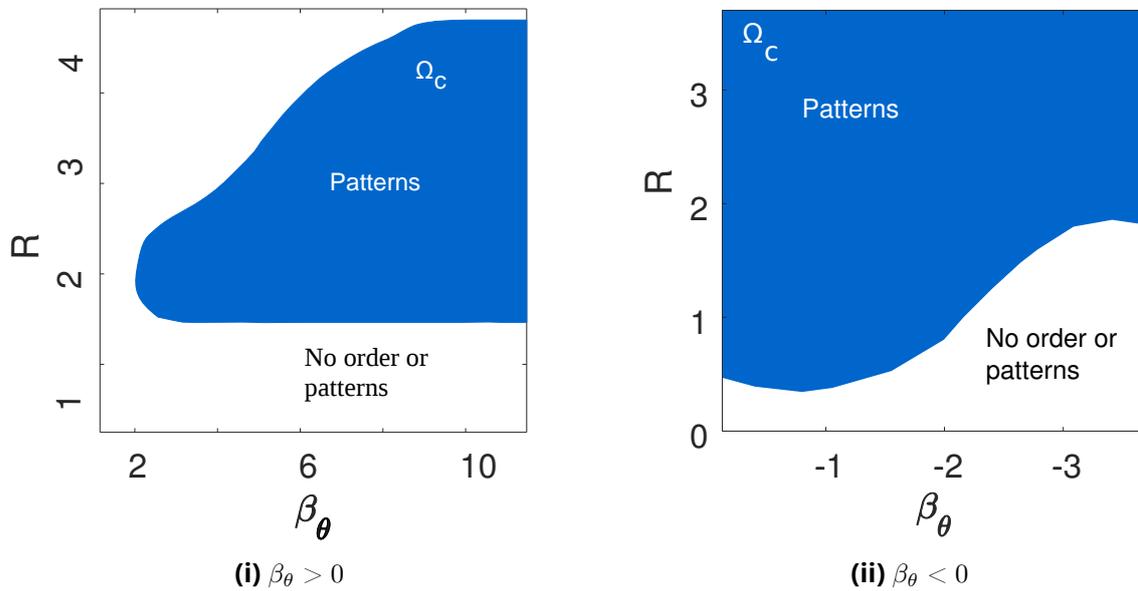
Figure 4.3.. Pattern formation regimes Ω_C in the interaction radius - sensitivity plane for positive and negative values of angular sensitivity. In (**a**) there exists an optimal regime where one can find pattern formation. But in (**b**) for negative values of beta can see patterns at smaller values of interaction radius and at smaller values of angular sensitivity.

The global nematic order parameter, given by Eq. (4.8) for the complete simulation domain, measures the tendency of all cells to align nematically, or along a single axis. These order parameters take a value of one when there is global order, while taking a value of zero when the system is completely disordered. It should be noted that polar order implies nematic order, but the reverse is not true.

Similarly to other velocity alignment models [124], the model shows an order-disorder transition with increasing noise amplitude and decreasing density (in Supplementary Material 7). More importantly, in the regime $\beta_\theta < 0$, the system also undergoes a transition towards polar order with decreasing β_θ. After the transition, most particles have a similar orientation (Figs. 4.4i and 4.4iii). In the regime $\beta_\theta > 0$, a phase transition is also observed towards nematic order with increasing β_θ. In this case, However, the nematic ordering is not perfect, as evidenced by the nematic order parameter reaching values of around 0.35 after transition (Fig. 4.4ii) compared to the value of 0.9 of the polar order parameter after transition in the $\beta_\theta < 0$ regime. This is further evidenced by the bimodal distribution of orientations with peak separation of approximately π radians (Fig. 4.4iv). These simulation results further corroborate the previous theoretical results.

In turn, the effect of speed sensitivity β_v in terms of phase transitions has studied. Fix the angular sensitivity β_θ, either positive or negative, thus the speed distribution will only depend on β_v values.

When β_v is positive then speed distribution become bimodal and for $\beta_v < 0$ the speed distribution becomes unimodal (see Fig.10 in Supplementary Material 7). Moreover, one can observe that if the radial sensitivity $\beta_v < 0$ decreases, then the average speed increases. On the other hand, for any value of $\beta_v > 0$, the first and the second moments of the speed distribution cannot be defined, since this is bimodal. Finally, for increasing cell densities, the average speed increases as well (see Supplementary Material 7).

4.3. Application : the spherical bacteria case

Collective motion of bacteria has been extensively studied and modeled. Most studies have focused on the collective properties of *S. enterica*, *E. coli*, and *M. xanthus*. These species of bacteria are similar since they have a high aspect ratio. It has been shown that volume exclusion, coupled with a high aspect ratio, is sufficient to induce velocity alignment in the system [124], and accordingly, ordered clusters of bacteria are observed at high densities.

However, it has been recently shown [130] that even spherical *S. marcescens* bacteria do display collective migration for experimental details please see SI section). The biophysical mechanism whereby spherical bacteria interact with one another must be different from the high body aspect ratio volume exclusion mechanism proposed for elongated bacterial species.
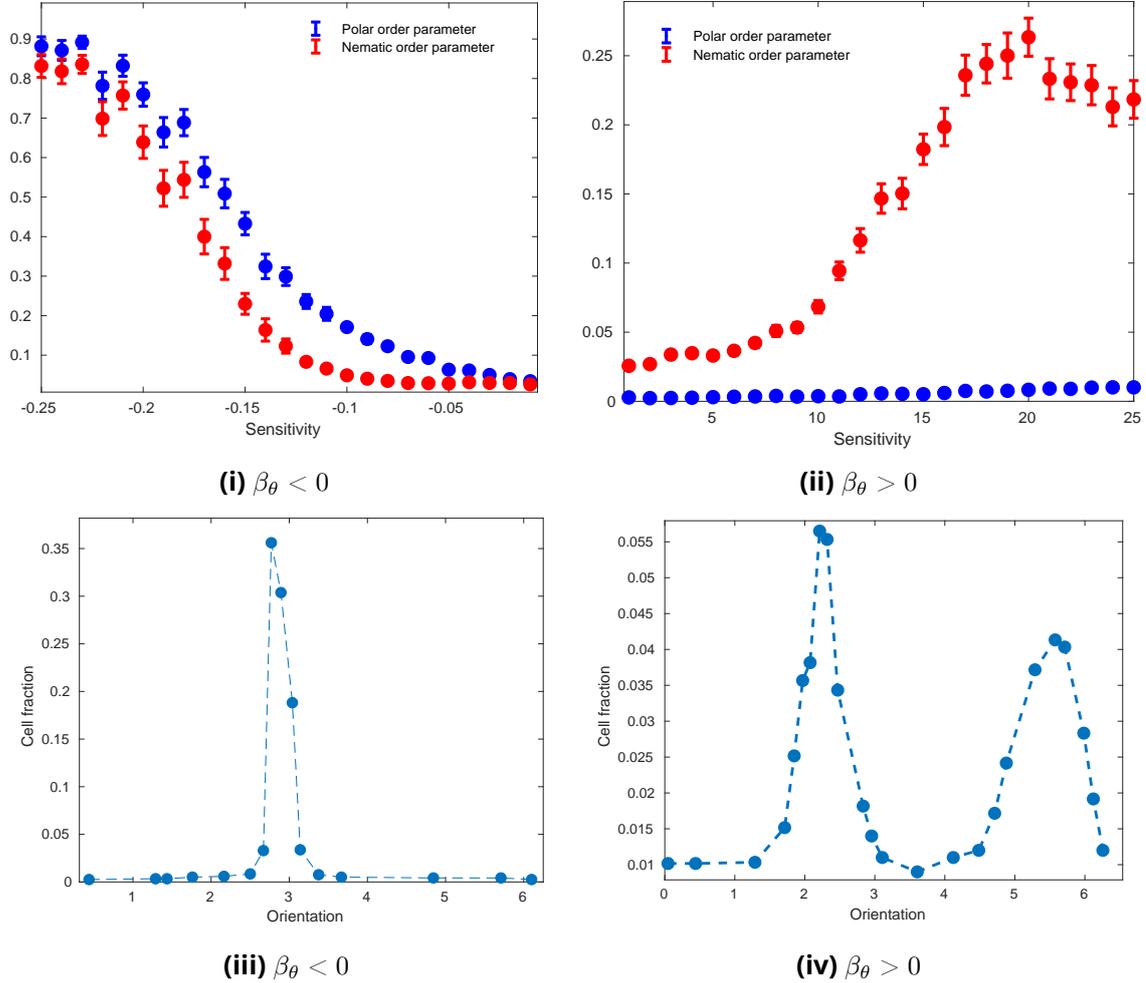
Figure 4.4.. Order-disorder phase transitions and orientation distributions in two parameter regimes. Here $g = 1$, $\beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. (i) In the regime $\beta_\theta < 0$, a phase transition towards polar order occurs at a critical value of the sensitivity. (iii) After the phase transition, polar order arises, and all cells have roughly the same orientation. (ii) In the regime $\beta_\theta > 0$, the phase transition towards nematic order occurs at critical value of the sensitivity. (iv) There is partial nematic order after the phase transition. Accordingly, several cells have opposite orientations. (i) and (ii) The number of particles was fixed at 10^3, noise standard deviation at 0.01, and interaction radius at 3. Values of the order parameters re averaged over 50 realizations after 1000 time steps. (iii) and (iv) The number of particles was fixed at 1000, noise standard deviation at 0, and interaction radius at 3. The histogram was created with data from 50 realizations after 1000 time steps.

Recently, a combination of biophysical agent-based and hydrodynamics model has been proposed to describe these experiments. In this study, the experimental observations re only partially reproduced. Therefore, the biophysical mechanisms underlying collective migration in spherical bacteria are still not well understood. An important aspect to consider is the bacterial speed v_i. It was found experimentally [130] that bacterial speed followed a Rayleigh distribution, dependent on bacterial density. Collective effects on cell orientations, on the other hand, re studied by observing the vortical behavior of the population [130].

To reproduce the experimentally Rayleigh distribution for cell speed, one choose the function $g(v_i) = v_i^{-1}$ as shown in [135]. It is important to note that this term is not impacting the qualitative behavior of the average bacteria speed, but only its variance (see Supplementary Material 7). Moreover, the interested reader could see the impact of the friction term in the average cell velocity in Supplementary Material 7. As shown in Fig. 4.5, the model qualitatively and quantitatively reproduces both the speed distribution and vorticity behavior of the experimental system. Interestingly, the behavior of the experimental system was replicated for high values of the sensitivities β_v and β_θ, and large interaction radii R.

The LEUP model not only allows for a quantitative reproduction of the experiments, but also provides insight into the potential biophysical mechanisms. Such values of the sensitivities and interaction radii indicate far-reaching, strong tendencies of bacteria to average their speeds while reorienting and traveling differently from their neighbors. Spherical, rear-propelled particles have been shown to destroy polar order as a result of hydrodynamic interactions [45], similarly to the model. Considering that *S. marcescens* is an example of a spherical, rear-propelled particle [75], the results agree with previous findings indicating that *S. marcescens* interacts through long-range hydrodynamics [153]. The long range interaction radius suggests the existence of hydrodynamically induced interaction (which has been suggested by Ariel et al. as well as by other studies ([75], [citesteager2008dynamics)) or self avoiding interaction [134].

4.4. Summary and outlook

In this work, an off-lattice model of LEUP-induced collective migration has introduced, based on the self-propelled particles modeling framework. It was assumed that individuals changed their radial and angular velocity components independently through LEUP. Reorientation is governed by a stochastic differential equation depending on a white noise term and a force arising from an interaction potential.

The exact form of the interaction potential can be very complex, and its specific form is dependent on particular mechanochemical details of the modeled system. While it has been shown that, in general, interactions between individuals can effectively drive the entropy of the entire system towards an extremum point [37, 62], here do the opposite. Instead of modeling the interaction potential biophysically,

(i)

(ii)

(iii)
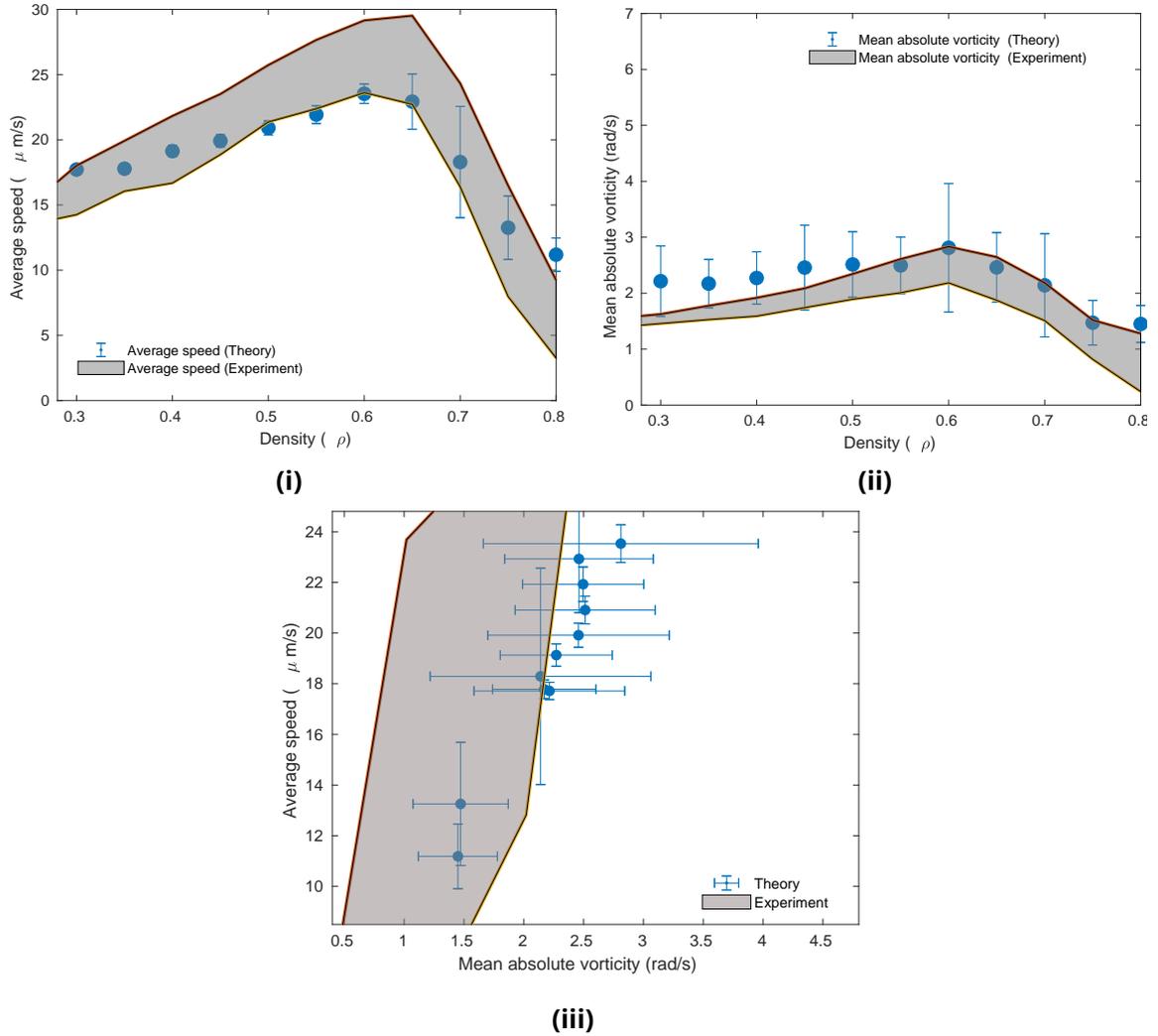
Figure 4.5.. Comparison between vorticity trends in experiments and in simulations. (i) Relation between the average speed and the density. The simulation values shown are averaged over fifty realizations. (ii) Dependence of the spatially normalized averaged absolute value of vorticity on the density. The simulation values shown are averaged over fifty realizations (iii) Relation between average speed versus mean absolute vorticity from simulations for various densities over fifty realizations. Experimental values re taken from [130]. Throughout all simulations, the standard deviation of the noise was set at 0.0001, interaction radius at $R = 10$, proportionality constant $\epsilon = 0.008$, radial sensitivity $\beta_v = -20$, $g = \frac{1}{v_i}$ and angular sensitivity at $\beta_\theta = 5$. Data was obtained after 500 time steps.

it was assumed that particles followed the LEUP, which dictates that cells change their internal states in order to minimize the uncertainty of the internal states of cells in their surroundings. Although LEUP has been conceptualized to deal with high-dimensional internal states involved in cell decision-making, here on physical internal states one restricts such as speed and orientation. While cell speed was assumed to always minimize uncertainty, there was no assumption made on the cell orientation. Particles are therefore free to reorient either towards or against the gradient of entropy of the orientation distribution of particles in their neighborhood, depending on the sign of the sensitivity parameter, which also dictates the strength of the interaction. The orientation distribution in the neighborhood was assumed to be wrapped Cauchy distributed. Such a distribution facilitates the mathematical analysis of the model. However, the usage of other wrapped distributions does not qualitatively change the general behavior of the model (see Supplementary Material 7). Please note that non-parametric methods for estimating entropies without assuming any underlying parametric distributions exist. For instance, such methods employ kernel density estimation, $k-$nearest neighbthes or regression methods [70].

It has been shown that when the parameter β_θ is negative, the model produces steady-state polar alignment patterns. Interestingly, that the classical formulation of Vicsek model [162] is a special case of LEUP. Conversely, when the parameter is β_θ is positive, particles tend to reorient against the mean velocity of their neighborhood. In this regime, the free energy diverges, indicating an out-of-equilibrium parameter regime. This kind of parameter-dependent dichotomy is similarly observed in systems with logarithmic potentials [40], involved in processes such as long range-interacting gases [23], optical lattices [99], and DNA denaturation [7]. The dichotomy arises from the logarithmic form of the entropy driving interaction in the model. It has been shown that, due to the non-normalizability of the steady state solution, such systems require a time-dependent expression for their analysis [87]. Therefore, an in-depth theoretical analysis of the model would require a similar multiparticle, time-dependent expression of the angular probability densities.

However, the LEUP migration model may go beyond the observed patterns in past Viscek-type models. In particular, in Barua et al. [12] have developed a discrete speed version of the LEUP migration model, where cells can have only zero or a finite speed. This model exhibits Turing patterns, i.e., dynamics clusters of non-motile cells of specific characteristic wavelength, where previously published Viscek-like models cannot may produce moving clusters of swirling cells (e.g., the milling Viscek model [34]) but never static ones.

As a proof of principle, one show that the model replicates the collective vortical behavior of spherical motile particles. Recently, the collective behavior of spherical particles been modeled as a combination of steric repulsion and hydrodynamic interactions [98]. Hydrodynamics and steric interactions has shown induce long-range microenvironmental entropy maximization, which coincides with the $\beta_\theta > 0$ LEUP regime. This generalizes the type of biophysical mechanisms required to produce vortical patterns.

4. Cellular orientation decisions: origin of pattern formations in collective cell migrations

It should be noted that, while spherical *S. marcescens* bacteria have been modeled biophysically, their collective behavior was partially reproduced [4]. This hints at an additional biological and/or biochemical interaction between cells. While the LEUP-based model is coarse-grained in terms of specific biophysical /biochemical interactions, it allows for a plausible reproduction of the experimentally observed collective velocity behavior by fitting a only few parameters. The application to spherical bacteria allows us to showcase the potential of the LEUP principle when the precise interaction mechanisms are not known.

As already mentioned, some assumptions can be made to simplify the model. Here, the model assumes a Gaussian, white noise term in the SDEs. This results in normal diffusive behavior in the absence of interactions. It has been observed experimentally, However, that in some conditions, cells perform Lévy walks resulting in superdiffusive behavior [106]. By changing the distribution or time correlations of the noise [30, 117], it would be possible to both replicate the non-Gaussian dynamics of single cells, and investigate the effect of single anomalous dynamics on collective behavior.

Another assumption has been taken i.e., particle velocities are the only internal states relevant for reorientation, for simplicity and as a proof of concept of the LEUP principle. However, it is reasonable to think that other states, such as relative position or adhesive state, may be relevant to include when modeling specific systems. This reveals an interesting point in the application of LEUP-driven models, which is the selection of the most relevant/dominant internal variables. Although experimental intuition could be the easiest approach, are currently developing a spatial principal component analysis method that would allow to select the most relevant internal variables using spatial data such as multiplexing biopsies or spatial RNA sequencing.

As stated above, LEUP circumvents the biophysical details of cell migration. The need to model systems of interacting agents without previous knowledge of the biophysical mechanisms involved has sparked at least another agent based model[43]. In this model, similarly to LEUP model, agents act without a mechanistic rule. Rather, they consider every possible action and penalize those which are not favorable to their internal standards. While both the aforementioned model and LEUP are defined in a similar spirit, modeling under LEUP consists in correctly identifying the relevant internal cellular states for entropy optimization, while in [43] modeling is concerned with defining suitable penalization for each possible decision scenario.

LEUP has additional appealing features. For instance, LEUP allows for replicating a plethora of collective migration patterns. In this particular case, have analytically derived the polar and nematic alignment Vicsek models for LEUP arguments. In this sense, LEUP acts as a generative model for collective migration mechanisms. This is particularly useful upon limited knowledge of such mechanisms, a problem called structural model uncertainty. Another advantage of LEUP is the mapping of biophysical mechanism combination to the $\beta > 0$ or $\beta < 0$ regimes. This allows for unifying the model analysis but for a better classification of migration mechanisms. Finally, known mechanisms or data could be easily integrated to the proposed framework by further constraining the LEUP dynamics.

5. Cell differentiation and sensing: tissue robustness from optimal environmental sensing

In this chapter [1] the LEUP based theory on cell differentiation phenomena at the level of tissue has been represented. Cell differentiation phenomena as a collective cell decision-making and tissue robustness will be discussed briefly in this chapter.

Decision-making is a process to identify important choices and responses which depends on some basic criteria [146]. Cell decision-making is a process where cells select a new state, such as cell fates or phenotypes, in response to their microenvironmental milieu. In this regard, pluripotent cell differentiation can be viewed as cell decision-making of inheritable fates. Typically, cells irreversibly acquire new fates by following a hierarchical lineage, where pluripotent stem cells in a proper microenvironment (stem cell niche) differentiate into, for example, bone, muscle, epithelial, and further specialized cells. The cell differentiation process encompasses the dramatic change of geometry, shape, gene expression inside the cell, etc. [147, 2]. It is yet to be fully understood how the information available to pluripotent progenitors, including its intrinsically determined state and extrinsic microenvironmental signals, is encoded and processed by progenitors to generate different differentiated cell types.

In 2006, Takahashi and Yamanaka [155] discovered that almost any differentiated cell can be sent back in time to a state of pluripotency by expressing appropriate transcription factors (the Nobel Prize in Medicine 2012). Cell reprogramming can be externally induced via the delivery of transcription factors, naturally or in-vitro [49, 154]. Such a reversion of a differentiated to a pluripotent state is the idea

---

[1]This chapter includes text and figures from the publication and corresponding supplementary information: Arnab Barua, Alireza Beygi and Haralampos Hatzikirou, *Entropy* 2021, 23(7), 86. Author contribution: Haralampos Hatzikirou conceptualized the project. Arnab Barua, Alireza Beygi and Haralampos Hatzikirou formulated the mathematical model. Arnab Barua and Alireza Beygi performed the analysis. All authors interpreted the results and wrote the manuscript. All authors read and approved the final manuscript.

behind cancer stem cells (CSC). The CSC theory proposes that, among all cancerous cells, a few act as stem cells that reproduce themselves and sustain cancer, much like normal stem cells that renew and sustain organs and tissues of the body. The process of somatic reprogramming using Yamanaka factors, where many of them are oncogenes, offers a glimpse into how cancer stem cells may originate. In particular, neurological cancers such as primary glioblastomas [126] and retinoblastomas [92] are resulting from dedifferentiation of the glial and photoreceptor (retinal) cells, respectively. In particular, retinoblastoma tumor cells lose their photoreceptorness and become malignant plastic cells, i.e., CSC [92]. Xu et al. also have shown that the cell of origin for retinoblastoma is a committed cone precursor an almost terminally differentiated photoreceptor that has lost an Rb gene, and not a pluripotent progenitor [166, 25]. Here, the example of photoreceptor mosaic of avian retina has been used to shed light on the following questions: **(Q1)** how do cell intrinsic dynamics and microenvironmental factors coordinate during development to produce organized tissues such as photoreceptor mosaics [26]; and **(Q2)** why is the differentiated mosaic so stable or how probable is the reversal of retinal tissue back to a pluripotent one, e.g., retinoblastoma?

The theory behind cell differentiation has been formalized in a metaphorical way by C. H. Waddington [60, 3, 29], which allows for developing a dynamical system's framework for modeling single-cell fate decisions [48, 121, 82]. Waddington has depicted the developmental process as a series of cell decisions that can be represented as bifurcations towards a differentiated state/phenotype. Practically, cell states, also called microstates, can be viewed as a vector of molecular expressions that are experimentally measured via high-throughput omics data, FACS, immunohistochemistry markers, etc. [54, 100, 29]. Please note that such cellular microstates are technically different from the classical statistical mechanics definition of microstates. In general, differentiated states can be viewed as the fixed point(s) of microstate attractors [74, 110]. Typically, these states can be associated with an appropriate probability distribution peaked around the fixed point, which is the deepest point of the valley in the Waddington potential. The situation is much more intricate in the case of pluripotent, stem cell-like states, where Waddington has also depicted them as attractors around a fixed point. However, this has been recently challenged by Furusawa and Kaneko [52], where they have shown that stem cell-like attractors can be viewed as limit-cycle attractors and not as stable fixed points. Biological observations of dynamic variability of single cells within pluripotent cell populations distinguish between pluripotency as a molecular state and pluripotency as a function, indicating that a pluripotent state is not unique but rather appears to be compatible with a wide variety of interchangeable molecular microstates (patterns of gene or protein expressions) [28, 18]. In this view, pluripotent cells independently explore a variety of molecular expression states, and this phenotypic/state exploration transiently primes each individual cell to respond to a range of various differentiation-inducing stimuli, depending upon its instantaneous molecular state [126]. Finally, Waddington theory does not take into account cell sensing and the corresponding interactions that take place in a tissue. The aforementioned limitations of the Waddington approach require further theoretical development to answer the questions **(Q1)** and **(Q2)**.

In order to answer **(Q1)**, in this paper, one employ the recently proposed the

Least microEnvironmental Uncertainty Principle (LEUP)—which is essentially a statistical mechanical theory for cell decision-making [66, 11, 12]—and apply it to the problem of cell differentiation. The LEUP is inspired by the theories of Bayesian brain hypothesis [90], the free-energy principle [50], and other dynamic Bayesian inference theories that try to explain human brain cognitive dynamics. Similar ideas have also been proposed in the influential work by Bialek [16]. Similar to these theories, the LEUP is based on the premise that cell internal molecular networks adapt to the sensed microenvironmental data and subsequently determine the relevant decisions. In turn, cells are encoding sensed information in genetic, epigenetic, translational, or transcriptional levels, where different timescales are related to each of these encoding levels, depending on the persistence of the microenvironmental stimuli. The reason to propose a theory such as the LEUP is the fact that the complexity of molecular networks does not allow us to know the exact involved dynamics, thus one use the LEUP as a kind of dynamic Bayesian inference to circumvent this complexity and to make predictions. The LEUP theory also implies a decrease in the local microenvironmental entropy of the cell decision-maker, which biologically translates into the actions of differentiating cells, which lead to more organized tissues during development.

The central cellular process related to the LEUP is cell sensing. Cells can acquire knowledge about their microenvironment by various sensing mechanisms such as the binding of their receptors to diffusible ligands [95], pseudopodia extension [88], mechanosensing [104], proton-pump channels [96], gap junctions, etc. Cells can sense rapid changes of their milieus, where they mainly exploit two ways to decrease sensing errors: **(1)** by increasing the number of receptors or the responses of the downstream signaling pathways [14], and/or **(2)** by increasing sensing area [97]. The latter can be spanned from the resting cell size to extensions via pseudopodia, blabbing, and other cell size regulation mechanisms. Pseudopodia or blebs can act as sensors, via a pressure sensing mechanism mediated by Piezo channels, allowing cells to decide when and where to migrate [152]. In this regard, the second question**(Q2)** is further specified as: can one calculate the limits of the cell sensing radius that ensure the robustness of differentiated tissue spatial order?

To answer **(Q2)**, thedevelop a thermodynamic-like theory using the tools of stochastic thermodynamics for a generic cell differentiation process. The main biological assumption is that differentiated cells can reverse to an undifferentiated state, for instance, by the process of carcinogenesis, see also Refs. [166, 92, 33]; a recent review on the latest advances in research on the process of dedifferentiation both at cell and tissue levels can be found in [168]. Stochastic thermodynamics allows us to understand the conditions that, even though single-cell dedifferentiation is possible (microscopic reversibility), the system (tissue) is still able to be robust, and it maintains its spatial order and differentiation integrity (macroscopic irreversibility). Stochastic thermodynamics is a suitable tool for systems where small scale dynamics matters (e.g., soft matters, active matters, and biological systems); in such cases, the higher-order moments dominate [141, 140, 160, 89, 142]; for different experimental applications of stochastic thermodynamics, see Refs. [31, 32]. In addition, formulating the laws of thermodynamics in the mesoscopic scale (specifically, at the level of trajectory) has also been investigated recently [161]. To describe the
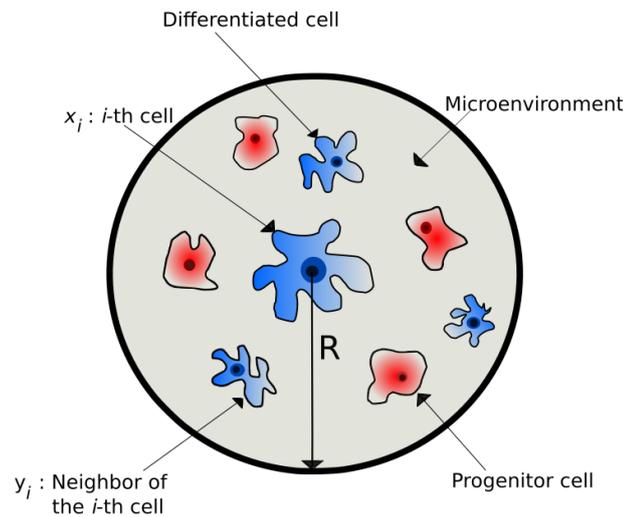
Figure 5.1.. Schematic diagram of the microenvironment of a differentiated cell, where R is the sensing radius. Microenvironment is being composed from a distribution of pluripotent progenitors and differentiated cells.

cell-level dedifferentiation process in the context of stochastic thermodynamics, the fluctuation theorem, specifically the Crooks' relation [35, 44] is being applied. The fluctuation theorem, which can be considered as the heart of stochastic thermodynamics, initially had been derived to explain how irreversibility at the macroscopic level emerges from the underlying reversible dynamics and to estimate the probability of the violation of the second law of thermodynamics within a short amount of time for systems at small scales [46, 47, 144].

By combining elements of stochastic thermodynamics with the LEUP, the Aspire to approach questions **(Q1)** and **(Q2)** within the context of the avian photoreceptor mosaic. This paper is organized as follows: in Section 5.1, the basic concepts of the LEUP is reviewed and its connection to statistical mechanics. In Section 5.2, how optimal microenvironmental sensing is associated with differentiated tissue spatial configuration is demonstrated; therein, examine the theory using the data obtained from the (avian) photoreceptor mosaic. The theory of fluctuation theorem is applied to cell differentiation in Section 5.3 and demonstrate the thermodynamic robustness of this process in the case of the avian retina development. In Section 5.4, how cell sensing radius and total entropy production are related and determine the limits of the sensing radius in order for the tissue development to be robust is shown. Finally, the conclusion and discussion of the results are presented in Section 5.5.

5.1. LEUP based mathematical model for cell differentiation

The cell collects information from its microenvironment and based on that it takes actions (phenotypic decisions). In other words, the cell reacts to the environmen-

tal information by changing its own state. By adopting the notations of Ref. [66], internal variables (such as gene expression, RNA molecules, metabolites, etc.) are denoted of the i-th cell as x_i and external variables (such as chemical signals, ligands, homotypic or heterotypic interactions, cellular densities, translational proteins, etc.) of the i-th cell as y_i. The latter includes all the different extrinsic variables within the interaction radius of the i-th cell, i.e., $y_i = \{y(r) : r \in (r_i, r_i + R]\}$, where r_i is the position vector of the i-th cell, and R is the maximum microenvironmental sensing radius of the i-th cell. Cell sensing radius is a tunable variable controlled by cell and is regulated by various biophysical mechanisms, as stated in the Introduction. This is precisely what R is modeling, where it has been considered as an intrinsic cell variable dictated by the LEUP dynamics. Please note that assumption of sensing radius can be identified with the typical interaction radius considered in agent-based models. In Fig. 5.1, the schematic diagram of the microenvironment has been shown of a differentiated cell as is being composed of both pluripotent progenitors and differentiated cells.

The behavior of a cell is considered as a Bayesian decision-maker that reacts to its microenvironment [1, 66], such that

$$P\left(x_i\left(t + \tau_{\text{decis.}}\right) \mid y_i\left(t\right)\right) = \frac{P\left(y_i\left(t\right) \mid x_i\left(t\right)\right) P\left(x_i\left(t\right)\right)}{P\left(y_i\left(t\right)\right)}, \tag{5.1}$$

where $\tau_{\text{decis.}}$ is the time needed that the cell makes a decision. The distribution $P\left(y_i\left(t\right) \mid x_i\left(t\right)\right)$ is the probability of microenvironmental information/data being collected by cell at time t. In other words, it is the probability that the cell perceives all other cells, chemicals, and nutrients in its surrounding. The distribution $P\left(x_i\left(t\right)\right)$ is the prior probability of the cell current internal states. If these two distributions are multiplied and be divided by the probability of external states $P\left(y_i\left(t\right)\right)$, then (5.1) implies that the resulting quantity is the posterior probability distribution of internal states $P\left(x_i\left(t + \tau_{\text{decis.}}\right) \mid y_i\left(t\right)\right)$. The latter describes the most likely decision to be made by cell over the internal variables after processing the available information in the time period of $\tau_{\text{decis.}}$.

As stated in the Introduction, the cell prior is assumed to continually being updated by the previous time step posterior, in the sense of Bayesian learning. We also have assumed a *perfect* transfer of the prior to the next time step posterior, although this process is highly noisy. Now, by taking the logarithm of (5.1) and integrating over the joint probability distribution $P\left(x_i\left(t\right), y_i\left(t\right)\right)$, for small decision times, we obtain:

$$\frac{\partial S\left(x_i \mid y_i\right)}{\partial t} = \frac{1}{\tau_{\text{decis.}}}\left[S\left(y_i \mid x_i\right) - S\left(y_i\right)\right] = -\frac{1}{\tau_{\text{decis.}}}I\left(x_i, y_i\right), \tag{5.2}$$

where for simplicity we have dropped t as an argument and $I\left(x_i, y_i\right)$ is the mutual information. The above equation reaches equilibrium when the mutual information vanishes. Now, the first crucial assumption is made of the work **(A1)** that the cell decision time is much smaller than the asymmetric division time, i.e., $\tau_{\text{decis.}} \ll \tau_{\text{div.}}$. This can be justified as cell division which is a prerequisite for differentiation takes around $\sim 24h$; on the other hand, the relaxation timescale of the molecular networks responsible for deciding a new state is much shorter ($\sim 1h$). This implies

that the dynamics of internal variables x_i's are much faster than the microenvironmental dynamics of y_i's. According to this *timescale separation*, can be assumed that variations of $S(y_i \mid x_i)$ and $S(y_i)$ belong to the slow manifold of the system. In turn, one observe that in order Eq. (5.2) reaches an equilibrium, the microenvironmental entropy sensed by cells $S(y_i \mid x_i)$ should be inevitably a decreasing quantity with respect to time. This fact can be interpreted as the entropy of cell sensing distribution should become more *focused* and more *independent* of the microenvironment as time goes on, since this information – microenvironmental data – has already been encoded in the cell prior. The latter is valid within biophysical context because of the following reasons: first, the collection of microenvironmental data, that is, the precise evaluation of $P(y_i \mid x_i)$ is energetically expensive, thus it is not favorable as cells would like to be energetically efficient; and second, fully differentiated cells are perfectly adapted to their microenvironment and the corresponding fluctuations implying an optimal prior distribution $P(x_i)$.

Now, let's generalize (2.3) a biologically relevant scenario of cell differentiation. It is well known that differentiation takes place during asymmetric divisions of pluripotent progenitor cells [113]. Let assume that $\mu \propto \tau_{\text{div.}}$ is the asymmetric division probability of a pluripotent cell. In a *mesoscopic* microenvironmental ensemble of $N(y_i \mid x_i)$ cells around the i-th cell, the probability of the central i-th cell is calculated to change its phenotype. This probability reads as

$$
\begin{aligned}
P(x_i) = & \text{Prob.}\{\text{to select the } i\text{-th cell out of } N \text{ cells}\} \times \\
& \text{Prob.}\{\text{the } i\text{-th cell divides asymmetrically once}\} \qquad (5.3) \\
& \times \text{Prob.}\{\text{the } i\text{-th cell decides over its phenotype}\}.
\end{aligned}
$$

The latter probability happens according to the LEUP, and it is the same as (2.3). Now, as the probability of asymmetric divisions is assumed within an ensemble of N cells follows a *Poisson distribution* $\text{Pois}[\mu N(y_i \mid x_i)]$, which is a limit of the Binomial distribution $\text{B}[N(y_i \mid x_i), \mu]$ for a small proliferation probability μ. Putting all these together, one can write:

$$
P(x_i) \propto \frac{1}{N(y_i \mid x_i)} \times N(y_i \mid x_i) e^{-\mu N(y_i \mid x_i)} \times e^{-\beta S(y_i \mid x_i)} = \frac{e^{-\beta S(y_i \mid x_i) - \mu N(y_i \mid x_i)}}{Z_p(\beta, \mu)}, \qquad (5.4)
$$

where the normalization factor of $P(x_i)$ is defined as $Z_p(\beta, \mu) = \int e^{-\beta S(y_i \mid x_i) - \mu N(y_i \mid x_i)} \mathrm{d}x_i$.

To illustrate the above with a concrete example: if the microenvironmental probability distribution sensed by the i-th cell follows a Gaussian distribution with the variance σ_n^2, assuming that y_i is scalar, and hence
$S(y_i \mid x_i) = (1/2) \ln[2\pi e \sigma_n^2(y_i \mid x_i)]$ and by considering $\mu = 0$, then (5.4) reduces to

$$
P(x_i) = \frac{\sigma_n^{-\beta}(y_i \mid x_i)}{\sum_j \sigma_j^{-\beta}(y_j \mid x_j)}. \qquad (5.5)
$$

Now, an explicit formula is established for cell internal entropy. To this end, by exploiting (5.4), we obtain

$$
S(x_i) = - \int P(x_i) \ln P(x_i) \, \mathrm{d}x_i = \beta \langle S(y_i \mid x_i) \rangle_{x_i} + \mu \langle N(y_i \mid x_i) \rangle_{x_i} + \ln Z_p, \qquad (5.6)
$$

where $\langle ... \rangle$ denotes the expectation value. By making usage of the notions which are developed within the context of thermodynamics, one can also define phenotypic internal energy sensed by cell as

$$U_{\text{int.}}(x_i, y_i) = \langle S(y_i \mid x_i) \rangle_{x_i} + \frac{\mu}{\beta} \langle N(y_i \mid x_i) \rangle_{x_i}. \tag{5.7}$$

Please note that for $\mu = 0$, the internal energy is the same as entropy for a particular realization. Thus, (5.4) resembles the Boltzmann distribution.

Interestingly, since the microenvironmental entropy $S(y_i \mid x_i)$ acts as an effective internal energy, it is expected to be a decreasing quantity in time. This result is in agreement with the discussion which follows (5.2). Thus, the LEUP is consistent with its premises.

5.1.1. Statistical results from LEUP

The steady-state distribution of (5.5) which is derived by the LEUP, under the assumption of Gaussian distributed microenvironment and $\mu = 0$, can be associated with different established statistical results. Here, several cases have been distinguished for specific values of β, which have particular meanings in statistics. For instance, when $\beta = 1$, one can reproduce distributions proportional to the well-known Jeffreys prior, see also Ref. [66]. Jeffreys prior is known as the most typical uninformative prior used in Bayesian inference [78].

Interestingly, if $\beta = 2$ is assumed, then one recovers the so-called *minimum variance estimator* when fusing multiple scalar estimates. Following [122], let us assume extrinsic variables $y_1, y_2, ..., y_i$, such as ligand concentrations, cell densities, etc., that are pairwise statistically uncorrelated and are normally distributed, i.e., $y_i \sim \mathcal{N}(\mu_i, \sigma_i^2(x_i))$, where $1 \le i \le n$. In turn, the assumptions of intrinsic variables $x_1, x_2, ..., x_i$, correspond to cellular sensors and downstream process of the aforementioned microenvironmental variables. The average sensed microenvironment is defined as the average of the extrinsic signals y_i's weighted by the distribution of the cell sensors x_i's, i.e., $Z(y_1, y_2, ..., y_i) = \sum_{j=1}^{n} P_j y_j$, where $\sum_{j=1}^{n} P_j = 1$. The distribution of internal variables P_i that minimizes the variance, aka noise of the sensed microenvironment, is given by the following formula:

$$P_i = \frac{\sigma_i^{-2}(x_i)}{\sum_j \sigma_j^{-2}(x_j)}, \tag{5.8}$$

which has the same form as the steady state of the LEUP (5.5), for $\beta = 2$. In the following, the latter part of the result is connected with a specific cell sensing scenario.

In this section, first the connection between theory of LEUP and cell sensing is made with respect to receptor-ligand binding. Then, one can apply the theory to avian photoreceptor mosaic and fit the parameter β to recover the photoreceptor percentages in retina.

5.2. Relation between LEUP and cell sensing

In this subsection, the relevance of the LEUP is shown within the context of cell sensing mechanisms. Here, the receptor-ligand sensing is focused as an apparatus of cell, where complexes are formed at cell membrane and subsequently internalized via endocytosis [95]. This sort of sensing mechanism is also relevant in the case of photoreceptors [151]. Receptors act as sensors for specific microenvironmental molecules (ligands) that bind together with a certain affinity to form complexes. Often different ligands may bind to the same receptor, for instance, the Notch-Delta-Jagged system where Notch receptor can bind to either Delta or Jagged molecules [20, 157]. The sensed information can be quantified by the concentration of internalized complex molecules. For simplicity, x is denoted as receptor concentration at the cell membrane and y_1 and y_2 as the corresponding ligands concentrations, which it is assumed that they are statistically independent. Moreover, the ligands concentrations are consumed by cells, since they bind to receptors and therefore depend on the concentration x (we omit this dependence for the sake of simplicity). Typical dynamics of such systems reads as

$$\frac{dc}{dt} = k_1 x y_1 + k_2 x y_2 - dc. \tag{5.9}$$

Under the assumption of fast decay rates $d \gg 1$, which is consistent with **(A1)**, the system behaves as in a steady state, that is,

$$c_{\text{Eq.}} = \frac{k_1}{d} x y_1 + \frac{k_2}{d} x y_2. \tag{5.10}$$

The terms, $(k_i/d)x$, $i = 1, 2$, define the percentage of receptors bound to ligands y_i's. In the context of the LEUP, $(k_i/d)x$ corresponds to $P(x \mid y_i)$, which under the assumption of Gaussian microenvironment – in a steady state – is:

$$P_i = \frac{\sigma_i^{-\beta}(x_i)}{\sigma_1^{-\beta}(x_1) + \sigma_2^{-\beta}(x_2)}, \qquad i = 1, 2, \tag{5.11}$$

which for $\beta = 2$ coincides with (5.8). Taking all the above together, the receptor-ligand cell sensing system is translated as a linear combination of complex formation estimates y_i's, where the LEUP probabilities P_i's are the corresponding coefficients/proportions of complexes bound to ligands y_i's, that is,

$$c_{\text{Eq.}} = \frac{\sigma_1^{-\beta}(x_1)}{\sigma_1^{-\beta}(x_1) + \sigma_2^{-\beta}(x_2)} y_1 + \frac{\sigma_2^{-\beta}(x_2)}{\sigma_1^{-\beta}(x_1) + \sigma_2^{-\beta}(x_2)} y_2, \tag{5.12}$$

which coincides with the definition of $Z(y_1, y_2, ..., y_i)$, for $n = 2$, defined in the previous section.

At this point, a sanity test for the results are tested. For simplicity, consider a single type of ligand concentration y that binds to a receptor of concentration x. Then, the complex formation dynamics is described as

$$\frac{dc}{dt} = k_+ x y - dc. \tag{5.13}$$

At the equilibrium, the steady-state complex concentration $c_{\text{Eq.}}$ reads

$$c_{\text{Eq.}} = \frac{k_+}{d}xy. \tag{5.14}$$

As before, the probability $P_x = (k_+/d)x$ is defined as the proportion of binding receptors. Then from (5.14), one observe that:

$$P_x \propto \frac{1}{y}. \tag{5.15}$$

The ligand concentration, $y = \langle Y \rangle$, is the expected value of the number of ligand molecules within the cell volume, which is denoted as Y. Using arguments similar to that of Berg and Purcell in their seminal paper [14], the ligand molecules are diffusing, and therefore it can be assumed that Y follows a Poisson distribution, thus $\sigma_Y^2 = y$, where the ligand variance is denoted as σ_Y^2. Now, by combining the latter with (5.15), one recover the non-normalized LEUP result for $\beta = 2$:

$$P_x \propto \sigma_y^{-2}(x). \tag{5.16}$$

This specific value of β complies with the *perfect monitoring/sensing* assumptions (for details, see [14]). Please note that in (5.16), the dependence of diffusible ligand concentration on the receptor concentration x is explicitly denoted. As a side remark, if one change the receptor-ligand binding term by introducing, for instance, finite number of receptors or covalent bonds, then the parameter β will be modified.

The above result is pivotal since it connects the cell fate decision-making with optimal microenvironmental sensing in terms of minimization of sensing noise, which results into a specific spatial phenotypic distribution. In the following, the validity of these results is explored in the case of the avian photoreceptor mosaic.

5.3. LEUP driven fluctuation theorem : confirms the thermodynamic robustness of differentiated tissues

In this section, the thermodynamic constraints are determined of two coarse-grained cell states that correspond to pluripotent (s) and differentiated (d) state. Then apply the resulting theory to the particular case of the avian cone cells differentiation. To this end, first it has been shown that how microstates (internal variables) are related to the microenvironmental information and heat transfer.

In this context, a cellular *microstate* corresponds to a cell's phenotype that lives in a tissue, which could be gene expression, RNA molecules, receptor distribution etc. In other words, microstate gives information about the internal states of the cell. label these internal variables as x_s and x_d corresponding to pluripotent and differentiated cells, respectively. A cellular *macrostate* is defined as a statistical
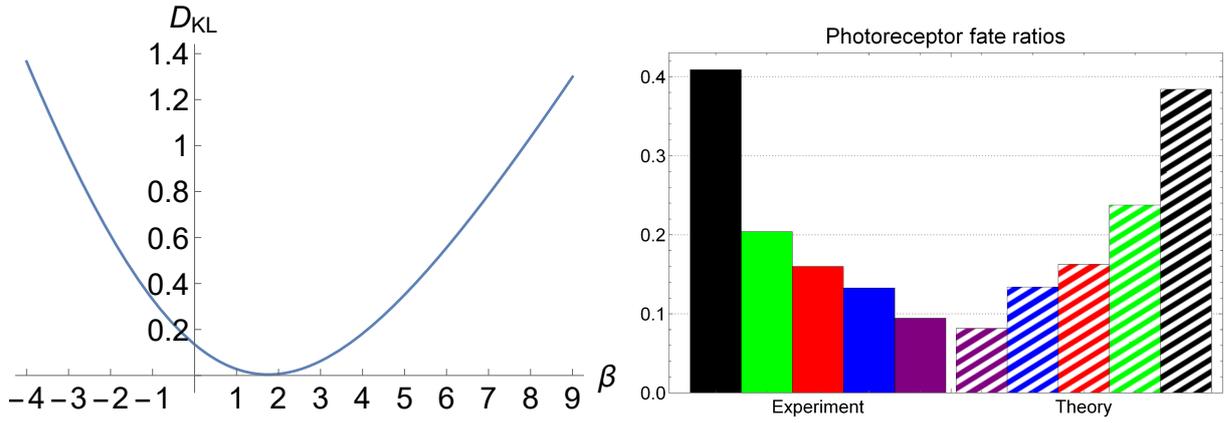
Figure 5.2.. In the left panel, the Kullback-Leibler divergence D_{KL} is depicted as a function of β. D_{KL} reaches its minimum of ≈ 0.004 at $\beta \approx 1.754$. The experimentally obtained photoreceptor fate ratios is compared to the LEUP for this particular β, in the right panel.

observable (e.g., average) of a cell's microenvironment that involves multiple cells of different phenotypes. Macrostates contain information about the external variables, which are labeled as y_s and y_d. The former macrostate y_s is assumed to describe a microenvironment of pluripotent progenitor or stem cells, i.e., characterized by the microstate x_s; the macrostate y_d accordingly describes a microenvironment of differentiated cells characterized by the microstate x_d. The number of pluripotent cells neighboring to a cell is defined with microstate x_s as, $N(y_s \mid x_s)$ and the number of differentiated cells neighboring to a cell is defined with microstate x_d as $N(y_d \mid x_d)$. The total number of pluripotent cells and differentiated cells inside the system are denoted as $N(s)$ and $N(d)$, respectively. Now based on (5.4), the probability of the cell can be written as the microstate x_s with the corresponding macrostate s as

$$P(x_s) = \frac{e^{-\beta_s S(y_s|x_s) - \mu_s N(y_s|x_s)}}{Z_1}, \tag{5.17}$$

and the probability of being in the microstate x_d with the corresponding macrostate d as

$$P(x_d) = \frac{e^{-\beta_d S(y_d|x_d) - \mu_d N(y_d|x_d)}}{Z_2}. \tag{5.18}$$

It is known that for a system which is coupled to a set of heat baths and is in a time-symmetrically driven nonequilibrium state, the *Crooks' theorem* is applicable [35, 44]. In this case, the dynamics follow Brownian motion, i.e., there are no extreme "jumps" in the system state. For any trajectory to be initially at $x_s(0)$ and is going through microstates $x_s(t)$ over time τ, the Crooks' theorem implies that [44],

$$\beta' \Delta Q = \ln \left[\frac{w[x_s(t)]}{w[x_s(\tau - t)]} \right], \tag{5.19}$$

where $\beta' \equiv 1/T$, which T is the temperature of the heat bath, ΔQ is the total heat released into the bath over the trajectory of $x_s(t)$, and $w[x_s(t)]$ is the probability of the trajectory $x_s(t)$. Eq. (5.19) demonstrates that when there is a forward state change, the system loses heat to the reservoir and in the case of the time-reversed path, there is a heat gain from the reservoir; this, in turn, implies that a forward

trajectory is more probable than the time-reversed one, thus, (5.19) substantiates a relation between heat and irreversibility at the microscopic level.As is demonstrated in Ref. [44], (5.19) is also valid for other kind of steady-state probability distributions besides the classical Boltzmann distribution. Here, such as Brownian "jumps" are interpreted as changes in the phenotype and the associated heat losses are assumed to be due to the cell metabolism. However, there are many other heat loss schemes that are disregarded such as physical friction, changes in the cytoskeleton etc.

By fixing the starting point of the trajectory as $x_s(0) = x_s$ and the ending point as $x_s(\tau) = x_d$, then taking the average over all trajectories from x_s to x_d, results in

$$\frac{w(x_s \rightarrow x_d; \tau)}{w(x_d \rightarrow x_s; \tau)} = \langle \exp\left[\beta' \Delta Q^{\tau}_{x_s \rightarrow x_d}\right]\rangle_{x_s \rightarrow x_d}, \tag{5.20}$$

where $w(x_s \rightarrow x_d; \tau)$ is the transition probability that the system is found to be in the microstate x_d at time τ, given that the system was initially in the microstate x_s (for the proof please see Supplementary Material 7).

Now, with the microscopic relation of (5.20), one can study the macroscopic (tissue) consequences of that. Indeed, the probabilistic description of macroscopic states of two distinct cell types, from which one can understand the phenomenon of irreversibility at the macroscopic level, can be constructed as in [44],

$$W(s \rightarrow d) = \int_d \mathrm{d}x_d \int_s \mathrm{d}x_s \, P(x_s \mid s)w(x_s \rightarrow x_d), \tag{5.21}$$

and

$$W(d \rightarrow s) = \int_s \mathrm{d}x_s \int_d \mathrm{d}x_d \, P(x_d \mid d)w(x_d \rightarrow x_s), \tag{5.22}$$

where $P(x_s \mid s)$ is the probability of the system to be in the microstate x_s, given that it is observed in the macrostate s, and $w(x_s \rightarrow x_d)$ is defined as before, where τ is omitted for the notational convenience. The transition probability $W(s \rightarrow d)$ in (5.21) implies the likelihood of the cell to be observed in the macrostate d while it was initially prepared in the macrostate s. Accordingly, Eq. (5.22) is understood in the same fashion, i.e., the likelihood that a microenvironment being prepared in the state d to satisfy the microenvironment s after another time interval τ. These processes are illustrated in Fig. 5.3.

By taking the ratio of (5.21) and (5.22), obtain

$$\frac{W(d \rightarrow s)}{W(s \rightarrow d)} = \frac{\int_s \mathrm{d}x_s \int_d \mathrm{d}x_d [P(x_d \mid d)/P(x_d)] P(x_d)\, w(x_d \rightarrow x_s)}{\int_d \mathrm{d}x_d \int_s \mathrm{d}x_s [P(x_s \mid s)/P(x_s)] P(x_s)\, w(x_s \rightarrow x_d)}, \tag{5.23}$$

where the numerator and the denominator have been multiplied and divided by $P(x_d)$ and $P(x_s)$, respectively. The point wise mutual information is defined for the individual trajectories as $i_1 = \ln[P(x_s \mid s)/P(x_s)]$ and $i_2 = \ln[P(x_d \mid d)/P(x_d)]$, and then by taking these definitions into account and replacing $P(x_d)$ by its correspond-
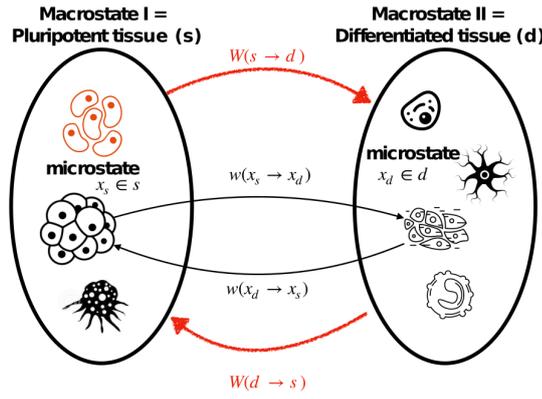
Figure 5.3.. Microscopic and macroscopic transitions between two distinct cell types.

ing relation in (5.18), (5.23) reduces to

$$
\begin{aligned}
\frac{W(d \to s)}{W(s \to d)} &= \frac{\int_s \mathrm{d}x_s \int_d \mathrm{d}x_d e^{i_2}\left[e^{-\beta_d S(y_d|x_d) - \mu_d N(y_d|x_d)}/Z_2\right]w(x_d \to x_s)}{\int_d \mathrm{d}x_d \int_s \mathrm{d}x_s e^{i_1}P(x_s)w(x_s \to x_d)} \\
&= \frac{\int_s \mathrm{d}x_s \int_d \mathrm{d}x_d e^{G}e^{i_2+i_1-i_1}(Z_1/Z_2)\left[e^{-\beta_s S(y_s|x_s) - \mu_s N(y_s|x_s)}/Z_1\right]w(x_d \to x_s)}{\int_d \mathrm{d}x_d \int_s \mathrm{d}x_s e^{i_1}P(x_s)w(x_s \to x_d)},
\end{aligned}
$$

$$(5.24)$$

where

$$
G \equiv -[\beta_d S\left(y_d \mid x_d\right) - \beta_s S\left(y_s \mid x_s\right)] - [\mu_d N\left(y_d \mid x_d\right) - \mu_s N\left(y_s \mid x_s\right)]. \tag{5.25}
$$

By exploiting (5.17) and (5.20), it is rewritten (5.24) as

$$
\frac{W(d \to s)}{W(s \to d)} = \frac{\int_s \mathrm{d}x_s \int_d \mathrm{d}x_d e^{G}e^{i_1}e^{\Delta i}(Z_1/Z_2)P(x_s)\langle e^{-\beta' \Delta Q^{\tau}_{x_s \to x_d}}\rangle_{x_s \to x_d}w(x_s \to x_d)}{\int_d \mathrm{d}x_d \int_s \mathrm{d}x_s e^{i_1}P(x_s)w(x_s \to x_d)}, \tag{5.26}
$$

where $\Delta i \equiv i_2 - i_1$.

Now, (5.26) can be expressed in terms of the average over all trajectories from the ensemble of microstates x_s which correspond to the macrostate s to the ensemble of microstates x_d which correspond to the macrostate d while each path is weighted by its probability as

$$
\frac{W(d \to s)}{W(s \to d)} = \langle\langle\exp\left[-\beta' \Delta Q^{\tau}_{x_s \to x_d}\right]\rangle_{x_s \to x_d} \exp[-\Delta S_{\mathsf{LEUP}}] \exp[-\Delta N] \exp[\Delta i] \exp[\ln(Z_1/Z_2)]\rangle_{s \to d},
$$

$$(5.27)$$

where

$$
\begin{aligned}
\Delta S_{\mathsf{LEUP}} &\equiv \beta_d S(y_d \mid x_d) - \beta_s S(y_s \mid x_s), \\
\Delta N &\equiv \mu_d N(y_d \mid x_d) - \mu_s N(y_s \mid x_s).
\end{aligned} \tag{5.28}
$$

By rearranging (5.27) as

$$
\begin{aligned}
\langle\langle\exp\left[-\beta' \Delta Q^{\tau}_{x_s \to x_d}\right]\rangle_{x_s \to x_d} &\exp[-\Delta S_{\mathsf{LEUP}}] \exp[-\Delta N] \\
&\times \exp[\Delta i] \exp[\ln(Z_1/Z_2)] \exp[\ln\{W(s \to d)/W(d \to s)\}]\rangle_{s \to d} = 1,
\end{aligned}
$$

$$(5.29)$$

and using the Jensen's inequality, i.e., $\exp[\langle X \rangle] \leq \langle \exp[X] \rangle$, and the fact that $e^x \geq 1 + x$, arrive at

$$\langle \beta' \Delta Q^{\tau}_{x_s \to x_d} \rangle_{s \to d} + \langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} + \langle \Delta N \rangle_{s \to d} - \langle \Delta i \rangle_{s \to d} - \ln \left[\frac{Z_1}{Z_2} \right] + \ln \left[\frac{W(d \to s)}{W(s \to d)} \right] \geq 0. \tag{5.30}$$

note that if s and d correspond to the same identical classes, i.e., $W(d \to s) = W(s \to d)$, then the last term of (5.30) vanishes. Now, by defining the remaining terms as the total entropy production, that is,

$$f \equiv \langle \beta' \Delta Q^{\tau}_{x_s \to x_d} \rangle_{s \to d} + \langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} + \langle \Delta N \rangle_{s \to d} - \langle \Delta i \rangle_{s \to d} - \ln \left[\frac{Z_1}{Z_2} \right], \tag{5.31}$$

then (5.30) implies that the total entropy production is always non-negative, in other words, (5.30) can be considered as a generalized Second Law of Thermodynamics (see also Ref. [44]). The above inequality constitutes a fluctuation theorem for tissue differentiation.

The inequality of (5.30) relates thermodynamic properties of the system to the LEUP; and as it is obtained under general assumptions, it is applicable to a general cell and tissue differentiation process. In the next subsection, it is shown how (5.30) implies the robustness of cell differentiation for the particular case of the avian cone photoreceptors.

5.3.1. Application: differentiated photoreceptor mosaics are thermodynamically robust

In this subsection, the fluctuation theorem has been used, culminated in (5.30), to illustrate the robustness of cell differentiation in the case of the avian cone photoreceptors differentiating process from progenitor cells. In particular, By defining $s \to d$ as progenitor cell differentiates to cone cell and denoting its corresponding forward transition probability as p_f, i.e., $W(s \to d) = p_f$, and $d \to s$ as cone cell dedifferentiates to progenitor cell with its backward transition probability of p_b, i.e., $W(d \to s) = p_b$, can rewrite (5.30) in terms of p_f and p_b as

$$\frac{p_f}{p_b} \leq \exp \left[\langle \beta' \Delta Q^{\tau}_{x_s \to x_d} \rangle_{s \to d} + \langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} + \langle \Delta N \rangle_{s \to d} - \langle \Delta i \rangle_{s \to d} - \ln \left[\frac{Z_1}{Z_2} \right] \right], \tag{5.32}$$

where the exponent of the exponential is already defined in the right-hand side of (5.32) as the total entropy production, see (5.31).

In order to obtain a thermodynamic constraint which ensures the robustness of cell differentiation, first assume that there exists a maximum forward transition probability from s to d in such a way that

$$\frac{p_{f\,\mathsf{max}}}{p_b} = \exp \left[\langle \beta' \Delta Q^{\tau}_{x_s \to x_d} \rangle_{s \to d} + \langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} + \langle \Delta N \rangle_{s \to d} - \langle \Delta i \rangle_{s \to d} - \ln \left[\frac{Z_1}{Z_2} \right] \right]. \tag{5.33}$$

5. Cell differentiation and sensing: tissue robustness from optimal environmental sensing

To simplify (5.33) more, note that based on (5.6), it can be written as

$$S\left(x_s\right) = \beta_s\langle S\left(y_s \mid x_s\right)\rangle_{x_s} + \mu_s\langle N\left(y_s \mid x_s\right)\rangle_{x_s} + \ln Z_1,$$
$$S\left(x_d\right) = \beta_d\langle S\left(y_d \mid x_d\right)\rangle_{x_d} + \mu_d\langle N\left(y_d \mid x_d\right)\rangle_{x_d} + \ln Z_2,$$

(5.34)

and then by subtracting these two and taking the average over all trajectories from s to d, obtain

$$\langle\Delta S\rangle_{s\to d} \equiv \langle S(x_d) - S(x_s)\rangle_{s\to d} = \langle\Delta S_{\text{LEUP}}\rangle_{s\to d} + \langle\Delta N\rangle_{s\to d} - \ln\left[\frac{Z_1}{Z_2}\right],$$

(5.35)

where the definitions of ΔS_{LEUP} and ΔN has been used as given in (5.28). Now, (5.33) can be written as

$$\frac{p_{f\max}}{p_b} = \exp\left[\langle\beta'\Delta Q^\tau_{x_s\to x_d}\rangle_{s\to d} + \langle\Delta S\rangle_{s\to d} - \langle\Delta i\rangle_{s\to d}\right].$$

(5.36)

The immediate implication of (5.36) is that the cell differentiation is robust if

$$\langle\beta'\Delta Q^\tau_{x_s\to x_d}\rangle_{s\to d} + \langle\Delta S\rangle_{s\to d} > \langle\Delta i\rangle_{s\to d}.$$

(5.37)

In the following, the avian cone cell differentiation is presented as an example of which the above inequality is satisfied, in other words, the robustness of the avian retina is demonstrated in development and the irreversibility of the time arrow for this particular process.

First, note that as the heat dissipation (the first term in the left-hand side of (5.37)) depends on the metabolic pathways, this implies the crucial role of cell metabolism in the process of cell differentiation. In addition, consumption of glucose depends upon cell types. Progenitor or pluripotent stem cells use glucose as the primary metabolites for anaerobic glycolysis (fermentation) pathway: Glucose + 2 ADP + 2 Phosphate $\to$ 2 Lactate + 2 H$^+$ + 2 ATP, whereas a result of the breakdown of glucose to lactic acid the amount of energy around 109.4 kJ/mol is released [103]. Differentiated cells use glucose to produce carbon dioxide and water by using aerobic glycolysis (respiration) reaction: Glucose + 6 O$_2$ + 36 ADP + 36 Phosphate $\to$ 6 CO$_2$ + 6 H$_2$O + 36 ATP, where the released energy is around 2820 kJ/mol [103]. Thus, $\langle\Delta Q^\tau_{x_s\to x_d}\rangle_{s\to d}$ which is the total heat released during the journey from s to d is 2929.4 kJ/mol. As these reactions have taken place in $T = 310$ K, have: $\langle\beta'\Delta Q^\tau_{x_s\to x_d}\rangle_{s\to d} \approx 9.450$, where the dimension has dropped as the Boltzmann constant is set to 1, i.e., $k_B = 1$.

Now, in order to verify (5.37) for the avian cone cell differentiation, one can calculate the value of $\langle\Delta S\rangle_{s\to d}$ for this particular process. As previously mentioned, Kram et al. [93] have reported the different color percentages of the green, red, blue, violet, and double avian cone cells inside the retina. By assuming these numbers as the probabilities of the corresponding colors, one can write: $P_g \approx 0.204$, $P_r \approx 0.160$, $P_b \approx 0.133$, $P_v \approx 0.094$, and $P_\delta \approx 0.409$. Thus, the entropies can be calculated for the individual cone cells and as a result that of $S(x_d)$ as

$$S(x_d) = -\sum_{i_p} P_{i_p}\ln P_{i_p}$$
$$\approx 0.324 + 0.293 + 0.268 + 0.222 + 0.366$$
$$= 1.473.$$

(5.38)

Now, the entropy difference between differentiated and stem cells reads

$$\langle \Delta S \rangle_{s \to d} = \langle S(x_d) - S(x_s) \rangle_{s \to d} = -0.136, \tag{5.39}$$

where $S(x_s)$ behaves as the entropy of a uniform distribution, i.e.,

$$S(x_s) = -\sum_{i_p=1}^{5} (1/5) \ln(1/5) = \ln 5 \tag{5.40}$$

this is due to the fact that for pluripotent stem cells there are no yet color preferences.

In order cell differentiation to be robust, (5.37) imposes a lower bound on the heat dissipation, as

$$\langle \beta' \Delta Q^\tau_{x_s \to x_d} \rangle_{s \to d} > -\langle \Delta S \rangle_{s \to d}, \tag{5.41}$$

where have set $\langle \Delta i \rangle_{s \to d} \to 0$ for simplicity. (5.41) is strongly holding for the values obtained here, that is, $9.450 \gg 0.136$. This implies that the development of the avian retina is highly robust, in other words, the arrow of time is almost irreversible in this process.

5.4. The limit for cell sensing radius

In this section, a relationship between total entropy production and sensing radius is derived and, in turn, it is applied in the particular case of progenitor cell differentiation into the avian cone photoreceptors. The limits of cell sensing radius is calculated and in the parametric space of the LEUP parameters suggest physically acceptable regions for robust tissue differentiation.

In (5.31), the total entropy production is introduced as,

$$f = \langle \beta' \Delta Q^\tau_{x_s \to x_d} \rangle_{s \to d} + \langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} + \langle \Delta N \rangle_{s \to d} - \langle \Delta i \rangle_{s \to d} - \ln \left[\frac{Z_1}{Z_2} \right], \tag{5.42}$$

where the thermodynamic properties of the system are related to the LEUP quantities.
In order to calculate $\langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d}$, the microenvironmental probability distributions are assumed as Gaussians, thus

$$\langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} = \frac{\beta_d}{2} \ln[2\pi e \sigma_d^2] - \frac{\beta_s}{2} \ln[2\pi e \sigma_s^2], \tag{5.43}$$

where (5.28) is used. The above can be simplified to

$$\langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} = (\beta_d - \beta_s) \ln[2\pi e]^{1/2} + \frac{\beta_d}{2} \ln \sigma_d^2 - \frac{\beta_s}{2} \ln \sigma_s^2. \tag{5.44}$$

Now, the forms of σ_d^2, σ_s^2 and their scaling with the corresponding cell sensing radius are postulated.

5. Cell differentiation and sensing: tissue robustness from optimal environmental sensing

Jiao et al. in [80] have found that the differentiated retina mosaic is hyperuniform, that is, $\sigma_d^2 = V R^A$, where $A < D$, and $D = 2, 3$; on the other hand, for the progenitor cells, assume a Poisson distribution, i.e., $\sigma_s^2 = U R^D$. Here, R is the sensing radius, and V and U are the densities of the cone and progenitor cells' neighbors, respectively. Plugging these formulas in (5.44) leads to

$$\langle \Delta S_{\mathsf{LEUP}} \rangle_{s \to d} = (\beta_d - \beta_s) \ln[2\pi e]^{1/2} + \frac{1}{2} \ln \left[\frac{V^{\beta_d}}{U^{\beta_s}} \right] + \frac{A\beta_d - D\beta_s}{2} \ln R. \tag{5.45}$$

Another term of (5.42) which needs to be dealt with is $\langle \Delta N \rangle_{s \to d}$. From (5.28), have

$$\langle \Delta N \rangle_{s \to d} = \langle \Sigma_{i=1}^5 (\mu_{i,d} N_{i,d} - \mu_{i,s} N_{i,s}) \rangle_{s \to d}, \tag{5.46}$$

where i counts different types of cones: green, red, blue, violet, and double. note that the population at each tissue reads as $N_{i,j} = \rho_{i,j} R^D$, where $\rho_{i,j}$ is the density of the microenvironment and $j \in \{s, d\}$. Thus, (5.46) reduces to

$$\langle \Delta N \rangle_{s \to d} = \langle \Sigma_{i=1}^5 (\mu_{i,d} \rho_{i,d} - \mu_{i,s} \rho_{i,s}) \rangle_{s \to d} R^D \equiv \Delta \tilde{\mu} R^D. \tag{5.47}$$

putting all the terms together, the total entropy production (5.42) becomes

$$f(R, \beta_d, \beta_s) = C_0 + C_1(\beta_d - \beta_s) + \frac{1}{2} \ln \left[\frac{V^{\beta_d}}{U^{\beta_s}} \right] + \frac{A\beta_d - D\beta_s}{2} \ln R + \Delta \tilde{\mu} R^D, \tag{5.48}$$

where it is assumed that $\langle \beta' \Delta Q^\tau_{x_s \to x_d} \rangle_{s \to d}$, $-\langle \Delta i \rangle_{s \to d}$, and $-\ln(Z_1/Z_2)$ are constants and have grouped them together as C_0, and $C_1 \equiv (1/2) \ln(2\pi e)$.

An explicit formula for the total entropy production is written, which can be obtained by the optimal sensing radius for which the total entropy production reaches its extrema. To this end, note that the first derivative of (5.48) vanishes at

$$R_c = \left(\frac{-\Gamma}{\Delta \tilde{\mu} D} \right)^{1/D}, \tag{5.49}$$

where $\Gamma \equiv (A\beta_d - D\beta_s)/2$. As R_c is a positive quantity, the following should always be satisfied: $\Gamma/\Delta \tilde{\mu} < 0$.
In order to determine the conditions for which R_c minimizes or maximizes the total entropy production, calculate the second derivative of (5.48) at (5.49), and obtain

$$\left. \frac{\partial^2 f}{\partial R^2} \right|_{R_c} = -\frac{\Gamma D}{R_c^2}. \tag{5.50}$$

Thus, R_c minimizes the total entropy production if $\Gamma < 0$ and it maximizes if $\Gamma > 0$.

In the left panel of Fig. 5.4, have illustrated the total entropy production (5.48) as a function of R for a particular set of the LEUP parameters which minimizes the total entropy production based on (5.50). The values of the curve above the R-axis (gray line) are biologically relevant as the total entropy production is positive, i.e., it ensures the robustness of the differentiation process. Moreover, this plot shows that as move away from R_c, the (positive) total entropy production is rapidly increasing. The right panel of the figure illustrates, f as a function of R and β_d, where the acceptable regions lie above the dark brown surface.
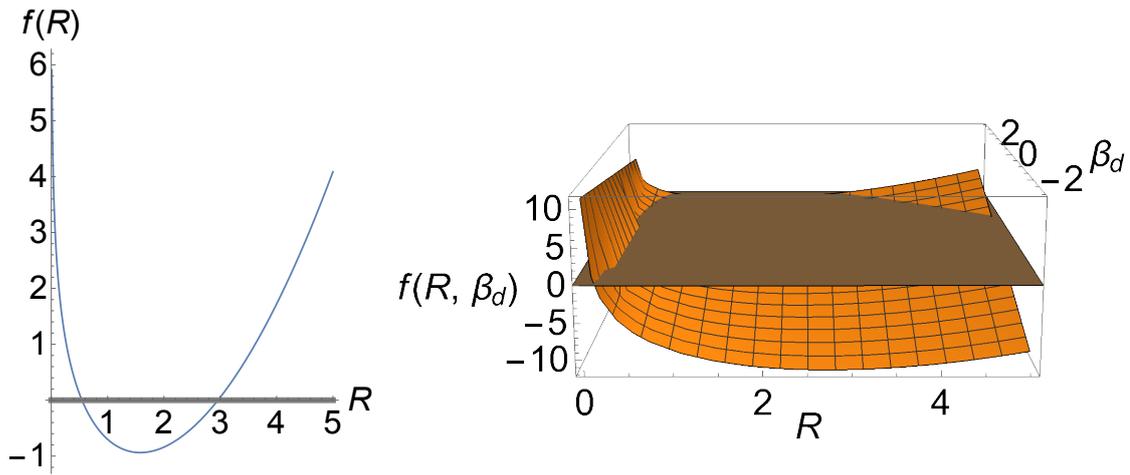
Figure 5.4.. The left panel shows the total entropy production as a function of the sensing radius for $C_0 = -1$, $\beta_d = \beta_s = 3$, $V = U = 1$, $A = 1$, $D = 2$, and $\Delta\tilde{\mu} = 0.3$. Based on (5.50), this specific set of parameters leads to a total entropy production which has a minimum of ≈ -0.937 at $R_c \approx 1.581$. In the right panel, β_d is also treated as a variable. Due to the fact that the total entropy production is always positive, the curve/surface above the gray line/plane is only physically acceptable.

In Fig. 5.5, the case for which the total entropy production reaches its maximum is shown. This figure illustrates that the total entropy production is bounded in this case. The immediate implication of this is tissue de-differentiation is possible (negative entropy production) for large sensing radii. If one want to have a positive total entropy production, in order to avoid reversibility of differentiated tissue, must have some sort of fine-tuning in order to restrict the values of R in such a way that they lead to a positive f. According to Bialek's postulated Biophysical principles, [16] fine-tuning in Nature is not favorable. In the case of avian photoreceptors, the parameters can be further constrained and can be identified further arguments to exclude this case.

5.4.1. Application: The average sensing radius of the avian cone cell

In this subsection, the average sensing radii of the avian cone photoreceptors is calculated and by exploiting (5.48) the regions can be seen in the parametric space of (β_s, β_d) and $(\beta_s, \beta_d, \Delta\tilde{\mu})$ which result in a positive total entropy production and ensure the robustness of the differentiated tissue.

In the retina of a bird, as explained in the previous sections, the five types of photoreceptor cells, namely, green, red, blue, violet, and double cones, form mosaic structures. From the experimental data (see Supporting Information of [93]), have the information about the average standard deviation of the Nearest Neighbor Distribution (NND) for each color as: $\sigma_g \approx 1.248$, $\sigma_r \approx 1.548$, $\sigma_b \approx 1.729$, $\sigma_v \approx 2.292$, and $\sigma_\delta \approx 0.948$.

In addition, Jiao et al. [80] have found how the variance of an avian cone σ^2 is related
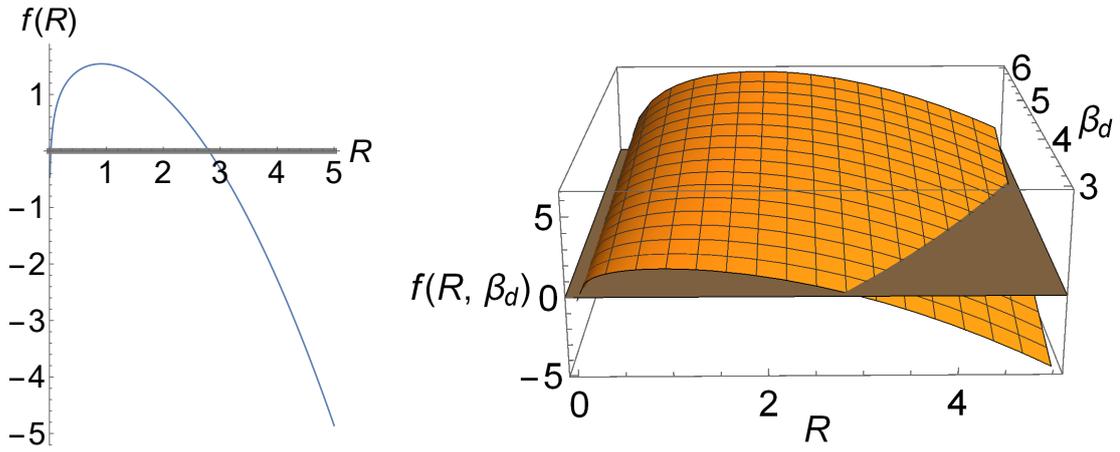
Figure 5.5.. The total entropy production f as a function of R and (R, β_d) is shown in the left and right panels, respectively. In the left panel, have fixed the parameters as: $C_0 = -1$, $\beta_d = 3$, $\beta_s = 1$, $V = U = 1$, $A = 1$, $D = 2$, and $\Delta\tilde{\mu} = -0.3$. $f(R)$ reaches its maximum of ≈ 1.542 at $R_c \approx 0.913$. In the right panel, β_d is also considered as a variable. In both panels, the only physically acceptable regions lie above the gray line and the gray surface, as the total entropy production should be positive; this makes the total entropy production a bounded function in contrast to the case depicted in Fig. 5.4.

to its sensing radius R as

$$\sigma^2(R) = M_1 R^2 + M_2 R \ln R + M_3 R, \tag{5.51}$$

where M_1, M_2, and M_3 have been calculated for each color (see Table I of Ref. [80]). By having this information at the disposal, are able to calculate the average sensing radius for each color as

$$\overline{R}_\delta \approx 0.786, \quad \overline{R}_g \approx 1.426, \quad \overline{R}_r \approx 2.006, \quad \overline{R}_b \approx 2.288, \quad \overline{R}_v \approx 3.553. \tag{5.52}$$

In the beginning of this section, the variance of the differentiated cells are introduced as $\sigma_d^2 = VR^A$; now, by approximating (5.51) to have such a particular form and setting $V = 1$, and using the values of (5.52), can find A for each color as

$$A_\delta \approx 0.440, \quad A_g \approx 1.247, \quad A_r \approx 1.256, \quad A_b \approx 1.322, \quad A_v \approx 1.308, \tag{5.53}$$

where they are in agreement with the assumption of hyperuniformity, that is, $A < 2$.

As an illustration, based on the values of (5.52) and (5.53), the physically acceptable regions of the total entropy production (5.48) are illustrated as in the parametric space of the LEUP parameters: (β_s, β_d) and $(\beta_s, \beta_d, \Delta\tilde{\mu})$ in Fig. 5.6 for the case of the double cone photoreceptor.

5.5. Summary and outlook

In this chapter, LEUP is used as a starting point which later leads to the results of generalised cell differentiation phenomena and apply it in the differentiation of
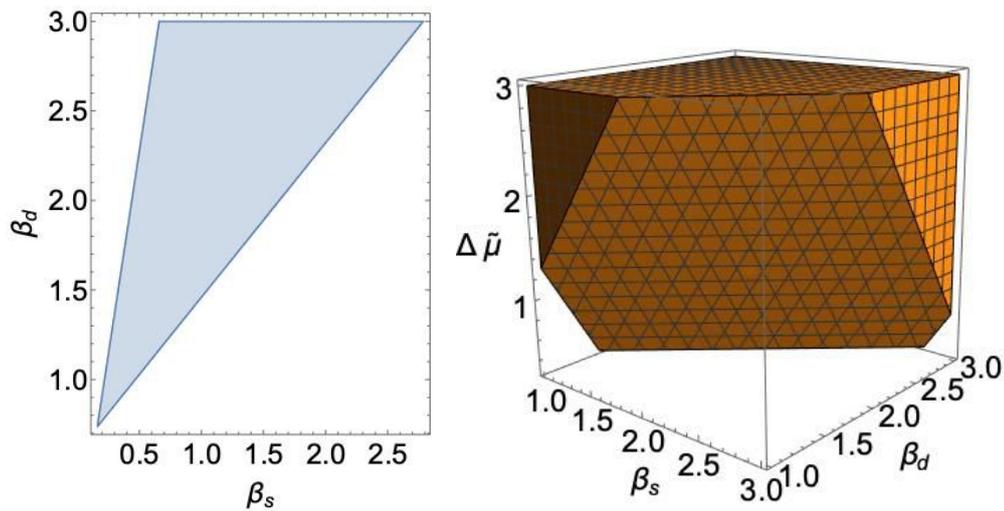
Figure 5.6.. The physically acceptable total entropy production (shaded regions) for the double cone photoreceptor of the avian retina. In the left panel, have fixed: $C_0 = -1$, $V = U = 1$, $D = 2$, and $\Delta\tilde{\mu} = 0.3$. This plot demonstrates the regions where (5.48) is positive in the parametric space of the LEUP parameters (β_s, β_d). note that have also imposed the condition regarding the existence of the optimal sensing radius R_c, which in this case reads as $\Gamma < 0$, see also (5.49) and (5.50). In the right panel, have relaxed the restriction on $\Delta\tilde{\mu}$.

avian photoreceptor mosaics. the question **(Q1)** is posed on how cells coordinate intrinsic and extrinsic variables to determine cell decisions that eventually lead to organized and stable tissues. To tackle this problem, the Least microEnvironmental Uncertainty Principle (LEUP) is employed, which has been recently proposed to understand cell decision-making in multicellular systems, and so far it has been applied to cell migration force distribution [66], collective cell migration [11], and binary phenotypic plasticity [12]. In the context of the LEUP, one can regard differentiation as a sort of Bayesian decision-making, where cells update their intrinsic variables by encoding microenvironmental information and producing relevant responses. This provides a distribution of internal states that depends explicitly on the information of the cell current microenvironment, which is represented by a mesoscopic microenvironmental entropy. Interestingly, it is shown here that local microenvironmental entropy should decrease in time leading to more organized cellular microenvironment, which is the case in differentiated tissues. As a proof of principle, the LEUP challenges the predictions to reproduce differentiated avian photoreceptor mosaics. Although, by fitting a single parameter β, the photoreceptor statistics is successfully reproduced, still this cannot be considered as a rigorous validation. To this end, recently an inter-species collection of photoreceptor mosaics was gathered to further investigate the potential of the LEUP to reproduce these tissues and possibly classify them.

By using the aforementioned results, one can attempt to shed light on the macroscopic transition between pluripotent and differentiated tissues and have specified it to the formation of photoreceptor mosaics, which is related to the question **(Q2)** posed in the Introduction. In this respect, a stochastic thermodynamic-like theory is developed, based on the Crooks' theorem, for a general cell and tissue differentiation process. It is shown that differentiated tissues are highly robust to dedif-

ferentiation, even though individual cells are allowed to go back into pluripotent phenotypes. Biologically, the robustness of differentiated tissues depends on reduced proliferation, change from anaerobic to aerobic metabolism, and increased cell sensing that leads to higher order of microenvironmental organization. In particular, the critical sensing radii of photoreceptor cones are estimated that ensure the thermodynamic robustness of differentiated mosaics, which turns out to be in the range of $0.8\mu m$ to $3.5\mu m$, see (5.52). Please note that the critical radius is the minimal radius required to ensure tissue robustness, and therefore should serve as a lower bound for the real values. Now, if one assume that the minimum sensing radius is equal to the cone size, then the predicted range correlates with the average cone size $2.59 \pm 1.05 \mu m$ (here, the $\pm 3\sigma$ rule was used for the data presented in Fig. 5A of Kram et al. [93]).

In summary, the LEUP-driven model is based on the four crucial assumptions: **(A1)** there is a timescale separation between the internal and microenvironmental variables dynamics, **(A2)** the multicellular system (tissue), where cell is differentiating, follows a Markovian dynamics with the assumption of microscopic reversibility, **(A3)** a flat cell state distribution is assumed for pluripotent cell states, and **(A4)** the spatial distribution of the early microenvironmental pluripotent cells follows a Poisson distribution. Based on these assumptions, one can arrive at three important results: **(I)** predicting the color percentage of the cone cells in the avian retina without any knowledge about the underlying biophysical and biochemical mechanisms, **(II)** demonstrating the robustness of cell-tissue differentiation in thermodynamic terms, and **(III)** determining the limits of the cell sensing radius by establishing a relation between total entropy production and microenvironmental sensing. In the following, these results are further elaborated.

Prediction of the cone color distribution: By calibrating a single parameter in the LEUP, one can predict the cone color percentage in the avian retina accurately. The finding regarding the LEUP parameter which reads as $\beta \approx 1.754$ (close to 2), gives a strong indication that cells sense their environment quasi-optimally when choosing a particular cell fate during differentiation process. For future study and investigation, someone can further examine the validity of this result for photoreceptor mosaics of other species [38, 53, 109, 27].

Robustness and measurement of information gain in differentiation: The fluctuation theorem is constructed for tissue differentiation and have derived a generalization of the second law of thermodynamics for this process based on a Markovian dynamics. It would be interesting and more realistic to relax this assumption and to analyze the problem using the LEUP on the basis of non-Markovian or memory processes [150], which can lead to different results. The only requirement is that system should have a unique stationary state. Please note that the LEUP-driven second law of thermodynamics can also be seen as a generalization of the Bayesian second law of thermodynamics [9] and the conditional second law of thermodynamics in a strongly coupled system [36].

The theory, which in the current paper has been applied to the specific case of the avian photoreceptor mosaics, suggests that differentiated tissue is (highly) thermodynamically robust, that is, the arrow of time is almost irreversible, and this robust-

ness depends on microenvironmental sensing and cell metabolism. It should be remarked that, as is demonstrated in (5.37), if one has the values of $\langle \beta' \Delta Q^\tau_{x_s \to x_d} \rangle_{s \to d}$ and $\langle \Delta S \rangle_{s \to d}$, then someone can determine the upper bound of point wise mutual information difference which is denoted as $\langle \Delta i \rangle_{s \to d}$ within the context of the LEUP. However, in the present work, transitions between equilibrium end-states were considered and have set $\langle \Delta i \rangle_{s \to d} = 0$. In a future study, it is a plan to investigate nonequilibrium dynamics of transitions between progenitor and differentiated cell states and establish the upper bound of information gain in cell differentiation. (please note that the value of $\langle \Delta i \rangle_{s \to d}$ can also be obtained directly from experiments, see Refs. [51, 73, 72, 158].)

Limits of sensing radius: By studying total entropy production as a function of cell sensing radius and the LEUP parameters, an understanding of how a cell regulates its sensing radius according to its microenvironment is provided to ensure the thermodynamic robustness of differentiated tissue. Two cases were shown where **(a)** the entropy production goes to infinity beyond a certain threshold radius which is depicted in Fig. 5.4 and **(b)** the entropy production goes to a maximum value as in Fig. 5.5. Interestingly, it is concluded that the former is the most biologically relevant case since it requires the division time of differentiated cells to be larger than that of the pluripotent ones and biologically systems operate away from fine-tuned parameter regimes to withstand noisy perturbations. On the technical side, one assumption was taken such that the spatial distribution of pluripotent tissue resembles a Poisson distribution. It would be interesting to relax this assumption and to derive this distribution from real tissue data.

One important issue is the range of validity of **(A1)** regarding the timescale separation between cell decision and cell cycle characteristic times. Although cell decisions may seem happening within one cell cycle, the underlying molecular expressions may evolve over many cell cycles [145, 118]. When these molecular expressions cross a threshold, then cell decision emerges very fast. Therefore, the definition of cell decision should be treated with care. In our case, it is specified that the cell decision is considered only when the cell state switches to another dynamic attractor, that induces at the same time some noticeable phenotypic changes. Such attractor transitions are manifested as switches with much shorter characteristic times than a cell cycle [74].

a brief comment on the relation of our theory is made to the commonly used approach of the maximum entropy production (MEP). The MEP formalism explains only the transition from a pluripotent state to a differentiating one, without realizing the corresponding dynamics. To build a connection between these two, one has to construct the LEUP theory for transition paths like maximum caliber principle [55]. Instead of internal and external variables, one uses internal and external paths of the corresponding evolution. Then, by exploiting the formulation of maximum caliber, one can write the time evolution of microenvironmental path entropy – as a conservation equation – in terms of sources and fluxes and subsequently in terms of path action and entropy production [42, 105]. In this regard, one can use maximum caliber principle to construct appropriate transition probabilities and even understand the spatio-temporal dynamics [116]. Finally, one should maximize the internal path entropy, which resembles the MEP approach. Working out

the details of this connection remains for a future work.

The proposed theory has important and interesting implications for cancer research and therapy. In particular, (5.42) states that the balance of metabolic, proliferative, and tissue organization changes (the LEUP term), needed to be taken place in order to destabilize the differentiated state, that is, to promote carcinogenesis. Until now, the majority of the therapies were focused on antiproliferative strategies, such as chemotherapy and radiotherapy, and more seldom to the metabolic conditions such as vasculature normalization. Here, that changes were proposed in the tissue organization, which plays a critical role. This fact has been very recently identified in the context of tumor evolution by West et al. [163]. Please note that in [114], it is realized that microenvironment normalization might be the key for immunotherapeutic success. The mechanistic connection between tissue architecture and cell sensing mechanisms is established in the context of our theory. Strikingly, the experimental work of M. Levin's group [111] shows that disrupting the ion channel sensing in a tissue can induce tumorigenesis. In this regard, one can put forward that investigating changes in the cell sensory processes deserve more attention and might be pivotal in treating cancers. The goal is to calibrate the existing theory to human photoreceptors data, thus one could apply these ideas to retinoblastoma tumors.

In a nutshell, how the LEUP facilitates the inference of cellular intrinsic states (or, cell phenotypes) by means of local microenvironmental entropies or fluctuations was shown. This allows the evaluation of cellular states without the detailed grasp of the underlying mechanisms. The sole knowledge about extrinsic variables distributions (or, collective cell decision-making) suffices. Therefore, one can apply the LEUP to cell differentiation problems where the biological or biophysical knowledge is unclear or unknown.

6. Discussions

Overall in this thesis, Least microEnvironmental Uncertainty Principle (LEUP) has been applied to understand open biological problems which had been discussed in the previous chapters. The internal and microenvironmental variables in the LEUP depend on the set of biological settings or interests.

In summary, in this thesis, LEUP was used to construct the set of mathematical models for different scenarios of cell decision-making to perceive the dynamics of cells at the tissue level. At first, this theory had been used in the area of phenotypic plasticity, which is the archetype for cell decision-making phenomena can be directly seen in chapter 3. The important findings are "Solid" and "Liquid" phases which can be distinguished through supercritical pitchfork in the "go-or-rest" case and the phases later can lead to Turing patterns in the "Go-or-Grow" case. In this chapter, one can understand how individual cell decision influences the tissue dynamics at the macroscopic level. Specifically, LEUP helps here to figure out the coarse-grained switching mechanism in Epithelial-Mesenchymal transition, which was driven by the microenvironment. After that, LEUP was used to understand collective cell migration mechanisms, which is the paradigm of collective orientation decision-making. To further figure out, a set of Langevin's equations for collective cell migration was built in chapter 4 to explain the emergence of pattern formations when the interactions among the cells are unknown. One can understand the influence and the adaptation of the neighbouring cells in the role of pattern formation. In addition, the aspect of the magnitude of sensitivity and sensing radius on pattern formation was examined. The macroscopic observables from the mathematical model were verified with experimental data sets. Interestingly, Vicsek model can be seen as a special case of LEUP driven mathematical model. At last, the tissue robustness at the thermodynamic level had been explained from the combined theory of arrow of time and LEUP in chapter 5. The probability distribution for each avian photoreceptor has been calculated from the LEUP theory, which later verified with respect to the experimental data sets. Also, the thermodynamic feasibility of cell differentiation can be figured out through the mathematical relation between entropy production and sensing radius. Subsequently, one can verify the notion of coarse-grained sensing radius with the spatial correlation func-

tion of external states around the cell. It can assist to observe the structure of the microenvironment.

The LEUP framework is a coarse-grained model which can be used to understand unknown biological facts when the mechanism behind the phenomena became illusive. Specifically, this framework can be used to understand biological phenomena when the precision about biophysical knowledge and inter-pretablity is fractional. The framework helps to find a mapping to phenomenological models which further connect the mechanistic models. In the future, one can connect this scheme with machine learning models to figure out the areas which contain less biophysical knowledge and less inter-pretablity.

It will be interesting to see if the non-ergodic recipe for microenvironmetal can make an influence on cell decision-making phenomena. Because of the size of the microenvironment and strong coupling between neighboring cells, the ergodicity of the microenvironment breaks. Due to this reason, one can formulate the LEUP framework using different definitions of entropy (i.e., Tsallis entropy, Renyi entropy etc.). This can be examined in future to gain a better understanding of cellular dynamics at the tissue level.

From an algorithmic point of view, one can find this groundwork as a strategy to develop the local search based algorithms which calculate the information of their neighborhood/microenvironment to update their knowledge for future decisions. Due to the local nature of the algorithm, the corresponding performance would depend on the interaction radius of the cell.

It will be fascinating in the future, if one can measure the minimum number of neighboring cells and their corresponding states which are responsible to make optimal decisions. Dynamics in cell decision-making play a huge role in understanding the robustness of decision-making. For further explanation, one can construct a Fokker-Planck equations from the microenvironmental entropy to see the evolution of time-dependent probability distribution of the internal states. In spatio-temporal cellular decisions, mutations can bring an extra piece of complexity. The appearance of mutations is sometimes beneficial and sometimes obnoxious. This supplementary piece of information can be used inside the variational principle as a constraint to figure out the rare decisions taken by the cell.

Another alluring point is that when the complex microenvironment in the vicinity of the cell is dynamic and fluctuating. In this special scenario, cells can grasp a decision in a memory dependent manner. So, the role of non-Markovian processes will come to the current coarse-grained framework. Memories can help cells to make better decisions. It can be hidden in the intra-cellular biological processes (i.e., Histone modifications, gene regulations etc.) and/or inter-cellular processes (i.e., the media/microenvironment where the cells take a decision is visco-elastic). It precisely tells us that there exists a lag in time during the flow of information through the cell while gathering information from the microenvironment. In a more precise way, one can say that the feedback between the microenvironment to cell and from cell to the microenvironment consists of a time difference. To test this lag, one can experimentally verify the response functions of the cell over time.

Bibliography

[1] S. Abe. "Conditional maximum-entropy method for selecting prior distributions in Bayesian statistics". In: *EPL* 108.4 (2014), p. 40008. DOI: 10.1209/0295-5075/108/40008.

[2] B. Alberts et al. *Molecular Biology of the Cell*. Sixth. New York City, NY: W. W. Norton & Company, 2014.

[3] M. Allen. "Compelled by the Diagram: Thinking through C. H. Waddington's Epigenetic Landscape". In: *Contemporaneity: Historical Presence in Visual Culture* 4 (2015), pp. 119–142. DOI: 10.5195/contemp.2015.143.

[4] G. Ariel et al. "Collective dynamics of two-dimensional swimming bacteria: Experiments and models". In: *Physical Review E* 98.3 (2018), p. 032415.

[5] S. Artavanis-Tsakonas, M. D. Rand, and R. J. Lake. "Notch Signaling: Cell Fate Control and Signal Integration in Development". In: *Science* 284.5415 (1999), pp. 770–776. ISSN: 0036-8075. DOI: 10.1126/science.284.5415.770. eprint: https://science.sciencemag.org/content/284/5415/770.full.pdf. URL: https://science.sciencemag.org/content/284/5415/770.

[6] J. Banyard and D. R. Bielenberg. "The role of EMT and MET in cancer dissemination". In: *Conn. Tis. Res.* 56.5 (2015). PMID: 26291767, pp. 403–413. DOI: 10.3109/03008207.2015.1060970. eprint: https://doi.org/10.3109/03008207.2015.1060970. URL: https://doi.org/10.3109/03008207.2015.1060970.

[7] A. Bar, Y. Kafri, and D. Mukamel. "Dynamics of DNA melting". In: *Journal of Physics: Condensed Matter* 21.3 (2008), p. 034110.

[8] G. Barriere et al. "Epithelial Mesenchymal Transition: a double-edged sword". In: *Clinic.Trans. Med.* 4.1 (2015), p. 14. ISSN: 2001-1326. DOI: 10.1186/s40169-015-0055-4. URL: https://doi.org/10.1186/s40169-015-0055-4.

[9] A. Bartolotta et al. "Bayesian second law of thermodynamics". In: *Phys. Rev. E* 94 (2016), p. 022102. DOI: 10.1103/PhysRevE.94.022102.

[10] A. Barua, A. Beygi, and H. Hatzikirou. "Close to Optimal Cell Sensing Ensures the Robustness of Tissue Differentiation Process: The Avian Photoreceptor Mosaic Case". In: *Entropy* 23.7 (2021). ISSN: 1099-4300. DOI: 10.3390/e23070867. URL: https://doi.org/10.3390/e23070867.

[11] A. Barua et al. "A least microenvironmental uncertainty principle (LEUP) as a generative model of collective cell migration mechanisms". In: *Sci. Rep.* 10 (2020), p. 22371. DOI: 10.1038/s41598-020-79119-y.

[12] A. Barua et al. "Entropy-driven cell decision-making predicts 'fluid-to-solid' transition in multicellular systems". In: *New J. Phys.* 22 (2020), p. 123034. DOI: 10.1088/1367-2630/abcb2e.

[13] D. A. Bennett and W. Tang. "Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone". In: *International Journal of Geographical Information Science* 20.9 (2006), pp. 1039–1066.

[14] H. C. Berg and E. M. Purcell. "Physics of chemoreception". In: *Biophys. J.* 20.2 (1977), pp. 193–219. DOI: 10.1016/S0006-3495(77)85544-6.

[15] P. L. Bhatnagar, E. P. Gross, and M. Krook. "A Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems". In: *Phys. Rev.* 94 (3 1954), pp. 511–525. DOI: 10.1103/PhysRev.94.511. URL: https://link.aps.org/doi/10.1103/PhysRev.94.511.

[16] W. Bialek. *Biophysics: Searching for principles*. Princeton University Press, 2012. ISBN: 9780691138916.

[17] W. Bialek and S. Setayeshgar. "Physical limits to biochemical signaling". In: *PNAS* 102.29 (2005), pp. 10040–10045. DOI: 10.1073/pnas.0504321102.

[18] H. M. Blau, T. R. Brazelton, and J. M. Weimann. "The Evolving Concept of a Stem Cell: Entity or Function?" In: *Cell* 105.7 (2001), pp. 829–841. DOI: 10.1016/S0092-8674(01)00409-3.

[19] M. Boareto et al. "Jagged–Delta asymmetry in Notch signaling can give rise to a Sender/Receiver hybrid phenotype". In: *PNAS* 112.5 (2015), E402–E409. ISSN: 0027-8424. DOI: 10.1073/pnas.1416287112. eprint: https://www.pnas.org/content/112/5/E402.full.pdf. URL: https://www.pnas.org/content/112/5/E402.

[20] M. Boareto et al. "Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype". In: *J. R. Soc. Int.* 13 (2016), pp. 1–11. DOI: 10.1098/rsif.2015.1106.

[21] F. Bocci, J. N. Onuchic, and M. K. Jolly. "Understanding the Principles of Pattern Formation Driven by Notch Signaling by Integrating Experiments and Theoretical Models". In: *Frontiers in Physiology* 11 (2020), p. 929. ISSN: 1664-042X. DOI: 10.3389/fphys.2020.00929. URL: https://www.frontiersin.org/article/10.3389/fphys.2020.00929.

[22] K. Böttger et al. "An Emerging Allee Effect Is Critical for Tumor Initiation and Persistence". In: *PLOS Comp. Bio.* 11.9 (Sept. 2015). Ed. by M. S. Alber, e1004366. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004366. URL: http://dx.plos.org/10.1371/journal.pcbi.1004366.

[23] F. Bouchet and T. Dauxois. "Kinetics of anomalous transport and algebraic correlations in a long-range interacting system". In: *Journal of Physics: Conference Series*. Vol. 7. IOP Publishing. 2005, p. 34.

[24] G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. 1979.

[25] R. Bremner and J. Sage. "The origin of human retinoblastoma". In: *Nature* 514 (2014), pp. 312–313. DOI: 10.1038/nature13748.

[26] J. A. Brzezinski and T. A. Reh. "Photoreceptor cell fate specification in vertebrates". In: *Development* 142.19 (2015), pp. 3263–3273. DOI: 10.1242/dev.127043.

[27] J. M. Bueno et al. "Arrangement of the photoreceptor mosaic in a diabetic rat model imaged with multiphoton microscopy". In: *Biomed. Opt. Express* 11 (2020), pp. 4901–4914. DOI: 10.1364/BOE.399835.

[28] P. Cahan and G. Q. Daley. "Origins and implications of pluripotent stem cell variability and heterogeneity". In: *Nat. Rev. Mol. Cell Biol.* 14.6 (2013), pp. 357–368. DOI: 10.1038/nrm3584.

[29] M. J. Casey, P. S. Stumpf, and B. D. MacArthur. "Theory of cell fate". In: *Wiley Interdiscip. Rev.: Syst. Biol. Med.* (2020). DOI: 10.1002/wsbm.1471.

[30] A. V. Chechkin and R. Klages. "Fluctuation relations for anomalous dynamics". In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.03 (2009), p. L03002.

[31] S. Ciliberto. "Experiments in Stochastic Thermodynamics: Short History and Perspectives". In: *Phys. Rev. X* 7 (2017), p. 021051. DOI: 10.1103/PhysRevX.7.021051.

[32] S. Ciliberto. "From space-time chaos to stochastic thermodynamics". In: *CR Phys.* 20.6 (2019), pp. 529–539. DOI: 10.1016/j.crhy.2019.09.001.

[33] D. Cobrinik. "Retinoblastoma Progression". In: *EBioMed.* 2.7 (2015), pp. 623–624. DOI: 10.1016/j.ebiom.2015.07.023.

[34] A. Costanzo and C. K. Hemelrijk. "Spontaneous emergence of milling (vortex state) in a Vicsek-like model". In: *Journal of Physics D: Applied Physics* 51.13 (Mar. 2018), p. 134004. ISSN: 0022-3727. DOI: 10.1088/1361-6463/aab0d4. URL: https://doi.org/10.1088/1361-6463/aab0d4.

[35] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". In: *Phys. Rev. E* 60 (3 1999), pp. 2721–2726. DOI: 10.1103/PhysRevE.60.2721.

[36] G. E. Crooks and S. Still. "Marginal and conditional second laws of thermodynamics". In: *EPL* 125.4 (2019), p. 40005. DOI: 10.1209/0295-5075/125/40005.

[37] E. Crosato et al. "Thermodynamics and computation during collective motion near criticality". In: *Physical Review E* 97.1 (2018), p. 012120.

[38] C. A. Curcio and K. R. Sloan. "Packing geometry of human cone photoreceptors: Variation with eccentricity and evidence for local anisotropy". In: *Visual Neurosci.* 9.2 (1992), pp. 169–180. DOI: 10.1017/S0952523800009639.

[39] E. Dahlman and Y. Watanabe. "SU-F-T-109: A Shortcoming of the Fisher-Kolmogorov Reaction-Diffusion Equation for Modeling Tumor Growth". In: *Med. Phys.* 43.6Part14 (2016), pp. 3486–3487. DOI: 10.1118/1.4956245. eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4956245. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4956245.

[40] A. Dechant et al. "Solution of the Fokker-Planck equation with a logarithmic potential". In: *Journal of Statistical Physics* 145.6 (2011), pp. 1524–1545.

[41] A. Del Sol et al. "Big-Data-Driven Stem Cell Science and Tissue Engineering: Vision and Unique Opportunities". In: *Cell Stem Cell* 20.2 (2017), pp. 157–160. DOI: 10.1016/j.stem.2017.01.006.

[42] R. Dewar. "Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states". In: *J. Phys. A: Math. Gen.* 36 (2003), pp. 631–641. DOI: 10.1088/0305-4470/36/3/303.

[43] M. Durve, F. Peruani, and A. Celani. "Learning to flock through reinforcement". In: *Physical Review E* 102.1 (2020), p. 012601.

[44] J. L. England. "Statistical physics of self-replication". In: *J. Chem. Phys.* 139.12 (2013), p. 121923. DOI: 10.1063/1.4818538.

[45] A. A. Evans et al. "Orientational order in concentrated suspensions of spherical microswimmers". In: *Physics of Fluids* 23.11 (2011), p. 111702.

[46] D. J. Evans, E. G. D. Cohen, and G. P. Morriss. "Probability of second law violations in shearing steady states". In: *Phys. Rev. Lett.* 71 (1993), p. 2401. DOI: 10.1103/PhysRevLett.71.2401.

[47] D. J. Evans and D. J. Searles. "The Fluctuation Theorem". In: *Adv. Phys.* 51.7 (2002), pp. 1529–1585. DOI: 10.1080/00018730210155133.

[48] J. E. Ferrell. "Bistability, Bifurcations, and Waddington's Epigenetic Landscape". In: *Curr. Biol.* 22.11 (2012), R458–R466. DOI: 10.1016/j.cub.2012.03.045.

[49] D. Friedmann-Morvinski and I. M. Verma. "Dedifferentiation and reprogramming: origins of cancer stem cells". In: *EMBO Rep.* 15.3 (2014), pp. 244–253. DOI: 10.1002/embr.201338254.

[50] K. Friston. "Is the free-energy principle neurocentric?" In: *Nat. Rev. Neurosci.* 11.8 (2010), p. 605. DOI: 10.1038/nrn2787-c2.

[51] D. Fuller et al. "External and internal constraints on eukaryotic chemotaxis". In: *PNAS* 107.21 (2010), pp. 9656–9659. DOI: 10.1073/pnas.0911178107.

[52] C. Furusawa and K. Kaneko. "A Dynamical-Systems View of Stem Cell Biology". In: *Science* 338.6104 (2012), pp. 215–217. DOI: 10.1126/science.1224311.

[53] L. Galli-Resta et al. "Modelling the mosaic organization of rod and cone photoreceptors with a minimal-spacing rule". In: *Eur. J. Neurosci.* 11.4 (1999), pp. 1461–1469. DOI: 10.1046/j.1460-9568.1999.00555.x.

[54] J. Garcia-Ojalvo and A. Martinez Arias. "Towards a statistical mechanics of cell fate decisions". In: *Curr. Opin. Genet. Dev.* 22.6 (2012), pp. 619–626. DOI: 10.1016/j.gde.2012.10.004.

[55] K. Ghosh et al. "The Maximum Caliber Variational Principle for Nonequilibria". In: *Ann. Rev. Phys. Chem.* 71 (2020), pp. 213–238. DOI: 10.1146/annurev-physchem-071119-040206.

[56] A. Giese et al. "Cost of Migration: Invasion of Malignant Gliomas and Implications for Treatment". In: *Journal of Clinical Oncology* 21.8 (Apr. 2003), pp. 1624–1636. ISSN: 0732-183X. DOI: 10.1200/JCO.2003.05.063. URL: https://doi.org/10.1200/JCO.2003.05.063.

[57] A. Giese et al. "Migration of human glioma cells on myelin". eng. In: *Neurosurgery* 38.4 (Apr. 1996), pp. 755–764. ISSN: 0148-396X. DOI: 10.1227/00006123-199604000-00026. URL: https://doi.org/10.1227/00006123-199604000-00026.

[58] A. Giese et al. "Dichotomy of astrocytoma migration and proliferation". In: *International Journal of Cancer* 67.2 (July 1996), pp. 275–282. ISSN: 0020-7136. DOI: 10.1002/(SICI)1097-0215(19960717)67:2<275::AID-IJC20>3.0.CO;2-9. URL: https://doi.org/10.1002/(SICI)1097-0215(19960717)67:2%3C275::AID-IJC20%3E3.0.CO;2-9.

[59] J. Godlewski et al. "MicroRNA-451 Regulates LKB1/AMPK Signaling and Allows Adaptation to Metabolic Stress in Glioma Cells". In: *Molecular Cell* 37.5 (Mar. 2010), pp. 620–632. ISSN: 1097-2765. URL: http://www.sciencedirect.com/science/article/pii/S1097276510001693.

[60] A. D. Goldberg, C. D. Allis, and E. Bernstein. "Epigenetics: A Landscape Takes Shape". In: *Cell* 128.4 (2007), pp. 635–638. DOI: 10.1016/j.cell.2007.02.006.

[61] P. Greulich, R. Smith, and B. D. MacArthur. "Chapter 8: The physics of cell fate". In: *Phenotypic Switching*. Academic Press, 2020, pp. 189–206.

[62] R. Grossmann, L. Schimansky-Geier, and P. Romanczuk. "Self-propelled particles with selective attraction–repulsion interaction: from microscopic dynamics to coarse-grained theories". In: *New Journal of Physics* 15.8 (2013), p. 085014.

[63] V. Hakim and P. Silberzan. "Collective cell migration: a physics perspective". In: *Rep. on Prog. in Phys.* 80.7 (Apr. 2017), p. 076601. DOI: 10.1088/1361-6633/aa65ef. URL: https://doi.org/10.1088%2F1361-6633%2Faa65ef.

[64] D. Hanahan and R. A. Weinberg. "The Hallmarks of Cancer". In: *Cell* 100.1 (2000), pp. 57–70. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(00)81683-9. URL: https://doi.org/10.1016/S0092-8674(00)81683-9.

[65] H. Hatzikirou et al. "'Go or Grow': the key to the emergence of invasion in tumour progression?" In: *Math. Med. Biol.* 29.1 (Mar. 2010), pp. 49–65. ISSN: 1477-8599. DOI: 10.1093/imammb/dqq011. URL: https://academic.oup.com/imammb/article-lookup/doi/10.1093/imammb/dqq011.

[66] H. Hatzikirou. "Statistical mechanics of cell decision-making: the cell migration force distribution". In: *J. Mech. Behav. Mater.* 27.1-2 (Apr. 2018), pp. 1–7. ISSN: 2191-0243. DOI: 10.1515/jmbm-2018-0001. URL: http://www.degruyter.com/view/j/jmbm.2018.27.issue-1-2/jmbm-2018-0001/jmbm-2018-0001.xml.

[67] E. Heller and E. Fuchs. "Tissue patterning and cellular mechanics". In: *J. Cell Bio.* 211.2 (2015), pp. 219–231. ISSN: 15408140. DOI: 10.1083/jcb.201506106.

[68] M. Herberg and I. Roeder. "Computational modelling of embryonic stem-cell fate control". In: *Development* 142 (2015), pp. 2250–2260. DOI: 10.1242/dev.116343.

[69] L. Herbert et al., eds. *Phenotypic switching: implications in biology and medicine*. 1st ed. Cambridge: Elsevier, 2020. ISBN: 9780128179963.

[70] N. M. Hideitsu Hino Kensuke Koshi jim. "Non-parametric entropy estimators based on simple linear regression". In: *Computational Statistics & Data Analysis* 89 (2015), pp. 72–84.

[71] K. S. Hoek et al. "In vivo switching of human melanoma cells between proliferative and invasive states". In: *Cancer Res.* 68.3 (Feb. 2008), pp. 650–656. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-07-2491. URL: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-07-2491.

[72] B. Hu et al. "Phenomenological approach to eukaryotic chemotactic efficiency". In: *Phys. Rev. E* 81 (2010), p. 031906. DOI: 10.1103/PhysRevE.81.031906.

[73] B. Hu et al. "Physical Limits on Cellular Sensing of Spatial Gradients". In: *Phys. Rev. Lett.* 105 (2010), p. 048104. DOI: 10.1103/PhysRevLett.105.048104.

[74] S. Huang et al. "Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network". In: *Phys. Rev. Lett.* 94 (2005), pp. 1–4. DOI: 10.1103/PhysRevLett.94.128701.

[75] T. Ishikawa. "Suspension biomechanics of swimming microbes". In: *Journal of The Royal Society Interface* (2009), rsif20090223.

[76] E. T. Jaynes. "Information Theory and Statistical Mechanics". In: *Phys. Rev.* 106 (4 May 1957), pp. 620–630. DOI: 10.1103/PhysRev.106.620. URL: https://link.aps.org/doi/10.1103/PhysRev.106.620.

[77] E. T. Jaynes. "Information Theory and Statistical Mechanics. II". In: *Phys. Rev.* 108 (2 Oct. 1957), pp. 171–190. DOI: 10.1103/PhysRev.108.171. URL: https://link.aps.org/doi/10.1103/PhysRev.108.171.

[78] E. T. Jaynes. "Prior Probabilities". In: *IEEE Trans. Sys. Sci. Cyb.* 4.3 (1968), pp. 227–241. DOI: 10.1109/TSSC.1968.300117.

[79] L. Jerby et al. "Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer". In: *Cancer Res.* 72.22 (Nov. 2012), pp. 5712–5720. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-12-2215. URL: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-12-2215.

[80] Y. Jiao et al. "Avian photoreceptor patterns represent a disordered hyperuniform solution to a multiscale packing problem". In: *Phys. Rev. E* 89 (2 2014), p. 022721. DOI: 10.1103/PhysRevE.89.022721.

[81] Y. Jiao et al. "Spatial Organization and Correlations of Cell Nuclei in Brain Tumors". In: *PLOS ONE* 6.11 (Nov. 2011), pp. 1–9. DOI: 10.1371/journal.pone.0027323. URL: https://doi.org/10.1371/journal.pone.0027323.

[82] A. Jilkine. "Mathematical Models of Stem Cell Differentiation and Dedifferentiation". In: *Curr. Stem Cell Rep.* 5.2 (2019), pp. 66–72. DOI: 10.1007/s40778-019-00156-z.

[83] M. K. Jolly and H. Levine. "Computational systems biology of epithelial-hybrid-mesenchymal transitions". In: *Current Opinion in Systems Biology* 3 (June 2017), pp. 1–6. ISSN: 2452-3100. URL: http://www.sciencedirect.com/science/article/pii/S2452310016300191.

[84] M. K. Jolly et al. "Towards elucidating the connection between epithelial - mesenchymal transitions and stemness". In: *J. Roy. Soc. Int.* 11.101 (2014), p. 20140962. DOI: 10.1098/rsif.2014.0962. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2014.0962. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2014.0962.

[85] D. L. Jones and A. J. Wagers. "No place like home: anatomy and function of the stem cell niche". In: *Nature Reviews Molecular Cell Biology* 9.1 (2008), pp. 11–21. DOI: 10.1038/nrm2319. URL: https://doi.org/10.1038/nrm2319.

[86] R. Kalluri and R. A. Weinberg. "The basics of epithelial-mesenchymal transition". In: *The J. of Cli. Inves.* 120.5 (May 2010), pp. 1786–1786. DOI: 10.1172/JCI39104C1. URL: https://www.jci.org/articles/view/39104C1.

[87] D. A. Kessler and E. Barkai. "Infinite covariant density for diffusion in logarithmic potentials and optical lattices". In: *Physical review letters* 105.12 (2010), p. 120602.

[88] S. J. King et al. "Lamellipodia are crucial for haptotactic sensing and response". In: *J. Cell Sci.* 129.12 (2016), pp. 2329–2342. DOI: 10.1242/jcs.184507.

[89] R. Klages, W. Just, and C. Jarzynski, eds. *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond*. Reviews of Nonlinear Dynamics and Complexity. Weinheim, Germany: Wiley-VCH, 2013.

[90] D. C. Knill and A. Pouget. "The Bayesian brain: the role of uncertainty in neural coding and computation". In: *Trends Neurosci.* 27.12 (2004), pp. 712–719. DOI: 10.1016/j.tins.2004.10.007.

[91] A. Q. Kohrman and D. Q. Matus. "Divide or Conquer: Cell Cycle Regulation of Invasive Behavior". In: *T. Cell Bio.* 27.1 (2017), pp. 12–25. ISSN: 18793088. DOI: 10.1016/j.tcb.2016.08.003. URL: http://dx.doi.org/10.1016/j.tcb.2016.08.003.

[92] I. E. Kooi et al. "Loss of photoreceptorness and gain of genomic alterations in retinoblastoma reveal tumor progression". In: *EBioMed.* 2.7 (2015), pp. 660–670. DOI: 10.1016/j.ebiom.2015.06.022.

[93] Y. A. Kram, S. Mantey, and J. C. Corbo. "Avian Cone Photoreceptors Tile the Retina as Five Independent, Self-Organizing Mosaics". In: *PLoS ONE* 5.2 (2010), pp. 1–13. DOI: 10.1371/journal.pone.0008992.

[94] R. Lanza. "Preface". In: *Essentials of Stem Cell Biology (Third Edition)*. Ed. by R. Lanza and A. Atala. Third Edition. Boston: Academic Press, 2014, p. xxi. ISBN: 978-0-12-409503-8. DOI: https://doi.org/10.1016/B978-0-12-409503-8.00046-9. URL: http://www.sciencedirect.com/science/article/pii/B9780124095038000

[95] D. A. Lauffenburger and J. J. Linderman. *Receptors: Models for Binding, Trafficking, and Signaling*. Oxford University Press, 1996.

[96] P. Li, M. Gu, and H. Xu. "Lysosomal Ion Channels as Decoders of Cellular Signals". In: *Trends Biochem. Sci.* 44.2 (2019), pp. 110–124. DOI: 10.1016/j.tibs.2018.10.006.

[97]   A. C. Lloyd. "The Regulation of Cell Size". In: *Cell* 154.6 (2013), pp. 1194–1205. DOI: 10.1016/j.cell.2013.08.053.

[98]   E. Lushi, H. Wioland, and R. E. Goldstein. "Fluid flows created by swimming bacteria drive self-organization in confined suspensions". In: *Proceedings of the National Academy of Sciences* (2014), p. 201405698.

[99]   E. Lutz. "Power-law tail distributions and nonergodicity". In: *Physical review letters* 93.19 (2004), p. 190602.

[100] B. D. MacArthur and I. R. Lemischka. "Statistical Mechanics of Pluripotency". In: *Cell* 154.3 (2013), pp. 484–489. DOI: 10.1016/j.cell.2013.07.024.

[101] M. Mak et al. "Single-Cell Migration in Complex Microenvironments: Mechanics and Signaling Dynamics". In: *Journal of Biomechanical Engineering* 138.2 (Jan. 2016). 021004. ISSN: 0148-0731. DOI: 10.1115/1.4032188. URL: https://doi.org/10.1115/1.4032188.

[102] C. Malinverno et al. "Endocytic reawakening of motility in jammed epithelia". In: *Nature Mat.* 16.5 (2017), pp. 587–596. ISSN: 1476-4660. DOI: 10.1038/nmat4848. URL: https://doi.org/10.1038/nmat4848.

[103] D. Marién and B. Sabater. "The cancer Warburg effect may be a testable example of the minimum entropy production rate principle". In: *Phys. Biol.* 14.2 (2017), p. 024001. DOI: 10.1088/1478-3975/aa64a7.

[104] F. Martino et al. "Cellular Mechanotransduction: From Tension to Function". In: *Front. Physiol.* 9 (2018), pp. 1–21. DOI: 10.3389/fphys.2018.00824.

[105] L. M. Martyushev and V. D. Seleznev. "Maximum entropy production principle in physics, chemistry and biology". In: *Phys. Rep.* 426.1 (2006), pp. 1–45. DOI: 10.1016/j.physrep.2005.12.001.

[106] F. Matthäus, M. Jagodič, and J. Dobnikar. "E. coli superdiffusion and chemotaxis—search strategy, precision, and motility". In: *Biophysical journal* 97.4 (2009), pp. 946–957.

[107] R. Mayor. "Cell fate decisions during development". In: *Science* 364.6444 (2019), pp. 937–938. ISSN: 0036-8075. DOI: 10.1126/science.aax7917.

[108] S. Milster et al. "Eliminating inertia in a stochastic model of a micro-swimmer with constant speed". In: *The Euro. Phys. Jour. Sp. Topics* 226.9 (2017), pp. 2039–2055. ISSN: 1951-6401. DOI: 10.1140/epjst/e2017-70052-8. URL: https://doi.org/10.1140/epjst/e2017-70052-8.

[109] M. Mitkus et al. "Retinal ganglion cell topography and spatial resolution of two parrot species: budgerigar (Melopsittacus undulatus) and Bourke's parrot (Neopsephotus bourkii)". In: *J. Comp. Physiol. A* 200 (2014), pp. 371–384. DOI: 10.1007/s00359-014-0894-2.

[110] M. Mojtahedi et al. "Cell Fate Decision as High-Dimensional Critical State Transition". In: *PLoS Biol.* 14.12 (2016), pp. 1–28. DOI: 10.1371/journal.pbio.2000640.

[111] D. Moore, S. I. Walker, and M. Levin. "Cancer as a disorder of patterning information: computational and biophysical perspectives on the cancer problem". In: *Converg. Sci. Phys. Oncol.* 3.4 (2017), p. 043001. DOI: 10.1088/2057-1739/aa8548.

[112] T. Mora and I. Nemenman. "Physical Limit to Concentration Sensing in a Changing Environment". In: *Phys. Rev. Lett.* 123 (2019), p. 198101. DOI: 10.1103/PhysRevLett.123.198101.

[113] S. J. Morrison and J. Kimble. "Asymmetric and symmetric stem-cell divisions in development and cancer". In: *Nature* 441 (2006), pp. 1068–1074. DOI: 10.1038/nature04956.

[114] Y. R. Murciano-Goroff, A. B. Warner, and J. D. Wolchok. "The future of cancer immunotherapy: microenvironment-targeting combinations". In: *Cell Res.* 30 (2020), pp. 507–519. DOI: 10.1038/s41422-020-0337-2.

[115] J. D. Murray. *Mathematical Biology II. Spatial Models and Biomedical Applications*. Ed. by J. D. Murray. 3rd ed. Vol. 18. Interdisciplinary Applied Mathematics. New York, NY: Springer New York, 2003, p. 1067. ISBN: 978-0-387-95228-4. DOI: 10.1007/b98869. URL: http://www.springerlink.com/index/10.1007/978-0-387-75847-3%20http://link.springer.com/10.1007/b98869.

[116] J. M. Nava-Sedeño et al. "Cellular automaton models for time-correlated random walks: derivation and analysis". In: *Sci. Rep.* 7 (2017), pp. 1–13. DOI: 10.1038/s41598-017-17317-x.

[117] J. M. Nava-sedeño et al. "Cellular automaton models for time-correlated random walks: derivation and analysis". In: *Scientific reports* 7.1 (2017), p. 16952.

[118] D. Nevozhay et al. "Mapping the Environmental Fitness Landscape of a Synthetic Gene Circuit". In: *PLoS Comput. Biol.* 8 (2012), e1002480. DOI: 10.1371/journal.pcbi.1002480.

[119] M. A. Nieto et al. "EMT: 2016". In: *Cell* 166.1 (June 2016), pp. 21–45. ISSN: 0092-8674. URL: http://www.sciencedirect.com/science/article/pii/S009286741630796

[120] M. Paola et al. "An end-to-end software solution for the analysis of high-throughput single-cell migration data". In: *Scientific Reports* 7.13 (2017), p. 42383.

[121] P. R. Paździorek. "Mathematical Model of Stem Cell Differentiation and Tissue Regeneration with Stochastic Noise". In: *Bull. Math. Biol.* 76.7 (2014), pp. 1642–1669. DOI: 10.1007/s11538-014-9971-5.

[122] Y. Pei et al. "An elementary introduction to Kalman filtering". In: *Commun. ACM* 62.11 (2019), pp. 1–12. DOI: 10.1145/3363294.

[123] D. Pellin et al. "Penalized inference of the hematopoietic cell differentiation network via high-dimensional clonal tracking". In: *Appl. Netw. Sci.* 4 (2019), p. 115. DOI: 10.1007/s41109-019-0225-1.

[124] F. Peruani, A. Deutsch, and M. Bär. "A mean-field theory for self-propelled particles interacting by velocity alignment mechanisms". In: *The European Physical Journal Special Topics* 157.1 (2008), pp. 111–122.

[125] K. Pham et al. "Density-dependent quiescence in glioma invasion: instability in a simple reaction–diffusion model for the migration/proliferation dichotomy". In: *J. Bio. Dyn.* June (June 2011), pp. 1–18. ISSN: 1751-3758. DOI: 10.1080/17513758.2011.590610. URL: http://www.tandfonline.com/doi/abs/10.1080/17513758.2011.590610.

[126] B. C. Prager et al. "Glioblastoma Stem Cells: Driving Resilience through Chaos". In: *Trends Cancer* 6.3 (2020), pp. 223–235. DOI: 10.1016/j.trecan.2020.01.009.

[127] S. Pressé et al. "Principles of maximum entropy and maximum caliber in statistical physics". In: *Reviews of Modern Physics* 85.3 (2013), p. 1115.

[128] T. D. Price, A. Qvarnström, and D. E. Irwin. "The role of phenotypic plasticity in driving genetic evolution". In: *Proc. R. Soc. London. Ser. B Biol. Sci.* 270.1523 (July 2003), pp. 1433–1440. ISSN: 0962-8452. DOI: 10.1098/rspb.2003.2372. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2003.2372.

[129] A. Puliafito et al. "Collective and single cell behavior in epithelial contact inhibition." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.3 (Jan. 2012), pp. 739–44. ISSN: 1091-6490. DOI: 10.1073/pnas.1007809109. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3271933%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[130] A. Rabani, G. Ariel, and A. Be'er. "Collective motion of spherical bacteria". In: *PloS one* 8.12 (2013), e83760.

[131] D. C. Radisky. "Epithelial-mesenchymal transition". In: *J. Cell Sci.* 118.19 (2005), pp. 4325–4326. ISSN: 0021-9533. DOI: 10.1242/jcs.02552. eprint: https://jcs.biologists.org/content/118/19/4325.full.pdf. URL: https://jcs.biologists.org/content/118/19/4325.

[132] W.-J. Rappel and L. Edelstein-Keshet. "Mechanisms of cell polarization". In: *Curr. Opin. Sys. Biol.* 3 (2017). Mathematical modelling Mathematical modelling, Dynamics of brain activity at the systems level Clinical and translational systems biology, pp. 43–53. ISSN: 2452-3100. DOI: https://doi.org/10.1016/j.coisb.2017.03.005. URL: http://www.sciencedirect.com/science/article/pii/S2452310016300166.

[133] A. Rebenshtok et al. "Non-normalizable densities in strong anomalous diffusion: Beyond the central limit theorem". In: *Physical review letters* 112.11 (2014), p. 110601.

[134] M. Robert and F. Schweitzer. "Multi-agent Model of Biological Swarming". In: *European Conference on Artificial Life* 2801 (2003), pp. 810–820.

[135] P. Romanczuk et al. "Active Brownian particles From individual to collective stochastic dynamics". In: *The European Physical Journal Special Topics* 202 (2012), pp. 1–162.

[136] Y. Sato et al. "Dynamic analysis of vascular morphogenesis using transgenic quail embryos". In: *PloS one* 5.9 (2010), e12674.

[137] M. Schienbein and H. Gruler. "Langevin equation, Fokker-Planck equation and cell migration". In: *Bulletin of Mathematical Biology* 55.3 (May 1993), pp. 585–608. ISSN: 1522-9602. DOI: 10.1007/BF02460652. URL: https://doi.org/10.1007/BF02460652.

[138] M. J. Schliekelman et al. "Molecular Portraits of Epithelial, Mesenchymal, and Hybrid States in Lung Adenocarcinoma and Their Relevance to Survival". In: *Cancer Research* 75.9 (May 2015), p. 1789. DOI: 10.1158/0008-5472.CAN-14-2535. URL: https://doi.org/10.1158/0008-5472.CAN-14-2535.

[139]   U. S. Schwarz and S. A. Safran. "Physics of adherent cells". In: *Rev. Mod. Phys.* 85 (3 Aug. 2013), pp. 1327–1381. DOI: 10.1103/RevModPhys.85.1327. URL: https://link.aps.org/doi/10.1103/RevModPhys.85.1327.

[140]   U. Seifert. "Stochastic thermodynamics, fluctuation theorems and molecular machines". In: *Rep. Prog. Phys.* 75.12 (2012), p. 126001. DOI: 10.1088/0034-4885/75/12/126001.

[141]   U. Seifert. "Stochastic thermodynamics: An introduction". In: *AIP Conference Proceedings* 1332 (2011), p. 56. DOI: 10.1063/1.3569487.

[142]   U. Seifert. "Stochastic thermodynamics: From principles to the cost of precision". In: *Physica A* 504 (2018), pp. 176–191. DOI: 10.1016/j.physa.2017.10.024.

[143]   D. Selmeczi, L. Li, and L. Pedersen. "Cell motility as random motion: A review". In: *The Eur. Phys. J S Top.* 157.15 (2008), pp. 1–15.

[144]   E. M. Sevick et al. "Fluctuation Theorems". In: *Ann. Rev. Phys. Chem.* 59 (2008), pp. 603–633. DOI: 10.1146/annurev.physchem.58.032806.104555.

[145]   A. Sigal et al. "Variability and memory of protein levels in human cells". In: *Nature* 444 (2006), pp. 643–646. DOI: 10.1038/nature05316.

[146]   H. A. Simon. *The new science of management decision.* en. New York: Harper & Brothers, Nov. 1960. URL: http://content.apa.org/books/13978-000.

[147]   J. M. W. Slack. "Metaplasia and transdifferentiation: from pure biology to the clinic". In: *Nat. Rev. Mol. Cell Biol.* 8.5 (2007), pp. 369–378. DOI: 10.1038/nrm2146.

[148]   R. Soldatov et al. "Spatiotemporal structure of cell fate decisions in murine neural crest". In: *Science* 364.6444 (2019). ISSN: 0036-8075. DOI: 10.1126/science.aas9536. eprint: https://science.sciencemag.org/content/364/6444/eaas9536.full.pdf. URL: https://science.sciencemag.org/content/364/6444/eaas9536.

[149]   V. Sourjik and H. C. Berg. "Receptor sensitivity in bacterial chemotaxis". In: *Proceedings of the National Academy of Sciences* 99.1 (2002), pp. 123–127. ISSN: 0027-8424. DOI: 10.1073/pnas.011589998. eprint: https://www.pnas.org/content/99/1/123.full.pdf. URL: https://www.pnas.org/content/99/1/123.

[150]   T. Speck and U. Seifert. "The Jarzynski relation, fluctuation theorems, and stochastic thermodynamics for non-Markovian processes". In: *J. Stat. Mech.* 2007 (2007), p. L09002. DOI: 10.1088/1742-5468/2007/09/l09002.

[151]   S. Srinivasan et al. "Ligand Binding Mechanisms in Human Cone Visual Pigments". In: *Trends Biochem. Sci.* 44.7 (2019), pp. 629–639. DOI: 10.1016/j.tibs.2019.02.001.

[152]   N. Srivastava et al. "Pressure sensing through Piezo channels controls whether cells migrate with blebs or pseudopods". In: *PNAS* 117.5 (2020), pp. 2506–2512. DOI: 10.1073/pnas.1905730117.

[153]   Steager et al. "Dynamics of pattern formation in bacterial swarms". In: *Physics of Fluids* 20.7 (2008), p. 073601.

[154] M. L. Suvà et al. "Reconstructing and Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-Like Cells". In: *Cell* 157.3 (2014), pp. 580–594. DOI: 10.1016/j.cell.2014.02.030.

[155] K. Takahashi and S. Yamanaka. "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors". In: *Cell* 126.4 (2006), pp. 663–676. DOI: 10.1016/j.cell.2006.07.024.

[156] J. P. Thiery. "Epithelial–mesenchymal transitions in development and pathologies". In: *C. Opi. Cell Bio.* 15.6 (2003), pp. 740–746. ISSN: 0955-0674. DOI: https://doi.org/10.1016/j.ceb.2003.10.006. URL: http://www.sciencedirect.com/science/article/pii/S0955067403001339.

[157] S. Tripathi, H. Levine, and M. K. Jolly. "The Physics of Cellular Decision Making During Epithelial-Mesenchymal Transition". In: *Annual Review of Biophysics* 49.1 (May 2020), pp. 1–18. ISSN: 1936-122X. DOI: 10.1146/annurev-biophys-121219-081557. URL: https://doi.org/10.1146/annurev-biophys-121219-081557.

[158] S. Uda. "Application of information theory in systems biology". In: *Biophys. Rev.* 12 (2020), pp. 377–384. DOI: 10.1007/s12551-020-00665-w.

[159] M. Ueda and T. Shibata. "Stochastic Signal Processing and Transduction in Chemotactic Response of Eukaryotic Cells". In: *Biophys, J.* 93.1 (July 2007), pp. 11–20. ISSN: 0006-3495. DOI: 10.1529/biophysj.106.100263. URL: https://doi.org/10.1529/biophysj.106.100263.

[160] C. Van den Broeck. "Stochastic thermodynamics: A brief introduction". In: *Proceedings of the International School of Physics "Enrico Fermi"* 184 (2013), pp. 155–193. DOI: 10.3254/978-1-61499-278-3-155.

[161] C. Van den Broeck and M. Esposito. "Ensemble and trajectory thermodynamics: A brief introduction". In: *Physica A* 418 (2015), pp. 6–16. DOI: 10.1016/j.physa.2014.04.035.

[162] T. Vicsek et al. "Novel type of phase transition in a system of self-driven particles". In: *Phys. rev. lett.* 75.6 (1995), p. 1226.

[163] J. West et al. "Normal tissue architecture determines the evolutionary course of cancer". In: *Nat. Commun.* 12 (2021), pp. 1–9. DOI: 10.1038/s41467-021-22123-1.

[164] P. R. ten Wolde et al. "Fundamental Limits to Cellular Sensing". In: *Journal of Statistical Physics* 162.5 (Mar. 2016), pp. 1395–1424. ISSN: 1572-9613. DOI: 10.1007/s10955-015-1440-5. URL: https://doi.org/10.1007/s10955-015-1440-5.

[165] L. Xu and J. Wang. "Chapter 7: Quantifying Waddington landscapes, paths, and kinetics of cell fate decision making of differentiation/development". In: *Phenotypic Switching*. Academic Press, 2020, pp. 157–187.

[166] X. L. Xu et al. "Rb suppresses human cone-precursor-derived retinoblastoma tumours". In: *Nature* 514.7522 (2014), pp. 385–388. DOI: 10.1038/nature13813.

[167] L. Yan and D. Bi. "Multicellular Rosettes Drive Fluid-solid Transition in Epithelial Tissues". In: *Phys. Rev. X* 9 (1 2019), p. 011029. DOI: 10.1103/PhysRevX.9.011029. URL: https://link.aps.org/doi/10.1103/PhysRevX.9.011029.

[168]   Y. Yao and C. Wang. "Dedifferentiation: inspiration for devising engineering strategies for regenerative medicine". In: *npj Regen. Med.* 5.14 (2020), pp. 1–11. DOI: 10.1038/s41536-020-00099-8.

[169]   M. Zhao et al. "dbEMT: an epithelial-mesenchymal transition associated gene resource". In: *Scientific Reports* 5.1 (June 2015), p. 11459. ISSN: 2045-2322. DOI: 10.1038/srep11459. URL: https://doi.org/10.1038/srep11459.

[170]   W. Zi et al. "Deep reinforcement learning of cell movement in the early stage of C.elegans embryogenesis". In: *Bioinformatics* 34.9 (2018), pp. 3169–3177.

7. Supplementary Material

Symbol	Explanation
X_i	Phenotype of i-th cell
Y_i	Phenotype of i-th neighbthehood cells
N_i^0	Number of cells in i-th cell's microenvironment having phenotype $(X_i = 0)$
N_i^1	Number of cells in i-th cell's microenvironment having phenotype $(X_i = 1)$
N_i^ϕ	Number of free slots in i-th cell's microenvironment
V	Total capacity
N_T	Total number of phenotypes
ℓ	Radius of the microenvironment
ν	Exchange rate
r	Growth rate
β	LEUP sensitivity
$\rho_0(x)$	Mean density of the resting cells at position x
$\rho_1(x)$	Mean density of the migratory cells at position x
D_0	Diffusion coefficient of resting cells
D_1	Diffusion coefficient of migratory cells

S.1. Calculation of microenvironmental entropy

Let assume that a maximum number of N cells is present inside the cell's microenvironment, where ℓ is the radius of the microenvironment (Fig.3.1i). Since there is the possibility of free slots, if there are less than N cells in the cell neighbthehood, have started by assuming a trinomial distribution for the i-th cell is $P\left(N_i^1, N_i^\phi\right)$ where the number N_i^0 of cells having phenotype $X_i = 0$ and by the number N_i^1 of cells having phenotype $X_i = 1$. In addition, a number N_i^ϕ of empty slots are included inside the

microenvironment. The joint probability is defined by

$$P\left(N_i^1, N_i^\phi\right) = \frac{N!}{N_i^1! N_i^\phi! \left(N - N_i^\phi - N_i^1\right)!} p^{N_i^1} \theta^{N_i^\phi} (1 - p - \theta)^{\left(N - N_i^\phi - N_i^1\right)}, \tag{7.1}$$

where p and θ are the probabilities of having a number N_i^1 of cells with phenotype $(X_i = 1)$ out of N cells and having a number N_i^ϕ of free slots. The conditional probability of having a number N_i^1 of cells present in the microenvironment given a number N_i^ϕ of free slots is

$$
\begin{aligned}
P\left(N_i^1 \mid N_i^\phi\right) &= \frac{P\left(N_i^1, N_i^\phi\right)}{P\left(N_i^\phi\right)} \\
&= \frac{\frac{N!}{N_i^1! N_i^\phi! \left(N - N_i^\phi - N_i^1\right)!} p^{N_i^1} \theta^{N_i^\phi} (1 - p - \theta)^{\left(N - N_i^\phi - N_i^1\right)}}{\frac{N!}{N_i^\phi! \left(N - N_i^\phi\right)!} \theta^{N_i^\phi} (1 - \theta)^{\left(N - N_i^\phi\right)}} \\
&= \binom{N - N_i^\phi}{N_i^1} \left(\frac{p}{1 - \theta}\right)^{N_i^1} \left(1 - \frac{p}{1 - \theta}\right)^{N - N_i^1 - N_i^\phi} \\
&= \boldsymbol{B}\left(N - N_i^\phi, \frac{p}{1 - \theta}\right) \\
&= \boldsymbol{B}\left(N_T, \frac{p}{1 - \theta}\right)
\end{aligned}
\tag{7.2}
$$

Now, one can use the form of binomial distribution to calculate the entropy of the Binomial distribution. Please note that $P(N_i^1 \mid N_i^\phi)$ can be written as $P(N_i^1 \mid N_T)$ due to Binomial distribution.

$$S\left(N_i^1 \mid N_T\right) = S\left(Y_i = [N_i^1 \mid N_T]\right) = \frac{1}{2} \log_2\left(2\pi e \frac{N_T p}{1 - \theta}\left(1 - \frac{p}{1 - \theta}\right)\right), \tag{7.3}$$

where

$$
\frac{p}{1 - \theta} = \frac{N_i^1}{N} \frac{1}{1 - \frac{N_i^\phi}{N}} = \frac{N_i^1}{N - N_i^\phi} = \frac{N_i^1}{N_T},
$$
$$
1 - \frac{p}{1 - \theta} = \frac{N_i^0}{N - N_i^\phi} = \frac{N_i^0}{N_T}.
\tag{7.4}
$$

According to LEUP have to evaluate the microenvironmental entropy of $S\left(Y_i \mid X_i = 0\right)$ and $S\left(Y_i \mid X_i = 1\right)$ to calculate the probability of the internal states $(X_i = 0)$ and $(X_i = 1)$

$$
S\left(Y_i = [N_i^1 \mid N_T] \mid X_i = 1\right) = \frac{1}{2} \log_2\left(2\pi e \left(\frac{(N_T - 1)(N_i^1 - 1)}{(N_T - 1)}\right)\left(\frac{N_i^0}{(N_T - 1)}\right)\right),
$$
$$
S\left(Y_i = [N_i^1 \mid N_T] \mid X_i = 0\right) = \frac{1}{2} \log_2\left(2\pi e \left(\frac{(N_T - 1)(N_i^1)}{(N_T - 1)}\right)\left(\frac{N_i^0 - 1}{(N_T - 1)}\right)\right).
\tag{7.5}
$$

From the Gaussian approximation can write the entropy difference as

$$
\begin{aligned}
\Delta S &= S(Y_i = [N_i^1 \mid N_T] \mid X_i = 1) - S(Y_i = [N_i^1 \mid N_T] \mid X_i = 0), \\
&= \frac{1}{2} \ln\left[\frac{N_i^0 (N_i^1 - 1)}{N_i^1 (N_i^0 - 1)}\right].
\end{aligned}
\tag{7.6}
$$

S.2. Calculation of cell proliferation rate

In the Go-or-Grow model, only resting cells are allowed to proliferate but not the moving cells. The growth rate depends in a specific way on the number of moving and resting cells in the microenvironment. In particular, assume that the per-capita growth rate for resting cells is a linearly decreasing function of N_i^0 and N_i^1, and is also decreasing with the number of migratory cells and a constant per-capita death rate d_1. Accordingly,

$$
\begin{aligned}
\frac{d\langle N_i^0 \rangle}{dt} &= \left(\langle W^+ \rangle_{N_i^0, N_i^1} - \langle W^- \rangle_{N_i^0, N_i^1} \right) \\
&= h_1 \rho_0 - q \left(\rho_0 + \rho_1 \right) \rho_0 - d_1 \rho_0 \\
&= r \rho_0 \left(1 - \rho_0 - \rho_1 \right)
\end{aligned}
\tag{7.7}
$$

where $\frac{q}{r} = 1$ and $r = h_1 - d_1$.

S.4. Phase space diagram for null sensitivity case
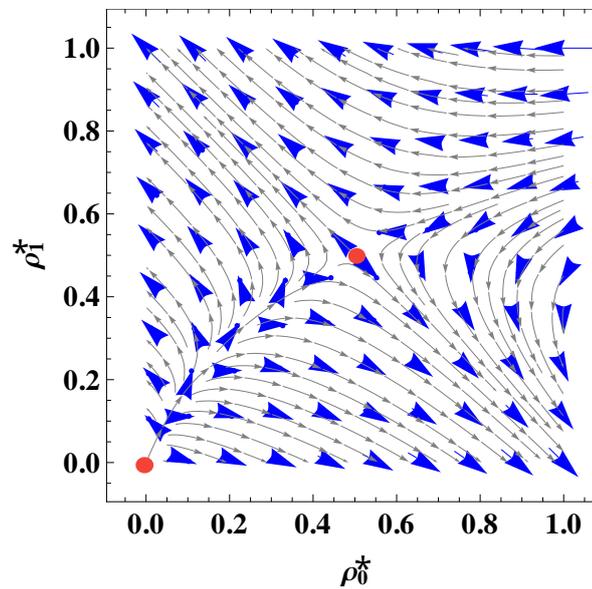
Figure 7.1.. Phase space diagram of ρ_0^* and ρ_1^* for, $r = 1$ where two fixed points are marked by red circles.

S.5. Turing space for pattern formation
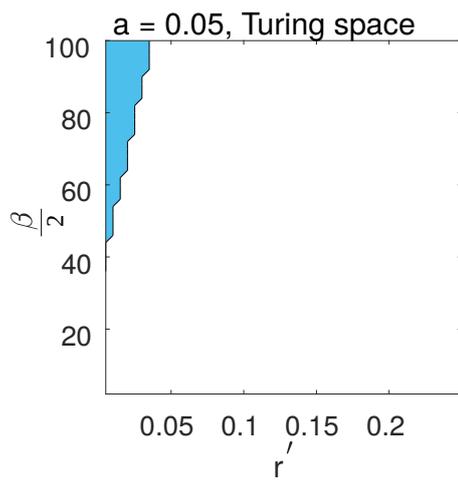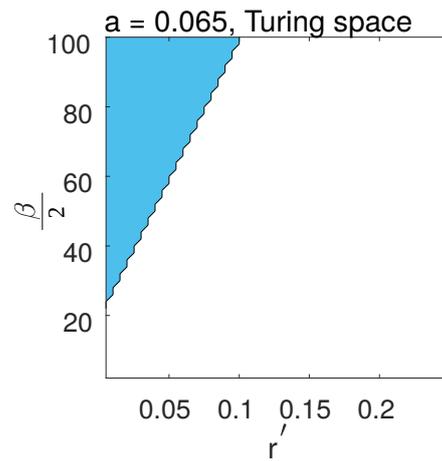
(i) a is fixed at 0.05.
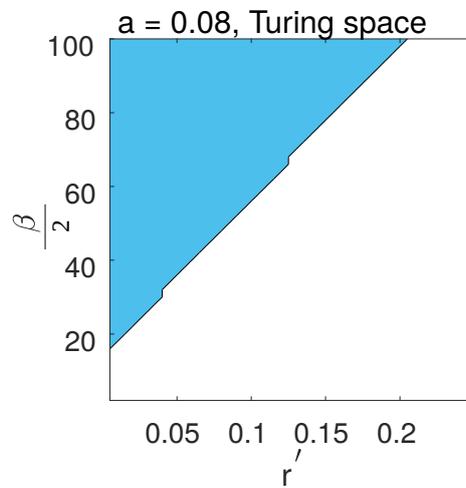
(ii) a is fixed at 0.065.

(iii) a is fixed at 0.08.

Figure 7.2.. Phase space diagram of β vs. r' for different values of a (**a-c**), where the Turing space is marked in blue. a is defined by $\frac{1}{V}$.

S.6. Critical wavelengths inside the Turing space for pattern formation

(i) a is fixed at 0.05.

(ii) a is fixed at 0.065.

(iii) a is fixed at 0.08.
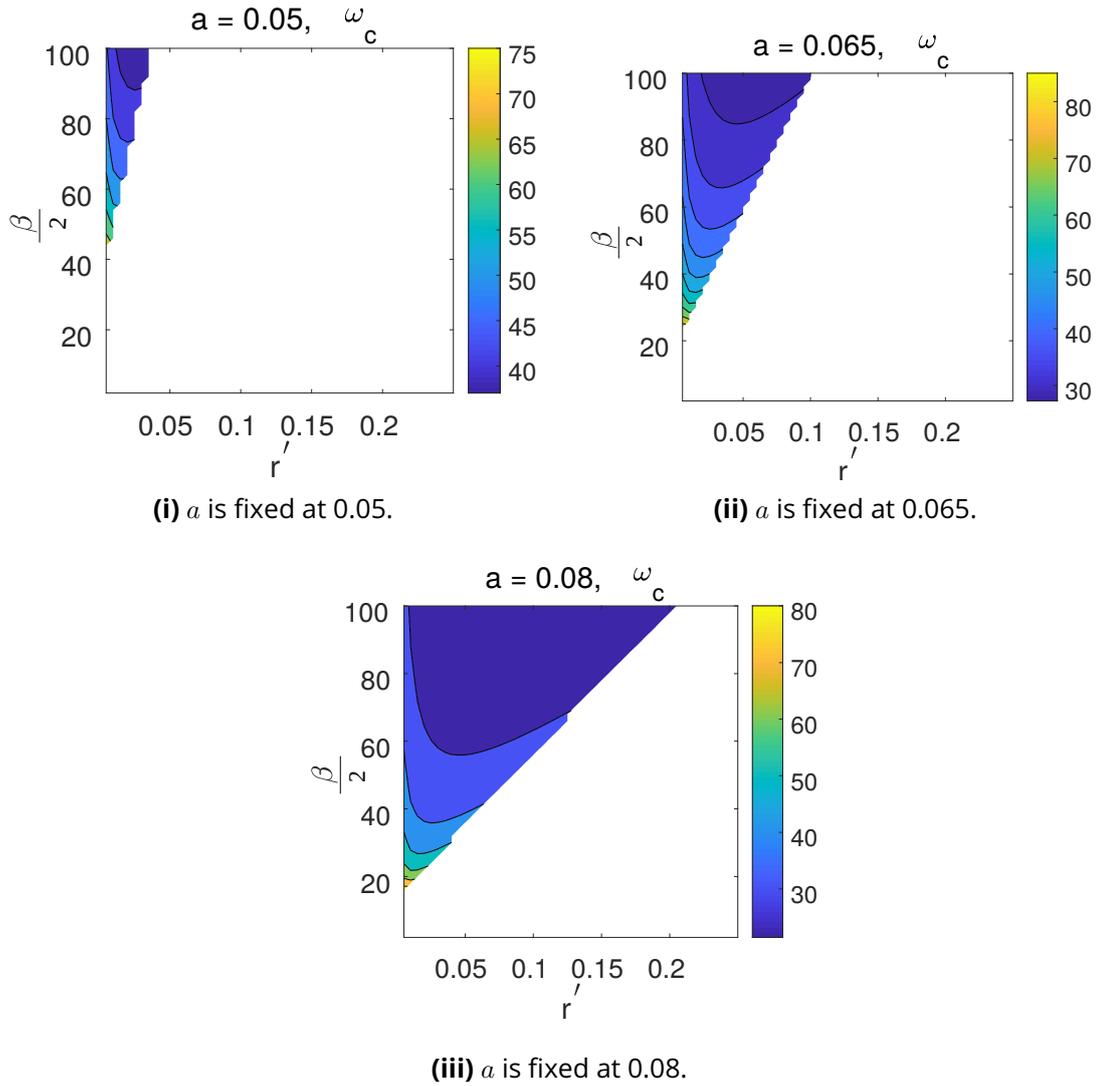
Figure 7.3.. Phase space diagram of β vs. r' for different values of a (**a-c**), where critical wavelengths (ω_c) have been plotted inside the Turing space. a is defined by $\frac{1}{V}$.

S.7. Plot of resting probability
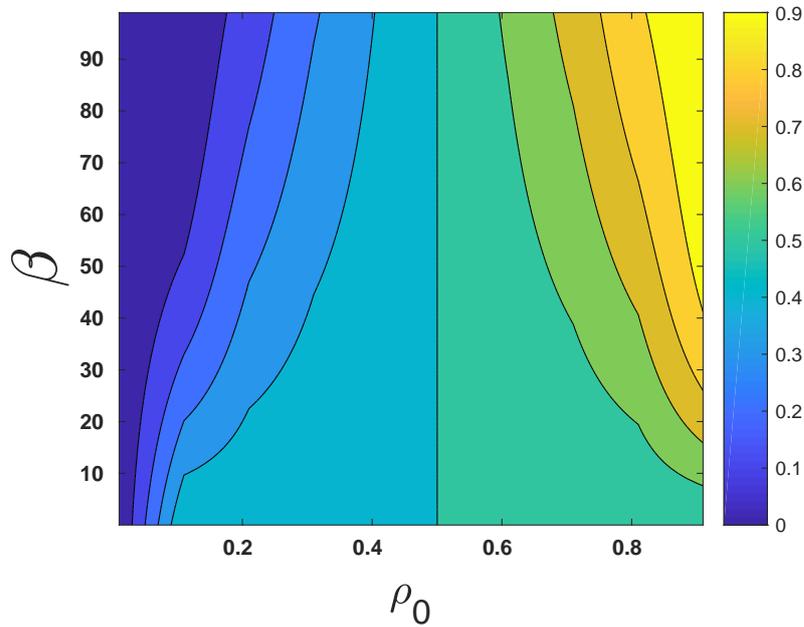
Figure 7.4.. Plot of resting probability with respect to β and ρ_0. have fixed the value of $\frac{1}{V}$ to 0.01.

S.8. Critical sensitivity vs. inverse capacity graph from IBM
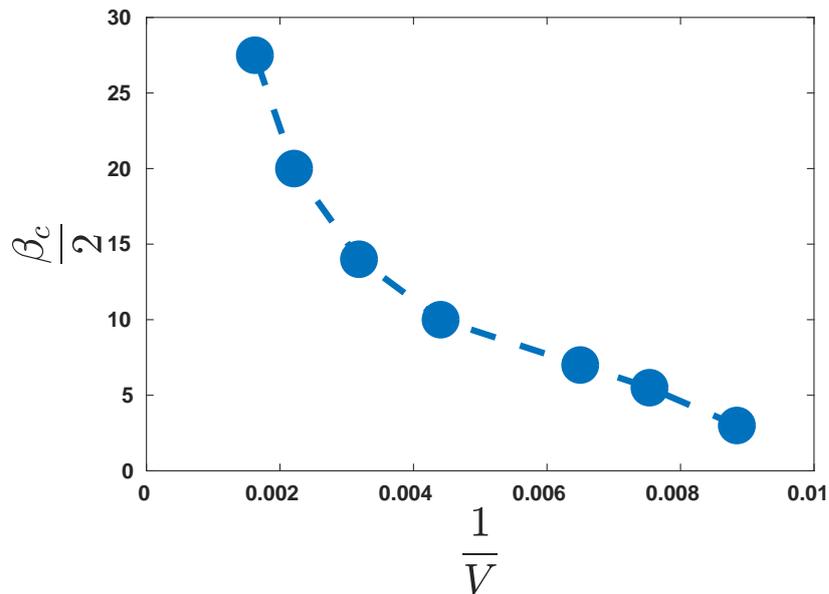
Figure 7.5.. Critical sensitivity with respect to the inverse capacity of the IBM over 5 simulations. Throughout the simulations, kept the total density at 0.2 and the total number of cells was 500.

S.9. Calculation of the curvature of microenvironmental entropy and Helmholtz free energy

The entropy of the wrapped Cauchy distribution is

$$S\left(\Theta_i \mid \theta_i\right) = \ln\left(2\pi\right)$$

$$+ \ln\left\{1 - \frac{1}{N_{C_{R,n}}^2}\left[\bar{v}_{y,n}^2 + \bar{v}_{x,n}^2 + 2\left(\bar{v}_{y,n}\sin\theta_i + \bar{v}_{x,n}\cos\theta_i\right)\right]\right\} \tag{7.8}$$

.

To discriminate the maximum from the minimum point, the value of $\frac{\partial^2}{\partial\theta_i^2}S\left(\Theta_i \mid \theta_i\right)$ must be evaluated at each extremum point. The second derivative is given by

$$\frac{\partial^2}{\partial\theta_i^2}S\left(\Theta_i \mid \theta_i\right) = \frac{\frac{2}{N_{C_{R,i}}^2}\left(\bar{v}_{y,i}\sin\theta_i + \bar{v}_{x,i}\cos\theta_i\right)}{1 - \frac{1}{N_{C_{R,i}}^2}\left[\bar{v}_{y,i}^2 + \bar{v}_{x,i}^2 + 2\left(\bar{v}_{y,i}\sin\theta_i + \bar{v}_{x,i}\cos\theta_i\right)\right]}$$

$$- \left\{\frac{\frac{2}{N_{C_{R,i}}^2}\left(\bar{v}_{y,i}\cos\theta_i - \bar{v}_{x,i}\sin\theta_i\right)}{1 - \frac{1}{N_{C_{R,i}}^2}\left[\bar{v}_{y,i}^2 + \bar{v}_{x,i}^2 + 2\left(\bar{v}_{y,i}\sin\theta_i + \bar{v}_{x,i}\cos\theta_i\right)\right]}\right\}^2. \tag{7.9}$$

In an extremum point, $\frac{\partial}{\partial\theta_i}S\left(\Theta_i \mid \theta_i\right) = 0$, therefore

$$\left.\frac{\partial^2}{\partial\theta_i^2}S\left(\Theta_i \mid \theta_i\right)\right|_{\text{ext}} = \frac{\frac{2}{N_{C_{R,i}}^2}\left(\bar{v}_{y,i}\sin\theta_i + \bar{v}_{x,i}\cos\theta_i\right)}{1 - \frac{1}{N_{C_{R,i}}^2}\left[\bar{v}_{y,i}^2 + \bar{v}_{x,i}^2 + 2\left(\bar{v}_{y,i}\sin\theta_i + \bar{v}_{x,i}\cos\theta_i\right)\right]}. \tag{7.10}$$

defining κ as the proportionality constant relating $\sin\theta_i$ and $\cos\theta_i$ with $\bar{v}_{y,i}$ and $\bar{v}_{x,i}$, respectively, the second derivative evaluated at $\theta_i = \bar{\theta}$ is

$$\left.\frac{\partial^2}{\partial\theta_i^2}S\left(\Theta_i \mid \theta_i\right)\right|_{\theta_i=\bar{\theta}} = \frac{\frac{2\kappa}{N_{C_{R,i}}^2}\left(\bar{v}_{y,i}^2 + \bar{v}_{x,i}^2\right)}{1 - \left(S_{C_{R,i}}^1\right)^2} > 0, \tag{7.11}$$

because the numerator is positive definite, and the denominator is positive given the bounds of the order parameters. The extremum point $\theta_i = \bar{\theta}$ therefore corresponds to an entropy minimum. Consequently, the behavior of the regime $\beta < 0$ is analogous to that of the Vicsek model. Conversely, at $\theta_i = \bar{\theta} + \pi$ find that

$$\left.\frac{\partial^2}{\partial\theta_i^2}S\left(\Theta_i \mid \theta_i\right)\right|_{\theta_i=\bar{\theta}} = \frac{\frac{-2\kappa}{N_{C_{R,i}}^2}\left(\bar{v}_{y,i}^2 + \bar{v}_{x,i}^2\right)}{1 - \left(S_{C_{R,i}}^1\right)^2} < 0, \tag{7.12}$$

using the same arguments as for the $\theta_i = \bar{\theta}$ point. Therefore, the point $\theta_i = \bar{\theta} + \pi$ corresponds to the entropy maximum. Then, the regime $\beta > 0$ corresponds to a nematic analog of the Vicsek model. Next, one can assume that the model has

a steady state, where the Helmholtz free energy per bacterium is given by $F = -\frac{1}{\beta_\theta} \ln Z$. Due to its extensivity, the Helmholtz free energy of the complete system is

$$F_T = -\frac{1}{\beta_\theta} \sum_{i=1}^{N} \ln Z_i = -\frac{1}{\beta_\theta} \ln \left(\prod_{i=1}^{N} Z_n \right),$$

where Z_n is the normalization constant of i-th cell(see Eq.(3) in chapter 2).

The effective normalization constant $Z_T := \prod_{i=1}^{N} Z_i$ is given by

$$Z_T = \int e^{-\beta_\theta \sum_{i=1}^{N} \left[\ln(2\pi) + \ln \left(1 - e^{-2\gamma_i} \right) \right]} \mathrm{d}\vartheta_i. \tag{7.13}$$

The integration is performed over the orientations of all cells in the system. Moreover, the dependency of each γ_i on all angles θ_i is complex and makes integration challenging. However, variation of θ_i for all i translates into a variation in all γ_i. Therefore, Eq. 7.13 is equivalent to

$$Z_T = \int e^{-\beta_\theta \sum_{i=1}^{N} \left[\ln(2\pi) + \ln \left(1 - e^{-2\gamma_i} \right) \right]} \mathrm{d}\gamma_i. \tag{7.14}$$

expanding up to linear terms around $\gamma_i = 0$ yields

$$Z_T = \int e^{-\beta_\theta \sum_{i=1}^{N} \left[\ln(2\pi) + \ln(2\gamma_i) \right]} \mathrm{d}\gamma_i,$$

which after rearranging terms and integrating reduces to

$$Z_T = \left[\frac{1}{(4\pi)^{\beta_\theta}} \frac{\gamma_i^{1-\beta_\theta}}{1 - \beta_\theta} \right]^{N}. \tag{7.15}$$

Substituting Eq. 7.15 into the expression of the Helmholtz free energy (Eq. (6) in the main text), and rearranging terms, yields the Helmholtz free energy

$$F = N \left[\left(1 - \frac{1}{\beta_\theta} \right) \ln \left(\gamma_i \right) + \ln(4\pi) + \frac{\ln \left(1 - \beta_\theta \right)}{\beta_\theta} \right]. \tag{7.16}$$

Eq. 7.16 is well-defined only for $\beta_\theta < 1$. This indicates that no steady state exists for $\beta_\theta \geq 1$, hinting at an out-of-equilibrium regime. The present model belongs to the class of models with logarithmic potentials.
The existence of a non-normalizable state in certain parameter regimes is a staple of systems with logarithmic potentials.

# S.10. Pattern formation in different β_θ regime (see Table 1) ($\beta_v = 0$)
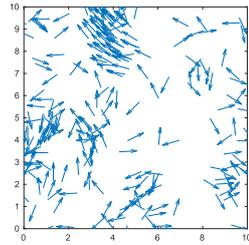
(i) Polar aligned street cells (P=0.6 to 0.7, N=0.5 to 0.7, V=0 to 0.04)
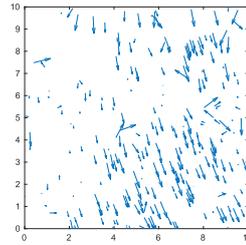
(ii) Compact polar aligned cells (P=0.8 to 1.0, N=0.7 to 1.0, V=0 to 0.065)
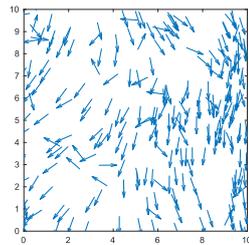
(iii) Scattered polar aligned cells (P=0.7 to 1.0, N=0.7 to 1.0, V=0 to 0.05)
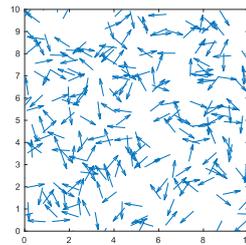
(iv) No order or patterns (P=0 to 0.09, N=0 to 0.07, V=0 to 0.065)
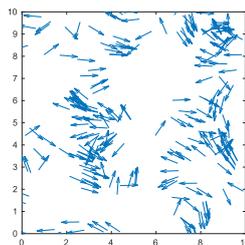
(v) Nematic streaming (P=0 to 0.05, N=0.3 to 0.5, V=0 to 0.065)
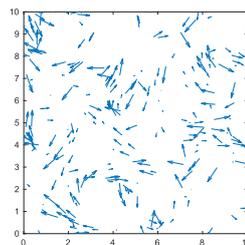
(vi) Nematic streaming and vorticules (P=0 to 0.03, N=0.2 to 0.4, V=0 to 0.04)
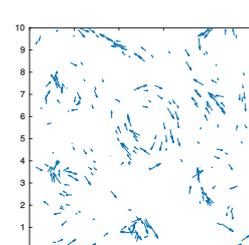
(vii) Vortices (P=0 to 0.03, N=0.2 to 0.4, V=0.1 to 0.35)

Figure 7.6.. The range of the polar order parameter is defined as P, nematic order parameter as N and mean absolute vorticity as V. All type of patterns are captured for different β_θ values. Patterns have been changed due to velocity distributions and interaction radius.

## S.11. Polar order parameter in angular sensitivity $(\beta_\theta < 0)$ regime
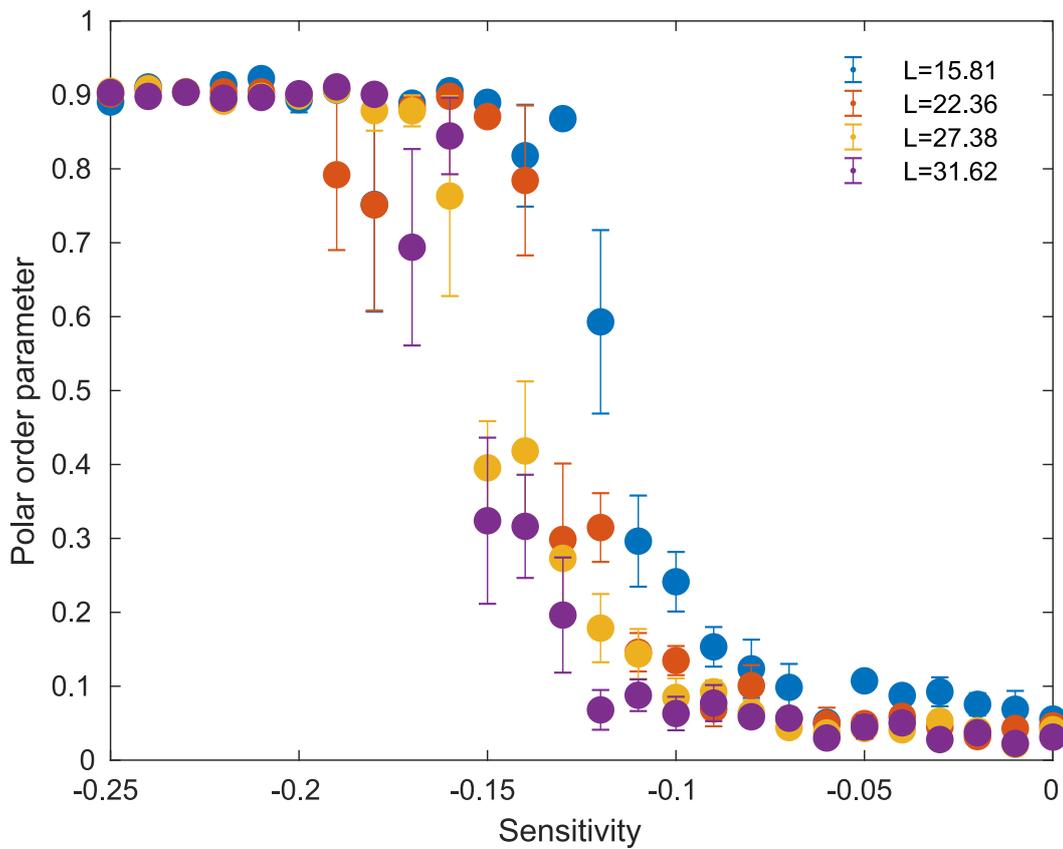
Figure 7.7.. Polar order parameter vs. angular sensitivity graph. Where interaction radius is at 3, and standard deviation of noise is at 0.1. Density is fixed at 1.0. Here, microenvironmental entropy has been taken from wrapped Cauchy distribution. Here $g = 1, \beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. All the order parameters re averaged over 5 realizations after 10^3 time steps.

S.12. Polar order parameter vs. angular noise graph
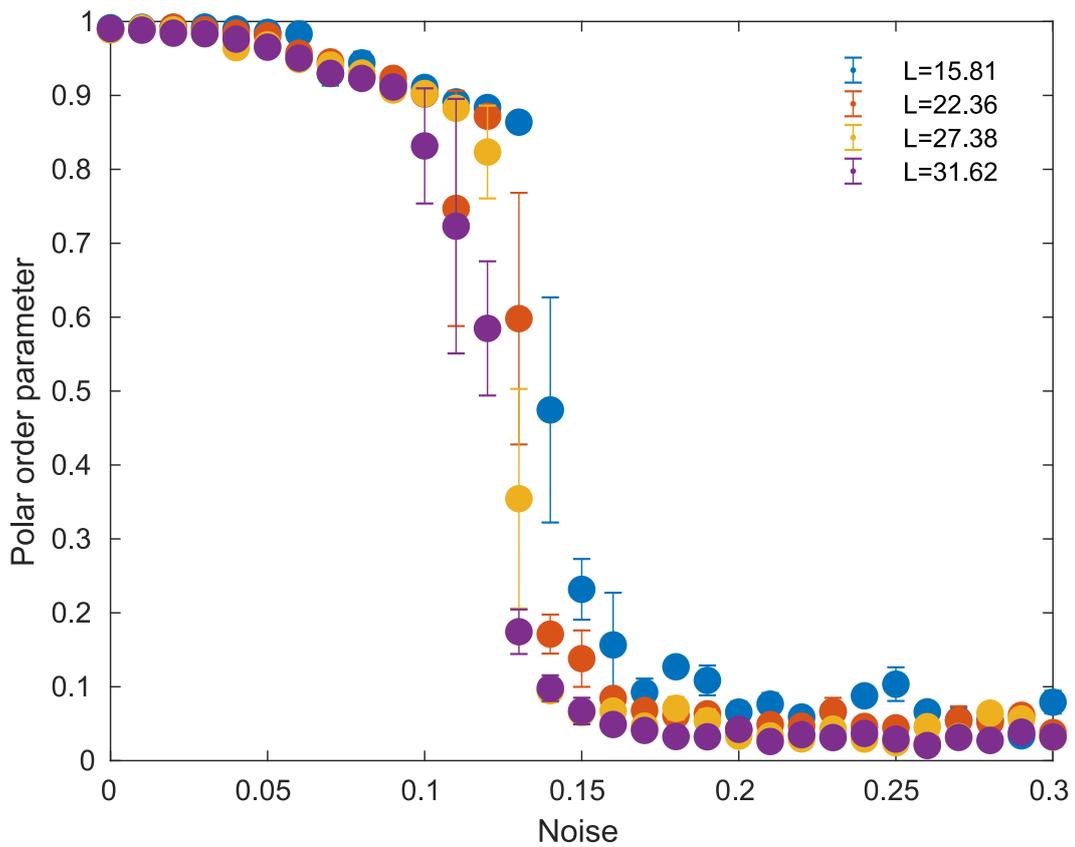
Figure 7.8.. Polar order parameter vs. angular noise graph. Where interaction radius is at 3 and angular sensitivity is at -0.2. Density is fixed at 1.0. Here, microenvironmental entropy has been taken from wrapped Cauchy distribution. Here $g = 1$, $\beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. All the order parameters re averaged over 5 realizations after 10^3 time steps.

S.13. Nematic order parameter in angular sensitivity $(\beta_\theta > 0)$ regime
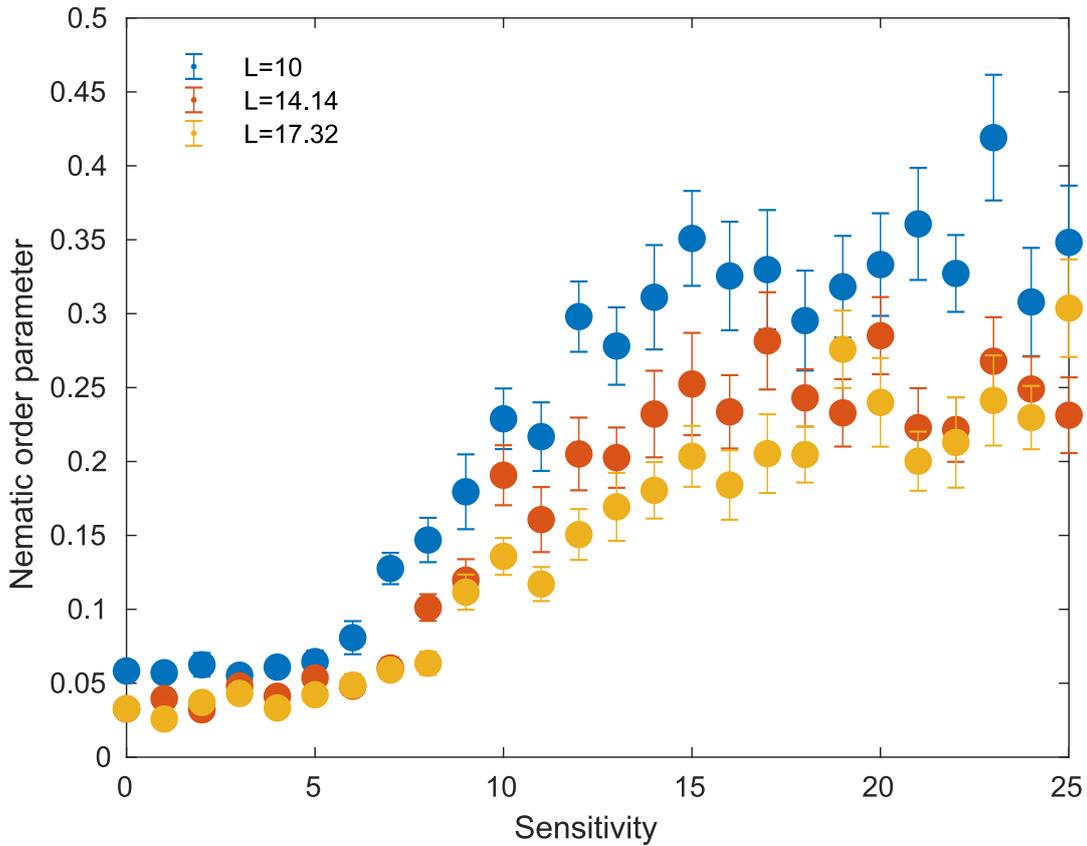
Figure 7.9.. Nematic order parameter vs. angular sensitivity graph. Where interaction radius is at 3 and standard deviation of angular noise is at 0.05. Density is fixed at 2.5. Here, microenvironmental entropy has been taken from wrapped Cauchy distribution. Here $g = 1, \beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. All the order parameters re averaged over 20 realizations after 10^3 time steps.

S.14. Nematic order parameter vs. angular noise graph
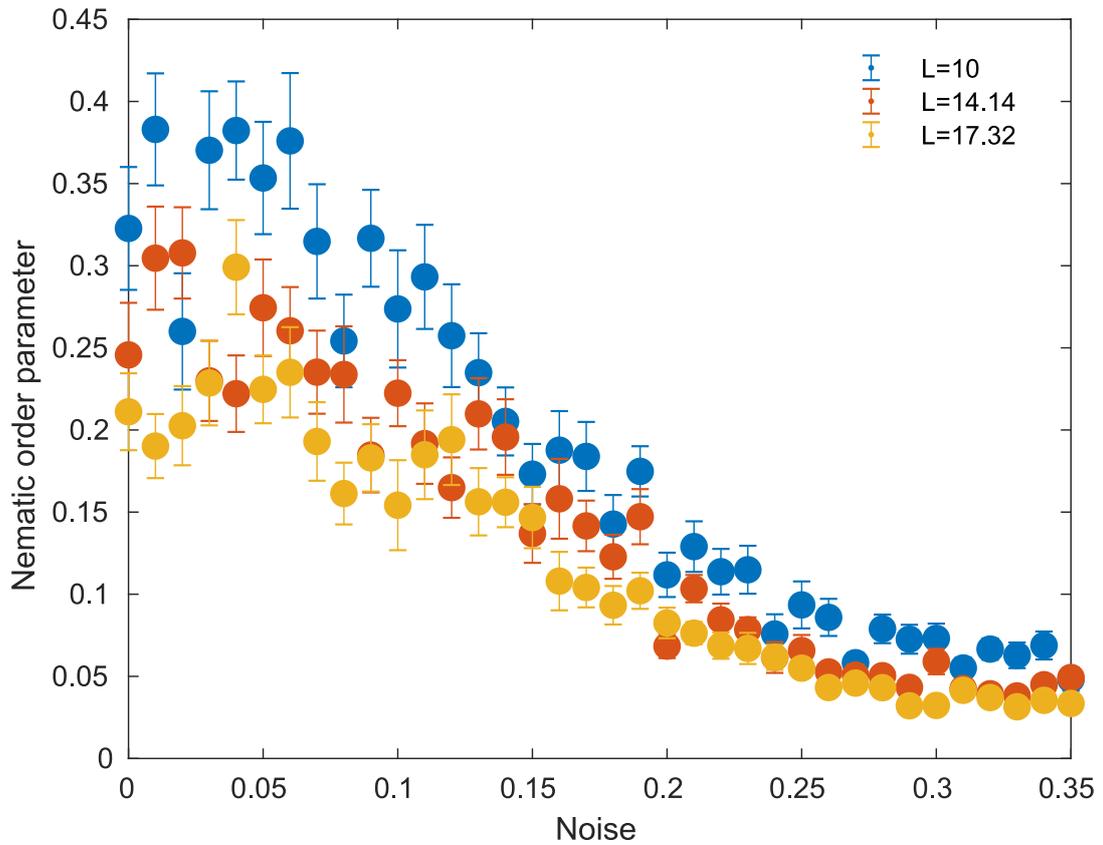
Figure 7.10.. Nematic order parameter vs. angular noise graph. Where interaction radius is at 3 and sensitivity is at 20.Density is fixed at 2.5. Here, microenvironmental entropy has been taken from wrapped Cauchy distribution. Here $g = 1, \beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. All the order parameters re averaged over 20 realizations after 10^3 time steps.

S.15. Order parameters vs. density graph
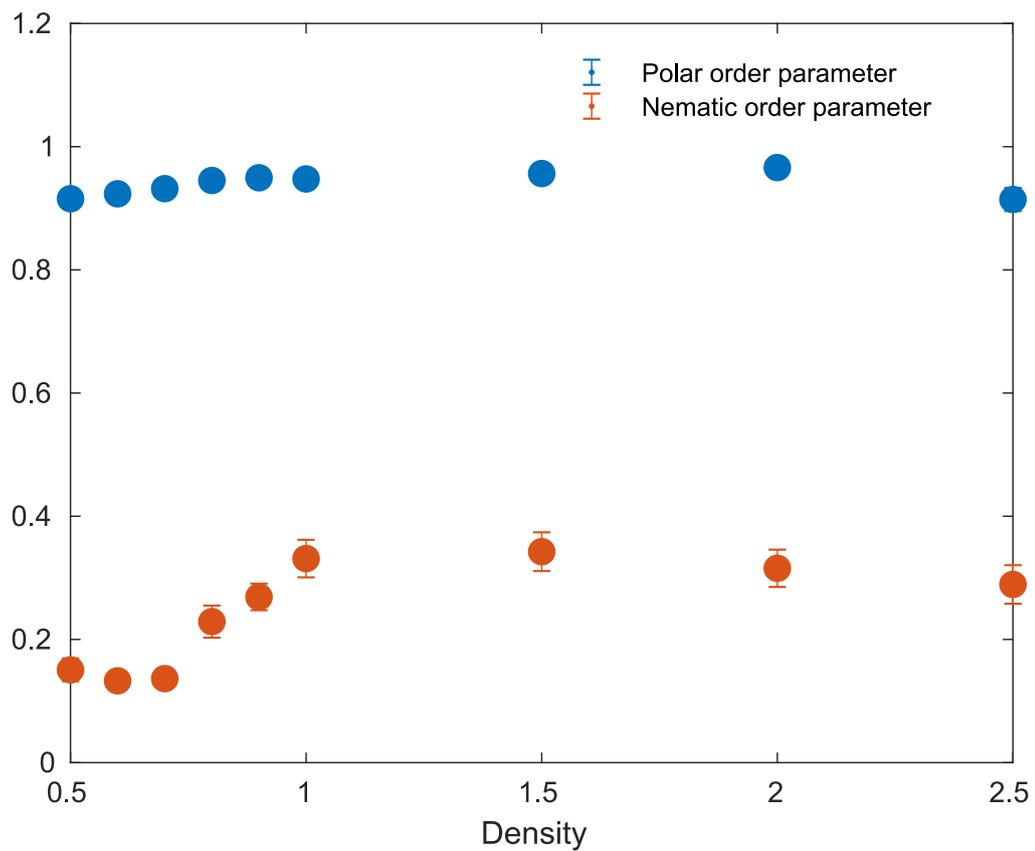
Figure 7.11.. Polar order parameter and nematic order parameter vs. density graph. Where interaction radius is at 3 and sensitivity is fixed at 15(for nematic order parameter) and -0.2 (for polar order parameter). Here, microenvironmental entropy has been taken from wrapped Cauchy distribution. Here $g = 1$, $\beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. All the order parameters re averaged over 20 realizations after 10^3 time steps.

S.16. Polar order parameter and nematic order parameter in angular sensitivity ($\beta_\theta < 0$) regime

One can take microenvironmental entropy from a wrapped exponential distribution, any qualitative change can not be observed in phase transition phenomena in $\beta_\theta < 0$ the regime.
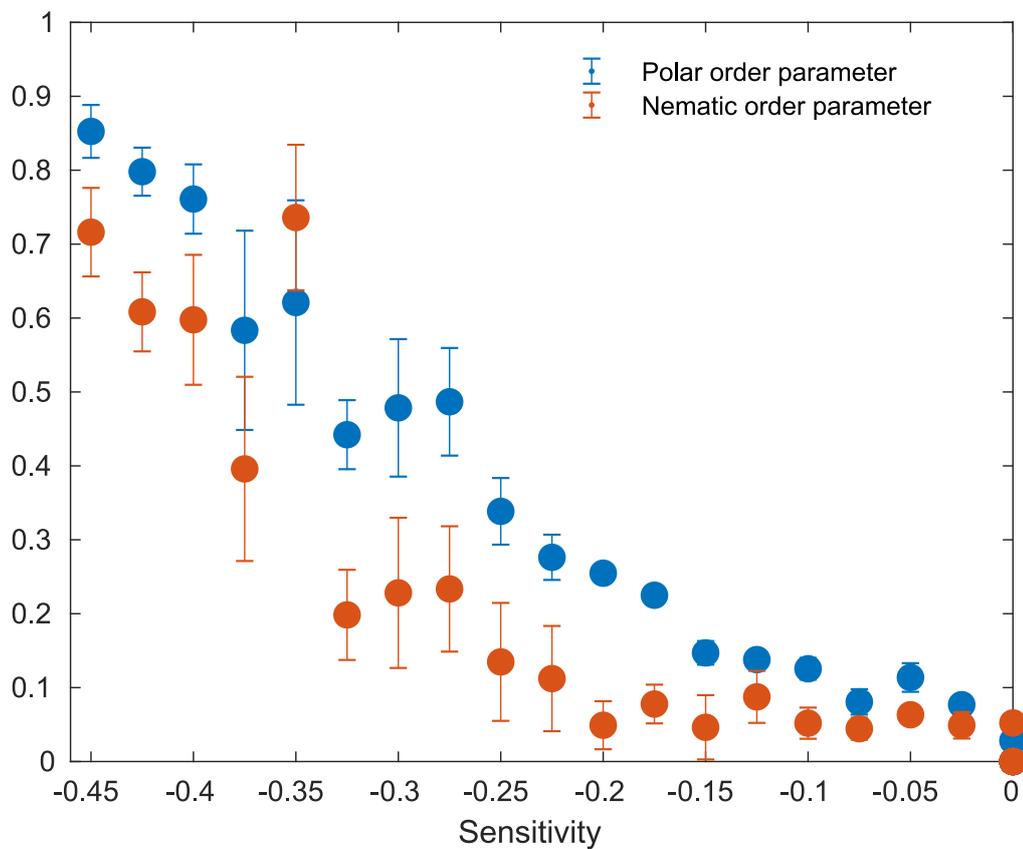

Figure 7.12.. Polar order parameter and nematic order parameter vs. angular sensitivity ($\beta_\theta < 0$) graph. Where interaction radius is at 3, and standard deviation of noise is at 0.05. Density is fixed at 2.5. Here, microenvironmental entropy has been taken from a wrapped exponential distribution. Here $g = 1$, $\beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. All the order parameters re averaged over 5 realizations after 250 time steps.

S.17. Polar order parameter and nematic order parameter in angular sensitivity ($\beta_\theta > 0$) regime

One can take microenvironmental entropy from a wrapped exponential distribution, any qualitative change can not be observed in phase transition phenomena in $\beta_\theta > 0$ regime.
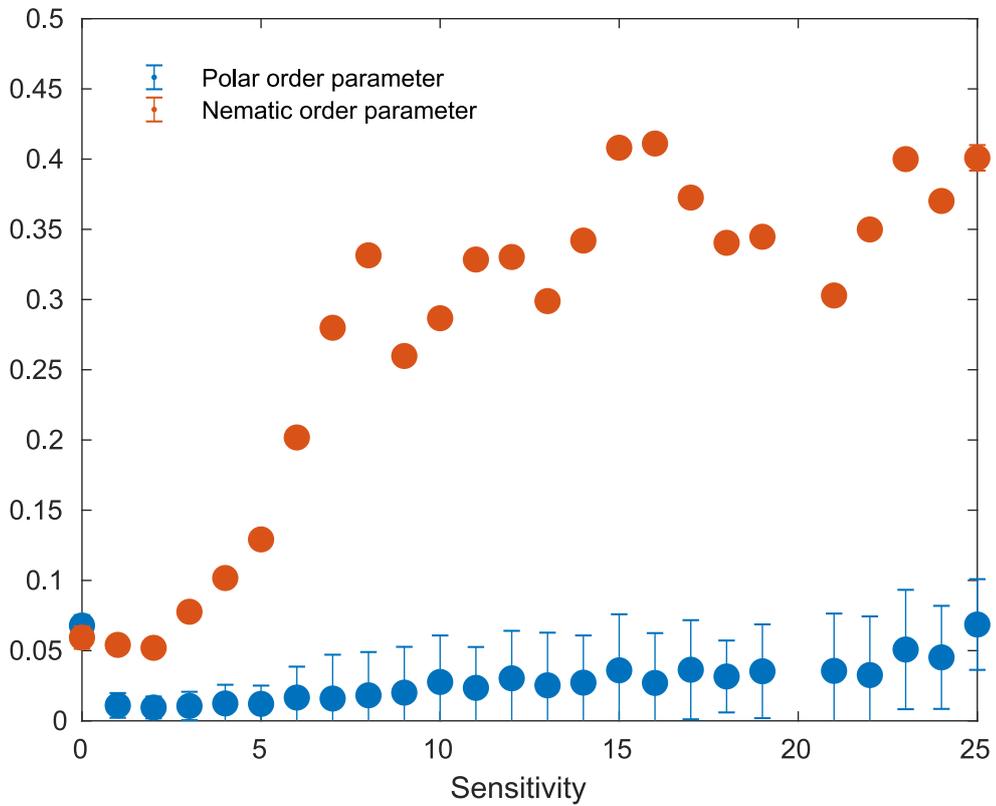

Figure 7.13.. Polar order parameter and nematic order parameter vs. angular sensitivity ($\beta > 0$) graph. Where interaction radius is at 3, and standard deviation of noise is at 0.05. Density is fixed at 2.5. Here, microenvironmental entropy has been taken from a wrapped exponential distribution. Here $g = 1$, $\beta_v = 0$, $\epsilon = 0$ and $\langle \xi_i^v(t)^2 \rangle = 0$. All the order parameters re averaged over 20 realizations after 10^3 time steps.

S.18. Average speed vs. radial sensitivity
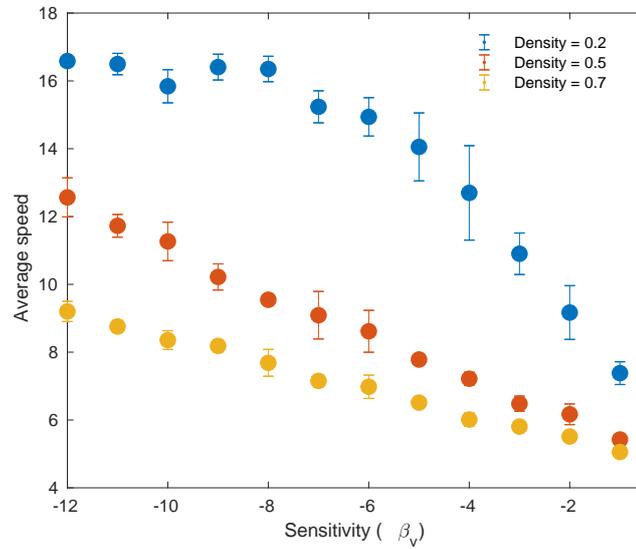
Figure 7.14.. Average speed vs. radial sensitivity (β_v) for different densities. The standard deviation of angular noise was fixed at 0.01, $\beta_\theta = 20$, $\psi = 0$, the box size has been fixed to 30 and the interaction radius was $R = 10$.

S.19. Probability distributions for speed

(i) $\beta_v > 0$

(ii) $\beta_v < 0$
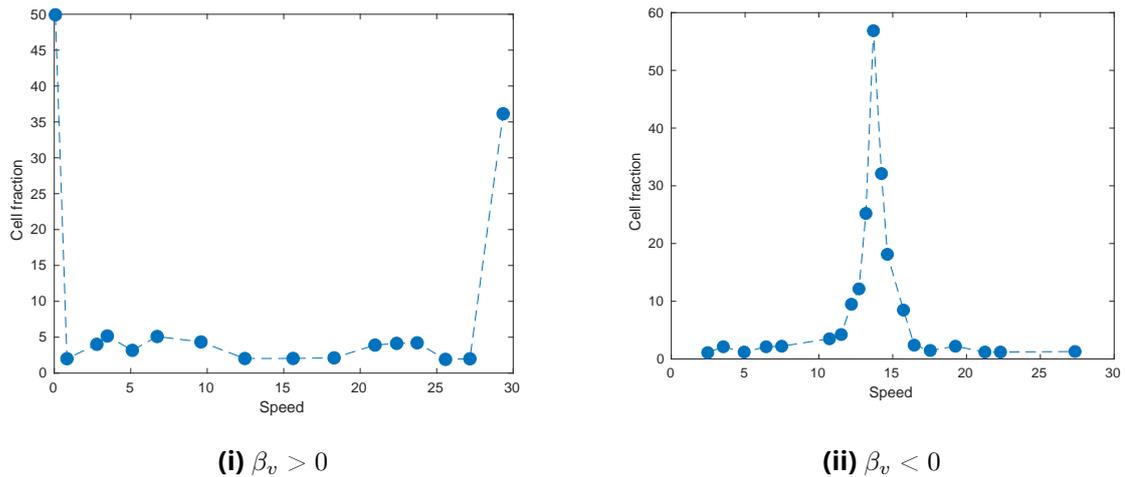
Figure 7.15.. The number of particles was fixed at 200, noise standard deviation at 0.05, β_θ is at -2 and interaction radius at 4. For (i) and (ii) β_v re fixed at 25 and -10. Simulations re averaged over 15 realizations after 200 time steps.

S.20. Average speed vs. density graph (depends on multiplicative noise)
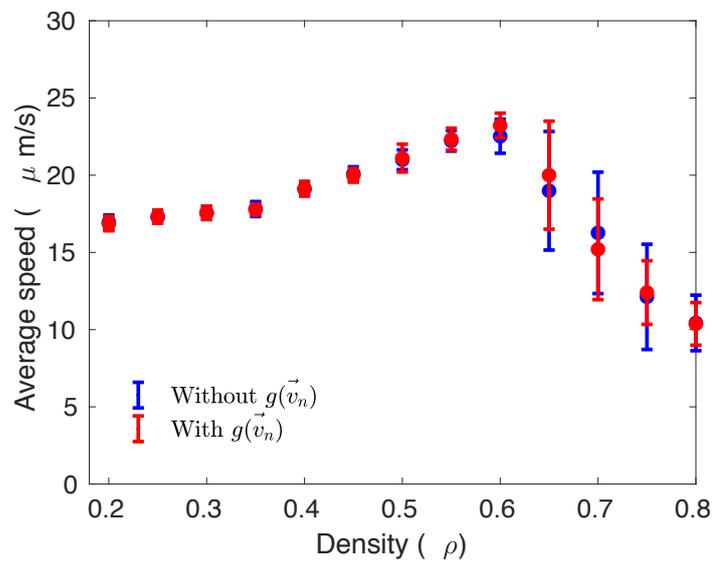

Figure 7.16.. Difference of average speed simulations, including or not the multiplicative noise term $g(\vec{v_i})$

.

## S.21. Averagve value of the friction term vs density graph
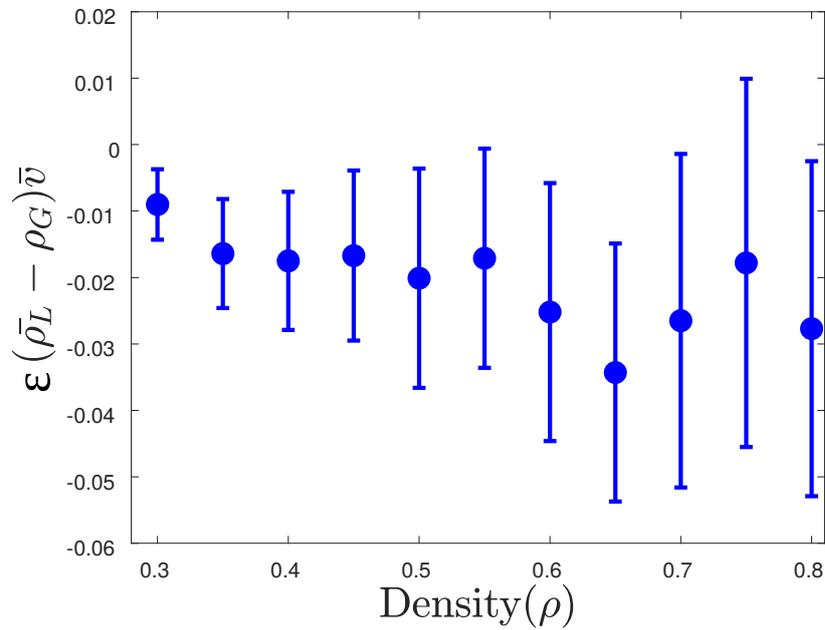
Figure 7.17.. Average value of the friction term in different densities. Throughout all simulations, the standard deviation of the noise was set at 0.0001, interaction radius at $R = 10$, proportionality constant $\epsilon = 0.008$, radial sensitivity $\beta_v = -20$, $g = \frac{1}{v_n}$ and angular sensitivity at $\beta_\theta = 20$. Data was obtained after 500 time steps.

S.22. Details on the experimental setup of *Serratia marcescens*

The experiment [130] has been done with gram negative bacteria *Serratia marcescens* 274. In the exponential growth phase at low density (i.e., 1×10^7 cells) the geometry of bacteria was rod shaped but due to starvation the geometry of the bacteria become spherical. The bacteria were grown to an OD650 of 2.0, corresponding to approximately 2×10^9 bacteria/ml. A 5-µl drop of an overnight (18 h) WT *S. marcescens* 274 culture have been placed on the glass slide. The density of the bacteria was fixed at 2×10^9 cells/ml initially and the aspect ratio was smaller than 1.1. Cells were coming and swimming on the upper surface of the drop. Surface density were increased from minimal to maximal, lasted approximately 20 min. The motion of the bacteria was independent of geometry of the drop, buyoancy, and gravity.

S.23. The proof of differentiation microreversibility relation
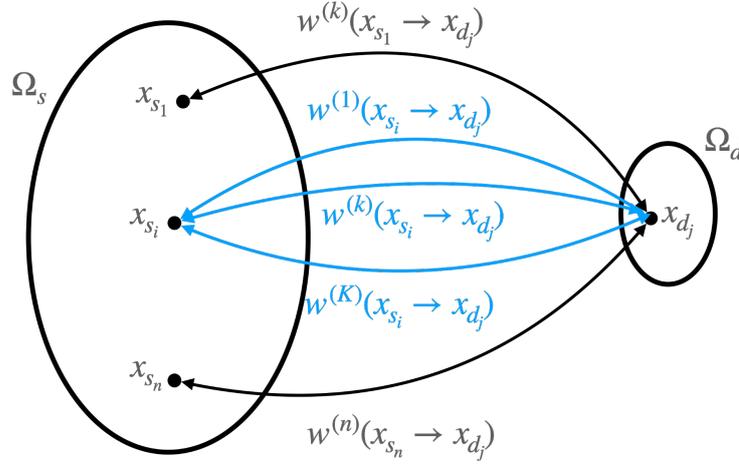
Figure 7.18.. Schematic transitions between pluripotent and differentiated microstates.

One can assume that a particular differentiated state belongs to the corresponding fixed point attractor which involves a number of realizations, $x_{d_j} \in \Omega_d = \{x_{d_1}, x_{d_2}, ..., x_{d_m}\}$. A pluripotent cell state belongs to an attractor – but not to a fixed point as is discussed in the main text – that involves the following realizations, $x_{s_i} \in \Omega_s = \{x_{s_1}, x_{s_2}, ..., x_{s_n}\}$, where $n \gg m$. The transition probability between a pluripotent x_{s_i} and a differentiated x_{d_j} microstate can be denoted as $w^{(k)}(x_{s_i} \to x_{d_j})$, where $k \in \{1, ..., K\}$ represents a possible path between the two states, see also Fig. 7.18. Now, by invoking the Crooks' theorem, the condition of microscopic reversibility can be written for a single path k as

$$\frac{w^{(k)}(x_{s_i} \to x_{d_j})}{w^{(k)}(x_{d_j} \to x_{s_i})} = \exp\left[\beta' Q_{ij}^{(k)}\right], \tag{7.17}$$

where $\beta' \equiv 1/T$, which T is the temperature of the heat bath. The quantity $Q_{ij}^{(k)}$ is the heat dissipation in the k-th path during the transition from x_{s_i} to x_{d_j}. Now, by averaging over all paths, the path-independent transition probability is written as

$$\frac{w(x_{s_i} \to x_{d_j})}{w(x_{d_j} \to x_{s_i})} = \left\langle \frac{w^{(k)}(x_{s_i} \to x_{d_j})}{w^{(k)}(x_{d_j} \to x_{s_i})} \right\rangle_k = \langle\exp\left[\beta' Q_{ij}^{(k)}\right]\rangle_k \geq \exp\left[\beta' Q_{ij}\right], \tag{7.18}$$

where $Q_{ij} = \langle Q_{ij}^{(k)}\rangle_k$ is the total heat dissipation over all paths (the lower bound is based on Jensen's inequality). The transition between a pluripotent state, x_s, to any differentiated state, x_d, can be interpreted as a transition from any x_{s_i} to any x_{d_j}. Therefore, one can write:

$$w(x_s \to x_d) = \sum_{i=1}^{n}\sum_{j=1}^{m} w(x_{s_i} \to x_{d_j}), \tag{7.19}$$

where the average over the corresponding paths is assumed. Now, the ratio of the forward/differentiation over the backward/dedifferentiation transition probability

reads

$$\frac{w(x_s \to x_d)}{w(x_d \to x_s)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} w(x_{d_j} \to x_{s_i}) \frac{w(x_{s_i} \to x_{d_j})}{w(x_{d_j} \to x_{s_i})}}{\sum_{i=1}^{n} \sum_{j=1}^{m} w(x_{d_j} \to x_{s_i})} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} w(x_{d_j} \to x_{s_i}) \langle \exp\left[\beta' Q_{ij}^{(k)}\right] \rangle_k}{\sum_{i=1}^{n} \sum_{j=1}^{m} w(x_{d_j} \to x_{s_i})}.$$

(7.20)

The last term in (7.20) can be viewed as a weighted average over all possible dedifferentiation paths between pluripotent and differentiated states, that is,

$$\frac{w(x_s \to x_d)}{w(x_d \to x_s)} = \left\langle \langle \exp\left[\beta' Q_{ij}^{(k)}\right] \rangle_k \right\rangle_{x_{d_j} \to x_{s_i}} = \langle \exp\left[\beta' \Delta Q_{x_s \to x_d}^{\tau}\right] \rangle_{x_d \to x_s},$$

(7.21)

where the latter has been used for notational simplicity. This is the microreversibility relation for a general differentiation process that corresponds to (5.20) in the main text.

8. Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässigeHilfe Dritterund ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlichgemacht.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die Dissertation wurde am Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH) der Technischen Universität Dresden unter Betreuung von Prof. Dr. Axel Voigt und Prof. Dr. Haralampos Hatzikirou angefertigt.

Es haben keine früheren erfolglosen Promotionsverfahren stattgefunden. Ich erkenne die Promotionsordnung der Fakultät Mathematik und Naturwissenschaften, Technische Universität Dresden vom 23.02.2011 an.

<div align="right">

Arnab Barua
16. August 2021

</div>