Bacterial chemotaxis: from information processing to behaviour

Gabriele Micali

Dissertation to obtain the degree PhilosophiæDoctor (PhD) and Diploma (DIC) of Imperial College London

Imperial College London

Department of Life Sciences

January 16, 2017
Supervisor:    Dr. Robert G. Entres
Advisor:       Dr. Giovanni Sena
Examiners:    Prof. Michael P.H. Stumpf
               Prof. Peter Swain
I declare that the work presented in this thesis which is not explicitly attributed to another person is my own.

Gabriele Micali

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Robert Endres, for the continuous support during all my Ph.D. His patience, positive mood and encouragements were an essential help to proceed with my work despite the obvious difficulties of academic research. Besides my supervisor, I would like to thank Prof. Michael Stumpf and Prof. Peter Swain for the stimulating discussion and insightful comments provided me during my viva. Furthermore, I thank my advisor Dr. Giovanni Sena for the funny and wise discussions through which I broadened my views over the field of biophysics. I would like to thank Prof. Victor Sourjik, Dr. Remy Colin and Dr. Silke Neumann for supplying the experimental data. Furthermore, I would like to acknowledge Prof. Stanislas Leibler, Prof. Thierry Emonet, and Dr. Yann Dufour for providing useful comments during conferences and meetings.

My sincere thanks also go to all the members of the biological physics group that I met over the years: Giovanna De Palo, Nikola Ojkic, Gerardo Aquino, David Richards and Linus Schumacher, and to my friends in Imperial College Laura Bouche and Velia Siciliano. Thanks for the feedback, the useful comments, inspiring conversations but also for the fun and support during coffee breaks, beers and (or better in particular) pizzas lunches. Particular thanks also go to Marianne Guenot for the feedback and help during the writing up.

A special thanks go also to my family, Carlo, Luciarita, Martina, Edoardo, Chiara, my girlfriend Raquel, my grandparents Anna, Antonio, Emilia, the rest of my large family and all my friends in London and elsewhere that always supported me and enjoyed with me many intense and beautiful moments during these intense years. Finally, I want to mention Lucia, Mario, and Pinotta which are not here anymore but are always with me.
Abstract

Chemotaxis allows flagellated bacteria to navigate in complex chemical environments, following nutrients and escaping toxins. The sensory system made up of chemoreceptors is constantly monitoring the extracellular concentrations of nutrients and toxins, while the signalling pathway processes and transmits the external information to the flagellated motors for movement. In the case of *Escherichia coli*, the chemotaxis pathway has been extensively characterised experimentally using genetics, biochemistry, and a wide range of imaging tools. This makes *E. coli* an ideal model organism for quantitative analysis and modelling. Several remarkable properties of the *E. coli* chemotaxis pathway have been summarised in terms of design principles. However, the swimming behaviour remains poorly understood, even for genetically identical cells in the artificial conditions normally used in a laboratory.

Here, I propose an interdisciplinary approach, which combines theory, computational simulations, and experimental data from my collaborators, to study *E. coli* chemotaxis from an information-theoretic point of view. I demonstrate that the *E. coli* chemotaxis pathway is designed to optimally transmit environmental information over a certain range of concentrations and gradients. To do so, I develop a theory that identifies both the responses and the environmental conditions that transmit maximal environmental information. Interestingly, when maximal information is transmitted, the behaviour characterised in terms of the drift velocity towards the nutrient is also maximised. A new design principle is proposed: maximal information transmission leads to maximal drift. Furthermore, the energetic cost of chemotaxis is much lower than the energy consumed to maintain the biological signalling pathway. Hence, thermodynamics does not seem to set constraints on information transmission and drift. However, to fully capitalise on my results, a closer connection with single-cell experiments is suggested.
## Contents

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract</strong></td>
<td>5</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>10</td>
</tr>
<tr>
<td>1.1 Decision making in bacteria</td>
<td>10</td>
</tr>
<tr>
<td>1.2 <em>Escherichia coli</em> chemotaxis</td>
<td>12</td>
</tr>
<tr>
<td>1.2.1 Classical view of <em>Escherichia coli</em> chemotaxis</td>
<td>13</td>
</tr>
<tr>
<td>1.2.2 Cell-to-cell variability and cell behaviour</td>
<td>18</td>
</tr>
<tr>
<td>1.2.3 Experimental measurements in literature</td>
<td>20</td>
</tr>
<tr>
<td>1.3 Information theory and biological channels</td>
<td>24</td>
</tr>
<tr>
<td>1.3.1 Introduction to information theory</td>
<td>24</td>
</tr>
<tr>
<td>1.3.2 Experimental and computational limitations</td>
<td>33</td>
</tr>
<tr>
<td>1.3.3 Successes of the theory in biology</td>
<td>36</td>
</tr>
<tr>
<td>1.3.4 Information processing and energy dissipation</td>
<td>40</td>
</tr>
<tr>
<td>1.4 Open questions and aims</td>
<td>45</td>
</tr>
<tr>
<td><strong>2 Instantaneous information transmission in the <em>Escherichia coli</em> chemotaxis pathway</strong></td>
<td>48</td>
</tr>
<tr>
<td>2.1 Overview</td>
<td>48</td>
</tr>
<tr>
<td>2.2 Classical methods to optimise mutual information</td>
<td>49</td>
</tr>
<tr>
<td>2.3 A novel, biologically inspired method for maximising mutual information</td>
<td>57</td>
</tr>
<tr>
<td>2.4 An adaptive algorithm</td>
<td>62</td>
</tr>
<tr>
<td>2.5 <em>Escherichia coli</em> chemotaxis responses and information processing</td>
<td>64</td>
</tr>
<tr>
<td>2.5.1 Information transmission at the chemoreceptors</td>
<td>65</td>
</tr>
<tr>
<td>2.5.2 Information transmission along the chemotaxis pathway</td>
<td>72</td>
</tr>
</tbody>
</table>
## CONTENTS

2.6 Discussion ................................................................. 77

3 Linking mutual information with chemotactic performance 81
   3.1 Overview: reversing the problem of sensing ................................. 81
   3.2 Identifying the distribution of attractant with maximal information transmission ................................................................. 83
   3.3 Identifying the gradients of attractant with high information transmission ................................................................. 88
   3.4 High information transmission leads to high drift velocity ................................. 91
   3.5 Discussion ................................................................. 96

4 Behaviours of individual *Escherichia coli* cells 99
   4.1 Overview ................................................................. 99
   4.2 Behaviour in terms of drift velocity ................................................................. 100
      4.2.1 An analytical theory for drift in shallow gradients ................................. 102
      4.2.2 The two peaks of drift velocity in exponential gradient ................................. 106
      4.2.3 Limits of the theory ................................................................. 108
   4.3 Experimental proof of the first peak ................................................................. 109
      4.3.1 Designing the experiments ................................................................. 113
      4.3.2 Analysing the experimental trajectories ................................................................. 114
   4.4 Additional insights on high-drift, high-information principle ................................................................. 116
   4.5 Discussion ................................................................. 121

5 Energy dissipation during chemotaxis ................................................................. 123
   5.1 Overview ................................................................. 123
   5.2 Entropy production during chemotaxis in exponential gradients ................................................................. 126
   5.3 Discussion ................................................................. 132

Conclusions ................................................................. 135

References ................................................................. 141

Appendix A: Description of the computational simulations ................................................................. 160

Appendix B: The linear theory for drift velocity ................................................................. 166
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Schematic of <em>E. coli</em> movement and chemotaxis pathway</td>
<td>14</td>
</tr>
<tr>
<td>1.2</td>
<td>Measuring the dose-response curves of the sensing channel</td>
<td>21</td>
</tr>
<tr>
<td>1.3</td>
<td>Interpretation of mutual information</td>
<td>29</td>
</tr>
<tr>
<td>1.4</td>
<td>Important observations about the use of mutual information in biology</td>
<td>33</td>
</tr>
<tr>
<td>2.1</td>
<td>Conventional ways of maximising mutual information</td>
<td>53</td>
</tr>
<tr>
<td>2.2</td>
<td>Total noise for optimal information transmission for different types of output noise</td>
<td>61</td>
</tr>
<tr>
<td>2.3</td>
<td>Adaptive evolutionary approach</td>
<td>64</td>
</tr>
<tr>
<td>2.4</td>
<td>Results of the adaptive algorithm</td>
<td>68</td>
</tr>
<tr>
<td>2.5</td>
<td>Comparison of optimal input-output curves with experimental data</td>
<td>70</td>
</tr>
<tr>
<td>2.6</td>
<td>Optimal instantaneous information transmission in <em>E. coli</em> chemotaxis pathway</td>
<td>78</td>
</tr>
<tr>
<td>3.1</td>
<td>The <em>Escherichia coli</em> sensory system is adapted to its chemical environment by evolution</td>
<td>82</td>
</tr>
<tr>
<td>3.2</td>
<td>Predicted distributions of chemical inputs and intracellular outputs</td>
<td>86</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulations of swimming bacteria in chemical gradients</td>
<td>91</td>
</tr>
<tr>
<td>3.4</td>
<td>Reconstruction of distributions of sampled gradients</td>
<td>93</td>
</tr>
<tr>
<td>4.1</td>
<td>Contributions to drift velocity</td>
<td>103</td>
</tr>
<tr>
<td>4.2</td>
<td>Swimming behaviour of identical cells in terms of trajectories and drift</td>
<td>107</td>
</tr>
<tr>
<td>4.3</td>
<td>Drift velocity in different solutions and for different phenotypes</td>
<td>110</td>
</tr>
<tr>
<td>4.4</td>
<td>Design of the experiments</td>
<td>112</td>
</tr>
<tr>
<td>4.5</td>
<td>Experimental setup and preliminary data analysis</td>
<td>115</td>
</tr>
<tr>
<td>4.6</td>
<td>Experimental verification of peak</td>
<td>117</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

4.7 Connection between behavior and information processing . . . . . . . . . 120

5.1 Connection between behavior and energy dissipation . . . . . . . . . . . 133
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Receptors sensitive regime for the input-output curve of cheRcheB mutants</td>
<td>65</td>
</tr>
<tr>
<td>2.2</td>
<td>The Hill coefficients of the input-output curves for different models</td>
<td>69</td>
</tr>
<tr>
<td>2.3</td>
<td>The fold increase in the transmitted mutual information</td>
<td>71</td>
</tr>
<tr>
<td>3.1</td>
<td>Fit parameters of the variance in FRET activity</td>
<td>87</td>
</tr>
<tr>
<td>3.2</td>
<td>Log-normal fits to predicted input distributions</td>
<td>88</td>
</tr>
<tr>
<td>4.1</td>
<td>Model for the E. coli chemotaxis pathway</td>
<td>101</td>
</tr>
<tr>
<td>4.2</td>
<td>Parameters in E. coli chemotaxis</td>
<td>102</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Decision making in bacteria

Decisions making is critical in nature, allowing organisms to identify and choose among alternatives based on existing knowledge. From animals to bacteria, living organisms are constantly processing information from their surroundings. This information is subsequently encoded internally, e.g. in internal target proteins levels, finally resulting in a behavioural response. An erroneous response usually penalises the individual. For example, a prey may need to escape when threatened, otherwise it may die or be injured. In addition, an unsuccessful escape might waste energy, penalising the prey in the next move.

Evolution has most likely selected for the most efficient decision-making. From the sensory system to the signalling processes involved in the decision-making, the whole pathway is under evolutionary pressure and it needs to be designed to provide reliable responses to selected stimuli. Despite the importance of understanding how organisms extract environmental information and make decisions, not much is known about the underlying principles. In most cases, the pathways involved are extremely complicated or unknown, which makes the identification of the internal ‘knowledge’, i.e. the level of the key proteins, impossible. In addition, decisions are usually not made based on a singular
environmental cue; they result from multiple stimuli being processed simultaneously. Moreover, biological processes are affected by noise, which may deteriorate the signal.

Being single-celled organisms, bacteria are often used as model organisms for their relative simplicity. However, even in bacteria, not much is known about the decision-making process orchestrating cellular behaviours. The study of decision-making in bacteria began in the 1940s, when Jacques Monod studied *Escherichia coli* cells growing in a medium with both glucose and lactose [Muller-Hill, 1996]. He observed that even though there was still lactose available to metabolise, the bacteria’s exponential growth stopped when the glucose was consumed, and only resumed after approximately an hour. The experiment highlighted that although the cells are genetically identical, the enzymes necessary to metabolise different nutrients are not all produced at the same time. In this case, *E. coli* only starts producing lactose-metabolising enzymes once the supply of glucose starts to diminish, suspending the bacterial growth until metabolism can resume. It is only decades later, with the help of fluorescence microscopy, that the heterogeneity of cellular behaviours in genetically identical cells was observed directly [Elowitz et al., 2002, Hekstra and Leibler, 2012]. Target experiments were designed to investigate the incidence of noise on cellular outputs. For instance, Elowitz and co-authors observed that two genes which were under control of identical mechanisms, were not necessarily expressed simultaneously [Elowitz et al., 2002]. Two different components were identified: a co-variable component which has a common effect on the expression of several genes, called extracellular or ‘extrinsic’ noise, and an independent component which is due to the intrinsic stochasticity in gene expression, called internal or ‘intrinsic’ noise [Elowitz et al., 2002]. This experiment demonstrated the complexity and multiplicity of biological noise.

It follows that while isogenic bacteria can adapt their responses to extracellular
changes in the environment, they can still behave differently under similar external conditions. Noise is unavoidable for cells and although it might be considered detrimental (as it deteriorates the signal), it also enhances the chance of survival, for example by allowing phenotypic switching [Balaban et al., 2004, Kussell and Leibler, 2005]. In order to understand the role of noise and decision-making in bacteria, experiments need to go hand-in-hand with theoretical modelling.

To quantify the information transmitted via a noisy channel, in his paper ‘A mathematical theory of communication’ published in the ‘40s, Claude Shannon formalised a theory called information theory [Shannon, 1948]. This theory, which was developed for telecommunications and later used in physics and neuroscience, has recently been applied to biological systems. Indeed, information theory provides a universal language to describe the information flow in cell signalling and hence may be applied to all levels of biology, regardless of temporal and spatial scales, across species and cellular pathways.

This thesis focuses on E. coli chemotaxis, i.e. the ability of E. coli to follow gradients of nutrients. The chemotaxis pathway has been well-characterised and dose-response curves have been measured both at the population and at the single-cell level. The signal is transduced by a mechanism spanning multiple time scales, from rapid receptor signalling, to the slower adaptation to persistent stimuli. Thus, E. coli chemotaxis is a perfect system to study how the environmental information is processed into behavioural responses.

1.2 Escherichia coli chemotaxis

Many reviews have recently emphasised the importance of E. coli chemotaxis as a model organism to reveal general design principles, which are believed to be applicable
1.2. ESCHERICHIA COLI CHEMOTAXIS

to many different systems [Kollmann et al., 2005; Sourjik and Wingreen, 2012; Tu, 2013; Parkinson et al., 2015; Micali and Endres, 2016]. This section introduces the chemotaxis pathway and summarises main findings. The aim is to motivate the claim that “E. coli chemotaxis can be considered the hydrogen atom of system biology” [Lan and Tu, 2016], i.e. a simple system from which we can learn general principles of how biology works at a quantitative level. This section conceptually follows the recent review Robert Endres and I published [Micali and Endres, 2016].

1.2.1 Classical view of Escherichia coli chemotaxis

E. coli is a Gram-negative bacterium. It is rod-shaped with typical dimension of $2\mu m \times 0.3 - 1\mu m$ and volume of around $0.7\mu m^3$. It is able to live both in aerobic and anaerobic conditions, habitually found inhabiting soils and gastrointestinal tracts of animals and humans. Inside the host, the majority of the strains contribute to the digestion of food and can provide protection against pathogens [Nataro et al., 2005]. However, some pathogenic strains can cause severe food poisoning. The bacteria strains with flagella are motile and able to perform chemotaxis, i.e. following gradients of chemical attractants, or able to escape lethal stresses, and they do so with a relatively simple chemotactic pathway.

Cell motility (swimming) is powered by 5 to 8 flagellar rotary motors, which can each rotate either in a clockwise direction (CW), or counterclockwise (CCW). The overall movement is determined by the rotation and interaction of these flagella, which result either in a ‘run’, a movement in a linear direction, or a ‘tumble’, a random reorientation [Darnton et al., 2007; Turner et al., 2000; Saragosti et al., 2011]. In absence of any chemical gradient, E. coli performs random walks, a combination of ‘runs’ and ‘tumbles’ without a particular direction. When a gradient is detected, this bacterium biases its movements
1.2. *ESCHERICHIA COLI* CHEMOTAXIS

Figure 1.1: Caption next page.
1.2. *ESCHERICHIA COLI* CHEMOTAXIS

Figure 1.1: **Schematic of *E. coli* movement and chemotaxis pathway.** (A) *E. coli* swimming behaviours from simulations in linear gradient. ‘Start’ and ‘End’ signify the beginning and end of a swimming trajectory. (B) *E. coli* chemotaxis pathway. The sensing module (dashed-lined red box) is made up of chemoreceptors of different types (here as an example, shown in brown and orange) which sense extracellular ligand concentration. Receptors are internally bounded to proteins CheA (A) and CheW (W). Proteins CheR (R) and CheB (B) regulate the methylation level of the receptors, methylating and demethylating the receptors, respectively. The methylation level of the receptors provides memory. When CheA is phosphorylated, it can in turn phosphorylate CheY (Y). CheY transduces the signal from the sensing module to the motility module (dashed-lined blue box). The motility module is made up of the motors, their C-ring and their flagella. CheY$_p$ randomly diffuses in the cytoplasm and binds to the C-ring of the motors via the FLiM protein. This regulates the flagella rotation, and enhances the possibility of the motor switching from CCW to CW rotation. By changing the number of FlIM proteins attached to the C-ring, the cell changes the bound affinity between CheY$_p$ and the C-ring, providing an additional mechanism of adaptation [Yuan et al., 2012]. CheY phosphorylation is further regulated by the protein CheZ (Z), which enhances CheY deactivation by dephosphorylation. Readapted from [Micali and Endres, 2016].

by having longer runs in the favourable direction (Fig. 1.1A).

To detect a gradient, the extracellular concentration of ligands needs to be measured and memorised. Both functions are carried out by the ‘sensing’ module (Fig. 1.1B, red box). The ligand is sensed by chemoreceptors. There are five types of chemoreceptors, the most abundant of which are the Tar and Tsr receptors, which respectively bind preferentially the amino acids aspartate and serine. Chemoreceptors are grouped in cooperative clusters of 10 to 20 receptors, which form an intracellular complex with the kinase CheA and the adapter protein CheW. Receptors in the active conformation activate the kinase CheA which autophosphorylates. In absence of any gradient of attractants, receptors maintain a default rate of activity (adapted activity). Once a change of extracellular concentration is detected, the receptors respond and the activity changes. The activity of the CheA is regulated by both the ligand binding events and the chemoreceptors’
methylation level. Counterintuitively, at fixed methylation level, an increased frequency of binding events eventually results in a decrease of receptor activity, i.e. a lower CheA_p. Each chemoreceptor has four methylation sites, and two receptors form a dimer with eight methylation levels in total. At fixed concentration (hence at fixed average binding events), the higher the methylation level, the more likely it is that CheA will be activated, the higher the receptor activity and vice versa [Tu, 2013, Endres, 2013]. Methylation level is controlled by CheR and CheB, which methylate and demethylate the chemoreceptors, respectively. There is a feedback loop between CheA activation and CheB activation (Fig. 1.1A) [Barkai and Leibler, 1997]. The methylation dynamics generates memory.

To perform chemotaxis, the signal from the ‘sensing’ module needs to be linked to the ‘motility’ module, which contains the flagellar motor (Fig. 1.1B, blue box). This is done by the regulating protein CheY. CheY is phosphorylated by CheA_p. CheY_p then randomly diffuses in the cytoplasm, eventually binding the FLiM protein which enhances the binding affinity between CheY_p and the internal part of the motors. Once attached at the C-ring of the motors, CheY_p changes the motor rotation from the default CCW to CW, which normally results in a tumble [Darnton et al., 2007, Turner et al., 2000, Sneddon et al., 2012]. For more information on the experimental characterisation of the pathway, see section 1.1.3.

Despite its simplicity, the chemotaxis pathway has been instrumental in the identification of many general sensory systems’ ‘design principles’ [Alon, 2006, Endres, 2013, Micali and Endres, 2016]. For instance, it detects small changes in chemical concentration with a sensitivity which is close to the physical limits of sensing [Berg and Purcell, 1977, Endres and Wingreen, 2009, Mao et al., 2003]. Furthermore, E. coli’s sensory system adapts to persistent stimulation, which is evidenced both by looking at the receptor activity
(using fluorescence resonance energy transfer, FRET, see Sec. 1.1.3) and by tracking the cell’s trajectories [Sourjik and Berg, 2002b, Berg and Tedesco, 1975]. The adaptation is considered ‘perfect’ for chemicals such as aspartate, meaning that there is a unique adapted activity level to which the activity returns after a persistent stimulation. Perfect adaptation has been explained by a robust mechanism, rather than a fine-tuning of the parameters involved [Barkai and Leibler, 1997]. This mechanism is called integral feedback control and it takes place at the sensing module where only inactive receptors are methylated by CheR and only active receptors are demethylated by CheB, [Barkai and Leibler, 1997, Alon et al., 1999]. The adaptation fulfils Weber’s law, i.e. $\Delta c/c = \text{const}$, highlighting that this bacterium measures relative changes in ligand concentration rather than absolute concentrations [Mesibov et al., 1973]. In addition, the temporal profile of the response during adaptation shows both fold-change detection [Tu et al., 2008, Shoval et al., 2010, Krembel et al., 2015] and Weber-Fechner law [Kalinin et al., 2009, Clausznitzer et al., 2014]. The former shows that the temporal profile is largely independent of the strength of stimulation, and the latter suggests that, in the sensitive regime, receptors perform logarithmic sensing, perceiving an exponential input as constant.

Cooperativity among receptors amplifies the signals, and the receptor activity at fixed methylation level, i.e. before adaptation, has been explained by the Monod-Wyman-Changeux (MWC) model [Keymer et al., 2006]. The MWC model was first introduced to describe the allosteric interaction in haemoglobin [Monod et al., 1965]. Briefly, a single receptor can be in one of two states, either active or inactive, and it can be either bound or unbound to the extracellular ligands. Furthermore, within a cluster, receptors are either all active or all inactive, i.e. there is full-cooperativity inside the cluster and zero with different clusters. Each of the four possible configurations (four methylation levels
per receptor) has its own energetic state and the free-energy difference between the active and inactive states has two contributions, one which depends on the ligand concentration and the other on the methylation level [Keymer et al., 2006]. Altogether, it is perfect adaptation and signal amplifications by cooperative receptors which allow *E. coli* to perform chemotaxis over a wide range of concentrations [Keymer et al., 2006, Endres and Wingreen, 2006]. However, note that adaptation is not always perfect [Masson et al., 2012, Min et al., 2012, Neumann et al., 2014] and the chemotaxis pathway is not only used to sense chemicals but also changes in temperature and pH [Oleksiuk et al., 2011, Yang and Sourjik, 2012]. Additional recent insights about variability and behaviours are highlighted in the next section.

### 1.2.2 Cell-to-cell variability and cell behaviour

Applying a theoretical approach to biology allows us to extract general principles which simplify molecular complexity and move the research towards a better understanding of the system. The design principles discussed in the previous section describe single cell sensing and signalling. However, bacteria live in large cellular communities. When accounting for cell-to-cell variability and different cells behaviours, new principles emerge.

Experimentally, the cell-to-cell variability is found at all scales [Spudich and Koshland, 1976]. At the level of a clonal cell population for instance, the adaptation time, the adapted activity and the dynamic range of CheY$_p$ can vary largely. This can be attributed to variation in copy-numbers of the molecular species involved in chemotaxis, in particular of CheR, CheB and CheY [Li and Hazelbauer, 2006, Vladimirov et al., 2008, Alon et al., 1999]. Furthermore, during their life cycle, cells might change behaviours [Leibler, 2016]. Variability at this level has been assumed to be helpful to face unpredictable changes in
the environment, whereby survival of the population is increased using the bet-hedging mechanism [Celani and Vergassola, 2010, Frankel et al., 2014].

Cells variability can also be seen on the level of total numbers and spatial distribution of receptors [Thiem and Sourjik, 2008] and motors [Cohen-Ben-Lulu et al., 2008]. The benefits of this heterogeneity is still not fully understood. Distributing receptors within small groups of clusters close to the flagella might be an effective mechanism to respond faster to extracellular changes, since the CheY$_p$ complex needs to migrate from the clusters to the flagella to transmit the information. The different number of flagella in the classical 'veto' model, in which one flagella rotating CW is sufficient to trigger a tumble, might drastically influence the tumble bias [Sneddon et al., 2012, Hu and Tu, 2013]. However, the use of optical traps, which allow long-time measurements of individual cells in response to external changes in ligand concentration, shows a coupling between flagella [Mears et al., 2014a]. According to these results, the veto model is valid, but only for an effective number of flagella, which is inferior to the number carried by the bacterium in reality (around three effective motors) [Mears et al., 2014a]. The tumble bias is therefore robust against an increase number of motors, which is a design principle which had not been identified before [Mears et al., 2014a].

Finally, a critical point for chemotactic cells is to tune their individual adapted CheY$_p$ levels towards the sensitive region of the motor dose-response curves, which are known to be ultra-steep, with a Hill coefficient of up to 20 [Cluzel et al., 2000, Yuan and Berg, 2013]. This seems to be achieved by adjusting the sensitive regime of the motor dose-response curves. Indeed, recent experiments highlight an additional mechanism of adaptation which acts at the motor level; within mutants with non-adapting receptors, Yuan and collaborators still observed an imperfect adaptation of the swimming behaviour, which
they attributed to the exchange of FliM molecules between the cytoplasm and the motor [Yuan et al., 2012]. This mechanism seems to be guided by a change in affinity between FliM and the motor substrate rather than a direct regulation by CheY\textsubscript{p} [Lele et al., 2012]. However, it is not clear whether this additional mechanism of adaptation can suffice to explain the matching between the adapted level of CheY\textsubscript{p} and sensitive region without fine-tuning.

### 1.2.3 Experimental measurements in literature

This section introduces the main experimental techniques reported in the literature which can be used to measure dose-response curves of the sensing and motility channels. For the purpose of this thesis, the dose-response curves are defined as the curves that describe the change in effect caused by differing levels of exposure (or doses) to a stressor, after a certain exposure time.

**FRET measurements for the receptors responses**

There are two types of stressors used experimentally, these can be a metabolisable chemoattractant, such as aspartate and serine, or a non-metabolisable chemoattractant, such as the methyl-DL-Aspartate (MeAsp), which mimics aspartate but is not degraded by the cells. The response of the receptors to these stressors is measured using fluorescence resonance energy transfer (FRET). A donor, a cyan fluorescent protein (CFP), is attached to CheZ and an acceptor, a yellow fluorescent protein (YFP), is attached to CheY [Sourjik and Berg, 2002a]. CheY and CheZ will only interact when CheY is phosphorylated [Sourjik and Berg, 2002a], at which point the fluorescent proteins they are coupled to will then exchange energy, and the fluorescence which results from the interaction will be...
Figure 1.2: **Measuring the dose-response curves of the sensing channel: Schematic of receptor response to changes in extracellular ligand concentration.** (A) Simulations of the sensing module response in the wild-type and mutant. Extracellular concentration of ligand profile in time (top), used as input for the wild-type (middle), and for non-adapting mutant, \( \text{cheR}\text{cheB} \), (bottom). Receptor response is measured over time. Blue dots signify the instantaneous response of a wild-type cell at fixed \( c^* \) (which correspond to fixed methylation level \( m^* \)), red ovals signify the response for the non-adapting cell at fixed \( c^* \) and \( m^* \). (B) Dose-response curves for the simulation shown in A. Theoretical prediction for fixed methylation level \( m^* \) (black solid line), instantaneous wild-type dose-response (blue dots, average of the initial response inside the blue dots in B), and mutant dose-response (red circles, average of the response in the red ovals in B). For experimental data, see [Sourjik and Berg, 2002b] and [Clausznitzer et al., 2014].
measured.

Wild-type cells respond to a step change in extracellular stressor concentration with an initial peak of activity. The system then changes the receptor methylation level thereby recovering the adapted activity level (Fig. 1.2A. For imperfect adaptation, see Neumann et al., 2014). By measuring the initial response, the dose-response curves at fixed methylation level, i.e. the instantaneous intracellular signalling response, can be acquired (Fig. 1.2B). Alternatively, to better quantify the dose-response curves at fixed methylation levels, it is possible to use non-adapting mutants. These mutants lack the CheR and CheB proteins (cheRcheB), and therefore cannot adapt by changing the receptor’s methylation level. Thus their internal proteins ‘follow’ the extracellular response, which makes the measurement more accurate (Fig. 1.2A). In order to mimic the various methylation level of the receptors, these mutant bacteria have been genetically engineered. Each of the amino acids glutamate (E) at the four receptor modification sites for methylation and demethylation can been substituted by a glutamine (Q). In absence of a CheR and CheB, E cannot be methylated, whereas Q naturally carries the methylation and this substitution effectively mimics a methylated site. Hence, genetically modifying some E into Q fixes the methylation level, e.g. mutant EEEE has 0 methylation level, mutant QEEE has 2 methylation level (one Q for each receptor in a dimer), QEQQ has methylation level 4, QEQQ 6, and finally QQQQ 8. Sourjik and Berg, 2002a. Non-adapting mutants’ average response is experimentally easier to measure than wild-type cells’ (Fig. 1.2B). However, the variability at the receptor level is lower for cheRcheB mutants than for wild-type cells Korobkova et al., 2004.
1.2. ESCHERICHIA COLI CHEMOTAXIS

Ultrasensitivity of the motor response

To measure the dose-response curve at the motility channel, i.e. the bias of individual motors as a function of CheY$_p$ at the single cell level, Cluzel and coauthors used *E. coli* mutants lacking the CheZ and CheB proteins, where the level of CheY-GFP is induced from plasmids [Cluzel et al., 2000]. In this way, the level of phosphorylated CheY in the cell is controlled, ranging from 0.8 to 6µM. The cells where immobilised with the flagella free to rotate, and the experiment was designed to measure the CheY$_p$ and the rotation of the flagella at the same time [Cluzel et al., 2000]. The resulting dose-response curve was fitted to a Hill function with Hill coefficient of 10.3 ± 1.1, [Cluzel et al., 2000].

More recently, an additional mechanism of adaptation at the motors was discovered. It acts in the time scale of minutes, thus much slower than all other time scales involved in chemotaxis. This mechanism depends on the number of FliM molecules attached to the intracellular part of the motor [Yuan et al., 2012]. This observation required new measurements of the motor dose-response curve where the number of FliM molecules attached to the motor is controlled. Using non adapting mutants (*cheR*-*cheB*), the authors measured the motor bias soon after a change in the attractant concentration. This short time-frame allowed them to neglect the change in number of FliM proteins bounds to the motor. The data show an even steeper response, with a Hill coefficient up to 20.7 ± 1.6 [Yuan and Berg, 2013]. Note that in this second work, the level of CheY$_p$ is not directly measured, but inferred from previous models.
1.3 Information theory and biological channels

1.3.1 Introduction to information theory

The key quantity of information theory is the Shannon entropy $H$ [Shannon, 1948]. Being $X$ a random variable, mathematically defined as a measurable function from a probability space $(\Omega_X, \mathcal{X}, p)$ (where $\Omega_X$ is the set of outcomes, $\mathcal{X}$ is the $\sigma$-algebra i.e. the collection of all the possible events and $p$ is the probability measure) into a measurable space, which is normally $\mathbb{R}$, the Shannon entropy of a discrete random variable $X$ is

$$H(X) = -\sum_{\Omega_X} p(x_i) \log_2 p(x_i). \quad (1.1)$$

The information content of one outcome $x_i \in \Omega_X$ is the logarithm of its probability of occurrence, and the entropy is the average over the all the possible events, i.e. $H(X) = -\langle \log_2 p(x_i) \rangle_{\Omega_X}$. For convenience, the logarithm is in base 2, providing $H$ conventionally in units of bits. Shannon information is the only measure which satisfies the following three natural postulates [Shannon, 1948]

1. The entropy is a statistical property of the probability of observing an outcome, hence $H$ should be a continuous function of $p$ only;

2. If the probability is uniform over the sample $\Omega_X$, then the bigger the sample the higher the entropy. Hence, $H$ should be a monotonic increasing function of $|\Omega_X|$ (where $|\Omega_X|$ denotes the dimension of the ensemble);

3. The entropy of two independent realisations is additive: $H(x_1, x_2) = H(x_1) + H(x_2)$. 

24
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

Note that the colloquial meaning of ‘high information’ might be misleading, since information is here linked with uncertainty: a rare event carries more information than a common event. To give an example, imagine that you want a friend to guess the word ‘example’. If you suggest to him/her that one of the letters is ‘x’, you are giving him/her more information than if you suggest that there is one ‘e’. That is because ‘e’ is more common (higher probability) than ‘x’ (lower probability) in the English language.

Shannon extended the concept of entropy and uncertainty to a transmitting noisy channel. Here, the channel is composed of a ‘transmitter’, which encodes and sends the message, i.e. the ‘input’, and a ‘receiver’ which receives and decodes the message, i.e. the ‘output’. All the steps are noisy. To describe the reduction of uncertainty of the random variable in the input, \( X \), after measuring the random variable in output \( Y \), Shannon introduced ‘mutual information’ \[\text{Shannon}, 1948\]. In particular, the information that the output \( Y \) communicates about the input \( X \), \( I[Y \to X] \), is defined by

\[
I[Y \to X] = H(X) - \langle H(X|Y) \rangle_{\Omega_Y},
\]

where \( \Omega_Y \) is the dominion of the random variable \( Y \) and the relative entropy is \( H(X|Y = y) = -\sum_{i \in \Omega_x} p(x_i|y) \log_2 p(x_i|y) \). Hence, using \( p(x_i, y_j) = p(y_j|x_i)p(x_i) = p(x_i|y_j)p(y_j) \)

\[
I[Y \to X] = \sum_{i \in \Omega_x} \sum_{j \in \Omega_y} p(x_i, y_j) \log_2 \left( \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right) = I[X; Y],
\]

which is symmetric in \( X \) and \( Y \). Note that Eq. \[1.3\] can be rewritten as

\[
I[X; Y] = \sum_{i \in \Omega_x} \sum_{j \in \Omega_y} p(x_i|y_j)p(y_j) \log_2 \left( \frac{p(x_i|y_j)}{p(x_i)} \right).
\]
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

Three properties follow directly from the definition of mutual information and are closely related with the three Shannon’s postulates for entropy.

1. $I[X;Y]$ depends on $p(x_i)$ and $p(x_i|y_j)$ only;

2. $I[X;Y]$ is symmetric in $x$ and $y$ (which allow us to write $I[Y \rightarrow X] = I[X;Y]$);

3. $I[X;Y]$ is invariant under reparametrisation (additive for independent variable).

Indeed, an alternative way of defining Shannon entropy is to first define the mutual information as in Eq. (1.2) with the respective properties as postulates. Then, the only $H$ which satisfies these conditions is the one in Eq. (1.1), effectively being an alternative formulation of Shannon theory (note that $I[X;X] = H(X)$). Note that the symmetry of mutual information states that this is a non-directional measure which does not provide any claim of causality. Although biological events might have a causal connection, it is expected that the uncertainty on the input after a given output has been observed is related with the uncertainty of observing an output when given a precise input. Alternative measures of information accounting for causality exist. For instance, directional information, commonly used in financial markets, has recently been used to describe evolution in biology [Marko, 1973, Massey, 1990, Rivoire and Leibler, 2011].

Mutual information quantifies statistical dependences between random variables. To show this, the Kullback-Leibler divergence (KL) needs to be introduced [Cover and Thomas, 1991]. Given $p(x_i)$ and $q(x_i)$ probability mass functions, the KL divergence is defined by:

$$\text{KL}(p||q) = \sum_{\Omega_X} p(x_i) \log \left( \frac{p(x_i)}{q(x_i)} \right).$$  \hspace{1cm} (1.5)

The KL divergence is always non-negative and zero if and only if $p = q$. However, it is not a mathematical distance, since it is not symmetric and it does not fulfil the triangle
inequality. By comparing Eqs. (1.4) and (1.5), the mutual information can be written in
term of KL, $I[X; Y] = KL(p(x, y)||p(x)p(y))$, then

4. $I[X; Y]$ is a non-negative quantity, which is zero if and only if $x$ and $y$ are independ-
ent.

Hence, mutual information is a measure of the statistical dependency between two vari-
ables which takes value in $0 \leq I[X; Y] \leq \max\{H(X), H(Y)\}$. As it was observed above,
mutual information is symmetric and invariant under reparametrisation, hence unit in-
dependent. A question arises spontaneously: which are the advantages of using mutual
information compared to conventional Pearson correlation as a measure of statistical de-
pendencies? Although both are symmetric and unit independent, Pearson correlation
quantifies linear dependences only \[\text{Steuer et al., 2002}].

In addition, mutual information provides meaningful interpretations about processing
information. Adding a step in information processing always decreases information (as
anyone who’s played the Chinese whispers game will know) \[\text{Cover and Thomas, 1991}].
Mathematically,

5. In a three step process $X \rightarrow Y \rightarrow Z$, $I[X; Z] \leq \min\{I[X; Y], I[Y; Z]\}$.

While by adding a process, information can only be lost not gained, multiple instantane-
ous measures of the same quantity may increase the information content. The chain
rule for information states that information increases by observing multiple inputs (out-
puts) \[\text{Cover and Thomas, 1991}\]

6. $I[X_1, X_2; Y] = I[X_1; Y] + I[X_2; Y|X_1]$,

where $I[X; Y|Z] = \sum_{\Omega_X} \sum_{\Omega_Y} \sum_{\Omega_Z} p(z) p(x; y|z) \log_2 \frac{p(x; y|z)}{p(x|z)p(y|z)}$. 

27
Finally, I mention one of the main successes of information theory: the noisy channel coding theorem. Imagine you are transmitting a message of length \( n \) (digits) by using a Shannon channel \( n \) times. Shannon showed that there is a code (i.e. a set of rules) to distinguish \( 2^nI \) input sequences, i.e. messages, for \( n \) sufficiently large. This way, mutual information can be seen as a rate of transmission of distinguishable digits out of a message of length \( n \) per use of the channel, given that the channel has been used multiple times. For example, 1 bit of information means that by using a channel \( n \) times (\( n \) sufficiently large), one message out of \( 2^n \) possibilities can be perfectly distinguished. The channel capacity \( C \) is the limiting information rate and it is given by

\[
C = \max_{p(X)} \{I[X;Y]\},
\]

i.e. the channel capacity is the mutual information optimised with respect to the input distribution. This theorem is used to optimise digital signal processing and to compress data.

**Extension to continuous random variables and interpretation of mutual information**

The Shannon entropy, Eq. \((1.1)\), and Shannon mutual information, Eq. \((1.4)\) can be extended to continuous random variables,

\[
H(X) = -\int_{\Omega_X} dx \ p(x) \log_2 p(x), \tag{1.7}
\]

\[
I[X;Y] = \int_{\Omega_X} \int_{\Omega_Y} dx \ dy \ p(x|y)p(x) \log_2 \left( \frac{p(x|y)}{p(x)} \right), \tag{1.8}
\]
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

Figure 1.3: **Interpretation of mutual information.** (A) Interpretation valid for discrete mutual information. Mutual information is the reduction of uncertainties after measuring the outputs, e.g. $I[Y;X] = H(X) - \langle H(X|Y) \rangle_Y$ (Eq. (1.2)). (B) Interpretation valid for both discrete and continuous mutual information. Mutual information measures the average ability to infer the inputs at given output. Input distribution (top), two different processes that have different likelihoods $p(o|i)$ (middle), the posterior $p(i|o)$ inferred for a single output $o = 0.4$ for both processes (bottom). On average, the process on the right, with the higher mutual information, enables better discrimination between the two inputs than the process on the left, as in the example presented here ($o=0.4$). Revised version of Fig.1 in [Bowsher and Swain, 2014].
when the integrals exist. It is important to note that in the continuous case, only mutual information is invariant under reparametrisation. Furthermore, Shannon entropy for continuous random variables can be negative, while mutual information remains non-negative [Cover and Thomas, 1991]. However, all the properties shown for the discrete case are still valid for the continuous mutual information. Note that for continuous random variables, $0 \leq \mathcal{I}[X;Y] \leq +\infty$.

In addition, it is worth mentioning that, for a Gaussian distributed random variable $X$ with standard deviation $\sigma$, the entropy is

$$H(X) = \frac{1}{2} \log_2(2\pi\sigma^2), \quad (1.9)$$

and for two Gaussian distributed random variables $X$, $Y$ with correlation coefficient $\rho$, the mutual information is

$$\mathcal{I}[X;Y] = -\frac{1}{2} \log_2 \left(1 - \rho^2\right). \quad (1.10)$$

Indeed, for Gaussian variables, entropy and mutual information are directly linked to variance and Pearson correlation, respectively, and mutual information can be seen as the generalisation of the correlation coefficient to nonlinear variables.

Mutual information is a measure of the reduction of uncertainty about the input after measuring the output only in the discrete case (Fig. 1.3A). This interpretation relates to the definition of mutual information in Eq. (1.2). However, for continuous random variables, the Shannon entropy can be negative, thus mutual information loses this interpretation [Bowsher and Swain, 2014]. However, both for the discrete and continuous case, mutual information quantifies the ability to infer the input from the output (Fig.
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

The higher the mutual information, the better the inference, on average [Bowsher and Swain, 2014]. Note that this is true under the assumption that the prior distribution for the inference is equal to the real input distribution [Bowsher and Swain, 2014].

For the properties and interpretations presented in this chapter, the mutual information is linked with both the efficiency of encoding the external inputs and reliability of transmitting them to an output. In particular, the channel capacity set the fundamental limits for both the optimal encoding and the optimal transmission reliability [Rivoire and Leibler, 2011]. Furthermore, Shannon’s theory of information is based on the events probabilities rather than accounting for the details of the channel. For those reasons, mutual information has become a central quantity for many systems of communications and more recently for many biological systems [Cover and Thomas, 1991, Schneidman et al., 2003, Levchenko and Nemenman, 2014, McMahon et al., 2014, Bowsher and Swain, 2014].

However, the Shannon channel is substantially different from many biological channels, e.g. the Shannon channel is memory-less. In some particular cases, alternative quantities have been proposed [Rivoire and Leibler, 2011, Rivoire and Leibler, 2014, Rivoire, 2016]. Furthermore, although this thesis focuses on the encoding rather than on the decoding, i.e. the estimation of the input given the output, it is worth pointing out that mutual information does not necessarily tell us anything about the decoding mechanism, there are several potential ways to decode the information which may or may not be used by living organisms [Schneidman et al., 2003].

Information theory and statistical independence

As it has been shown, mutual information is a measure of statistical dependence, which is zero if and only if the two random variables are independent. Schneidman and
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

co-authors focused on disentangling correlations of neurons activity, and identified three types of independence [Schneidman et al., 2003]. Take, for instance, two outputs, \( y_1, y_2 \) and a single input \( x \), there is:

- **Activity independence**: The two random variables in output \( Y_1 \) and \( Y_2 \) are activity independent, i.e. \( p(y_1, y_2) = p(y_1)p(y_2) \), if and only if \( I[Y_1; Y_2] = 0 \).

- **Conditional independence**: Two outputs \( Y_1 \) and \( Y_2 \) which are independent at given input \( X = x \), i.e. \( p(y_1, y_2|x) = p(y_1|x)p(y_2|x) \), are conditionally independent. A measure of conditional independence is

\[
I[Y_1; Y_2|X] = \int dy_1 dy_2 p(y_1, y_2|x) \log_2 \frac{p(y_1, y_2|x)}{p(y_1|x)p(y_2|x)}.
\]

Its average over the possible inputs \( \langle I[Y_1; Y_2|x]\rangle_X \) is also a valuable measure of conditional independence.

- **Information independence**: Two outputs can capture different features of the input. If the sum of the information encoded in the two outputs separately is the same as the information encoded in the two coupled outputs, i.e. \( I[Y_1, Y_2; X] = I[Y_1; X] + I[Y_2; X] \), the two outputs are information independent. Synergy, mathematically defined by

\[
\text{Syn} = I[Y_1, Y_2; X] - I[Y_1; X] - I[Y_2; X],
\]

measures the information independence, which can be a positive or negative quantity. Negative values indicate redundancy while positive values indicate synergy.

Overall, information theory provides an excellent framework to classify independent events, to distinguish if the statistical dependencies arise from shared inputs or from
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

Figure 1.4: Important observations about the use of mutual information in biology.
(A) Schematic of the experimental underestimation of mutual information. Experimentally, the mutual information between the measured stimulus $s_m$ and the measured response $r_m$ is measured. For the data processing inequality, this is a lower estimation of the mutual information between the stimulus $s$ and the response $r$. (B) Schematic of the computational overestimation of mutual information. 30 (left) and 300 (right) randomly distributed points in the 2d plane from simulations in R. Histogram of mutual information for 1000 simulations (bottom). Neither the histogram for the 30 points simulation (red) nor the one for 300 points (blue) match the expected mutual information which is zero (independent events). (C) Three possible topologies of a three node network starting in $g_1$. Since processing signals decrease information, the minimum mutual information is $\mathcal{I}[g_1; g_3]$ for A, $\mathcal{I}[g_1; g_2]$ for B and $\mathcal{I}[g_2; g_3]$ for C.

shared sources of noise, and to understand how multiple outputs encode the same input, which are all useful tools when applied in a biological contest.

1.3.2 Experimental and computational limitations

Two subtle problems occur when mutual information is measured either experimentally or from computational simulations. Firstly, the mutual information between measured inputs and measured outputs is always an underestimation of the mutual informa-
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

tion between the inputs and the outputs. In contrast, advanced computational techniques have to be carefully implemented not to overestimate mutual information.

To understand to what extent experiments limit the access to the mutual information, let us consider a noisy biological channel where a signal \( s \) is converted in a response \( r \), and imagine an experiment in which someone is attempting to measure the mutual information between \( s \) and \( r \). Since the person does not have access to \( s \) and \( r \), they need to experimentally measure the signal distribution \( p(s_m) \) which is related to the real distribution of signal \( p(s) \), and the responses distribution \( p(r_m) \), which relates to the ‘real’ \( p(r) \) (Fig. 1.4A). A good experiment gives a correlation between real and measured signal (response) of 90%, \( \rho = 0.9 \). Independently of the actual distributions of signal and response, using the data-process inequality, the upper limit of \( I[s_m; r_m] \) is given by the minimum between \( I[s; s_m] \), \( I[s; r] \) and \( I[r; r_m] \). Hence, assuming \( s (r) \) and \( s_m (r_m) \) Gaussian random variables with correlation coefficient \( \rho \), using Eq. (1.10),

\[
I[s; s_m] = I[r; r_m] = \frac{1}{2} \log_2(1 - 0.81) \approx 1 \text{ bit}, \tag{1.11}
\]

which means that only \( I[s; r] \) lower than about 1 bit can be estimated from \( I[s_m; r_m] \). This example shows how hard it is to measure the mutual information from experiments based only on one input and one output. To hope to get at least 3 bits of information, an extremely good experiment with \( \rho \approx 0.99 \) is needed. One way to increase the mutual information is to account for multiple signals and multiple responses, or interpret the input and the output as a time series [Cheong et al., 2011]. In this way, by implementing the chain rule (property 6 in the previous section), it is possible to raise the information of the system. However, careful experimental measurements are needed to estimate this information.
Mutual information is a measure of the full dependency between two random variables. This property makes mutual information hard to be evaluated from relatively small datasets, thus requiring large amounts of experimental or computational data (Fig. 1.4B). In mathematical terms, to evaluate the mutual information, the sum in Eq. (1.4) or the integral in Eq. (1.8) needs to be solved. However, such an integral is not linear in the probability distribution, which makes it hard to find an unbiased estimator for the mutual information. Many attempts have been done in recent years to find such unbiased estimators [Steuer et al., 2002]. Here, I initially present the problem which arises from a naive approach based on evaluating the probability distributions from fixed bin histograms, and subsequently, I report an improvement of this algorithm [Kraskov et al., 2004].

Figure 1.4B shows the quantification of the mutual information evaluated from datasets of 30 and 300 measurements of two independent continuous random variables $x$ and $y$. In this case, it is known that mutual information is zero. However, by accounting for the probability distribution from a histogram of fixed size bins, $M$, and performing the calculation for different datasets, a systematic overestimation of mutual information is observed (Fig. 1.4B) [Steuer et al., 2002]. Although this overestimation is due to finite size estimation of entropy, and hence can be overcome by enlarging the dataset, this might be hard to implement in experiments and simulations of biological problems [Grassberger, 1988]. The overestimation can be quantified for both entropies and mutual information [Herzel et al., 1994]. In particular,

$$\langle I^{\text{obs}} \rangle = I + \frac{M_{x,y} - M_x - M_y + 1}{2N},$$

where $N$ is the number of points in the dataset, and $M_x$, $M_y$ and $M_{x,y}$ denote the
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

number of discrete states (histogram bins) with a non-zero probability. Note that this approximation can be considered true only for a dataset of which the size is much bigger than the number of bins.

Possible improvements can be achieved by using an adaptive partitioning algorithm or a kernel density estimator [Steuer et al., 2002]. However, the best known estimator is the $k$-nearest neighbours [Kraskov et al., 2004]. This algorithm is based on evaluating the entropy from the average distance to the $k$-nearest neighbours. It is adaptive (higher resolution for bigger dataset) and has minimal bias compared with the others estimators. In particular, the systematic error scales as a function of $k/N$ has two main implications. First, the systematic errors are minimised by using small $k$; second, this approach can be used even for small size dataset (even fewer than 30 points [Kraskov et al., 2004]). However, this estimator needs to be tuned, since small $k$ reduces systematic but increases statistical errors (compared with higher $k$). Note that this algorithm assumes that the probability distribution is uniform inside the $k$-neighbour ball.

Finally, for a particular set of problems where the probability distribution can be approximated with joint Gaussian distributions, mutual information can be evaluated by using Eq. (1.10). Furthermore, even when probabilities are not Gaussian, this equation set a lower limit for mutual information (see methods in [Mitra and Stark, 2001] for the proof of $I[G;Y] \geq I[X;G;Y]$, where $G$ refers to Gaussian distributions).

1.3.3 Successes of the theory in biology

Despite the limitations identified in the previous section, mutual information has achieved important results when applied to biology. For instance, mutual information is used to reconstruct gene and protein interactions. Indeed, mutual information measures
the statistical dependency between random variables, e.g. concentration of chemical species. Using the data process inequality (property 5 in Sec. 1.3.1), it is possible to make predictions on the topology of one unknown chemical pathway. Take the three proteins $g_1$, $g_2$ and $g_3$ of Fig. 1.4C. These three proteins interact with each other and can be seen as nodes of a network where the links represent the interaction. If the interactions are unknown, the network can be in one of the three topologies shown in Fig. 1.4.

From the information theory point of view, the data process inequality guarantees that further species in the network have lower mutual information. By calculating mutual information between each pair of proteins, it is possible to select the unknown topology. For instance, in Fig. 1.4C if $I[g_1; g_3] < \min\{I[g_2; g_3], I[g_1; g_2]\}$ the selected topology is (a), if $I[g_1; g_2] < \min\{I[g_2; g_3], I[g_1; g_3]\}$ the topology corresponds to (b), otherwise the network is (b). By extending this idea, Margolin et al. built a software called ARACNE, which is able to reconstruct the gene regulatory networks from the experimental data, under certain conditions [Margolin et al., 2006]. In particular, they need to assume that loops have minor effects in such a network. Furthermore, by measuring the mutual information between the input and outputs of the tumour necrosis factor (TNF) pathway, Cheong et al. were able to discard the possibility of a direct regulation (‘bush’ network), proving the existence of an upstream bottleneck (‘tree’ network) in the pathway [Cheong et al., 2011].

In a developing embryo, identical stem cells acquire information according to their spatial position in order to adopt the appropriate fate. It is known that this positional information is achieved by measuring the local concentrations and activity of certain genes. This process is remarkable, since these crucial molecules are present at low concentrations, thus noise must limit the transmission of information. Formalising the prob-
lem using information theory has recently improved our understanding [Dubuis et al., 2013, Tkacik et al., 2014]. While in a *Drosophila* embryo, most of the cells’ genes are observably similar, depending on their position, a small set of gap genes provide a clear and reproducible path along the anterior/posterior axis of the embryo. By measuring mutual information between the gene activities and the position along the anterior/posterior axis, each gap gene encodes at least about 2 bits of information, which means that the on/off representation of this gene is inaccurate [Dubuis et al., 2013]. Furthermore, the patterns of the four gap genes’ expressions provide enough information to specify the position of individual cells with a precision lower than the average distance between two close nuclei [Dubuis et al., 2013]. However, the total information is not enough to specify the absolute position of every cell uniquely and the additional information might come from spatial comparison with the other cells.

Mutual information can be useful to distinguish the meaningfull variation in isogenic cells from confounding noise. For example, mutual information can be used to evaluate the relative impact of intrinsic and extrinsic noise in a signalling pathway [Mc Mahon et al., 2015], and the relative impact and interference generated by different inputs transduced into a single output [Mehta et al., 2009, Bowsher and Swain, 2012].

Other important successes of the theory are reported in recent reviews [Levchenko and Nemenman, 2014, Mc Mahon et al., 2014]. In [Levchenko and Nemenman, 2014], authors suggest that a change in how information is processed which results in a change of mutual information might be related to cell diseases. For instance, cancer can be interpreted as an information disease where the output (over-proliferation of the cells) is highly uninformative of the actual external level of growth factors. In this context, evaluating

---

1 A Boolean description (a two state active/inactive) of a gene activity is inaccurate when the information carried by the gene is larger than one.
mutual information is preferable to using correlations or gathering direct measurements of dose-response curves. That is because the sources of information loss can be varied and non-linear. For instance, by measuring the average dose-response curve a change in the noise behaviour might be missed, while mutual information takes into account all changes.

Furthermore, information theory can help design the best experiment to evaluate an unknown parameter [Liepe et al., 2013, Liepe et al., 2014]. In Mc Mahon et al., 2014, the authors show two simple examples. Firstly, they proved that in order to measure the rate at which the transcription factor Hes1 enters the nucleus, measuring Hes1 mRNA would be preferable to measuring Hes1 nuclear concentration, since the mutual information between the transcription rate and the mRNA is higher than the one between the transcription rate and the nuclear concentration of Hes1 protein. Secondly, when studying the downstream response to an epidermal growth factor signal under conditions which are not possible to implement experimentally, the authors selected, from a range of 12 available experimental set-ups, the one which most increases the predictions of behaviours.

Finally, Shannon information and other information theoretical quantities have been related over the years to evolutionary advantages of cellular populations [Kelly, 1956, Bergstrom and Lachmann, 2004, Kussell and Leibler, 2005]. In particular, more recently, Taylor and co-authors suggested that there is a minimum required level of mutual information between growth factors and extracellular stimuli to reach a certain grow rate, thus linking the abstract concept of mutual information with biological fitness [Taylor et al., 2007].
1.3.4 Information processing and energy dissipation

The Shannon entropy for a discrete random variable $X$, $H = -\sum p(x) \log_2 p(x)$, is mathematically proportional to the thermodynamic entropy, $S = -k_B \sum p(x) \ln p(x)$, where $k_B$ is the Boltzmann constant and in thermodynamics, the natural logarithm is used instead of base 2 logarithm. While the Shannon entropy is defined for any probability distribution, the thermodynamic entropy relates to the thermodynamic probabilities. Indeed, if the random variable $X$ represents the microstate of the system, the two entropies are conceptually identical. However, the link between information theory and thermodynamics does not end here, and goes all the way back to the Gedanken experiment proposed by Maxwell in the 19th century [Parrondo et al., 2015].

The second law of thermodynamics states that it is impossible to produce order without doing some work and hence consuming energy. However, Maxwell imagined a container internally divided in two parts ($A$ and $B$) filled with molecules of gas at the same temperature. A demon, the so-called ‘Maxwell demon’, starts to let the fast particles move from $A$ to $B$, and the slow particles from $B$ to $A$, for instance by opening the gate between the two compartments at the right time. At the end of the experiments, compartment $A$ will have lower temperature than $B$. Hence, the demon is able to make order without energy consumption, breaking the second law of thermodynamics. However, in order to do that, the demon has to measure the position and the velocities of the particles, i.e. acquiring information, thus this example shows the existence of a link between thermodynamics and information. Thanks to the use of fluctuations theorems and stochastic thermodynamics on the theoretical side, and technological advances for ex-

---

2For instance, a system coupled with a thermal bath is described by the canonical potential where the thermodynamic probability, i.e. the microstate of the system, is distributed following the Gibbs distribution.
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

Experimental purposes, progress has been done to clarify the link between thermodynamics and information theory [Parrondo et al., 2015].

Taking a measurement moves the system out of equilibrium. To illustrate this point, let us imagine a particle in a box. At equilibrium, it is assumed that the particle is equally likely to be in any position, with the statistical state being \( p(x) \) uniform. However, once a measure of the position of the particle has been done, the statistical state is suddenly \( p(x|m(t)) \) (state of \( x \) conditional to the measure \( m \) at time \( t \)) which is non-uniform, and so the system is in an out-of-equilibrium state. The extension of the thermodynamics concepts from equilibrium to out-of-equilibrium systems is controversial, since often there is no satisfying definition of physical entropy for these systems [De Groot and Mazur, 2013, Seifert, 2012]. However, Shannon’s entropy takes a thermodynamical meaning, not only for the equilibrium system, but also for the out-of-equilibrium systems coupled with thermodynamic reservoirs [Seifert, 2012].

The energetic state of a thermodynamic system coupled with a thermal bath at the equilibrium is described by the free energy \( F_0 \)

\[
F_0 = E_0 - TS_0, \tag{1.12}
\]

where \( E_0 \) is the energy of the system, \( T \) is the temperature, and \( S_0 \) is the entropy. If the equilibrium distribution of states is \( p_0(x) = e^{-\beta \mathcal{H}_0(x)}/Z_0 \) (i.e. canonical equilibrium state), then the entropy \( S_0 \) is the Shannon entropy (apart from a constant that I am neglecting here). \( \mathcal{H}_0(x) \) is the Hamiltonian of the system, i.e. the mechanical description of the system, \( \langle \mathcal{H}_0 \rangle_{p_0} = E_0 \), and \( Z_0 \) the partition function. Furthermore, Eq. (1.12) can be generalised for a non-equilibrium system coupled with a thermal bath where the states
are in a generic distribution $p$

$$\mathcal{F} := \langle H_0 \rangle_p - T S_p. \quad (1.13)$$

Note that this extension is valid for a non-equilibrium isothermal process. The system itself might have no well-defined temperature, but it is essential to the coupling with the thermal bath [Parrondo et al., 2015]. The extension of the second law of thermodynamics to non-equilibrium states connected by an isothermal process is [Parrondo et al., 2015, Gaveau and Schulman, 1997, Esposito and Van den Broeck, 2011]

$$T \Delta S = W_{\text{diss}} = W - \Delta \mathcal{F} \geq 0, \quad (1.14)$$

where $W$ is the average work performed on the system. This equation is found by considering the change of entropy of a non-equilibrium system made by a reversible contribution due to heat flow and an irreversible non-negative contribution.

The free-energy change due to only a measurement $m$ of the state of the system $p(x)$, i.e. $p(x) \rightarrow p(x|m)$, averaged over all the possible outcomes of the measurement is

$$\Delta \mathcal{F}^{\text{meas}} = \sum_m p(m) \mathcal{F}_{p(x|m)} - \mathcal{F}_{p(x)}. \quad (1.15)$$

Since no work has been done on or performed by the system, the Hamiltonian is the same
and the microstate has not changed, \( \sum_m p(m) \langle \mathcal{H}_0 \rangle_{p(x|m)} = \langle \mathcal{H}_0 \rangle_{p(x)} \), Eq. (1.15) becomes

\[
\Delta F_{\text{meas}} = -T \Delta S_{\text{meas}}
\]

\[
= kT \left( \sum_{m,x} p(m)p(x|m) \log p(x|m) - \sum_x p(x) \log p(x) \right)
\]

\[
= kT \mathcal{I}[X, M] \geq 0. \tag{1.16}
\]

Note that the constant \( k \) appears as conversion between the Shannon entropy \( H \) and thermodynamic entropy \( S \). This equation provides the link between thermodynamics and information. However, is this information a physical entity?

Landauer provided a positive answer to this question by observing that the outcomes of a measurement need to be stored in a physical memory \cite{Landauer1992}. The status of a memory can be seen as a random variable \( M \). There is an equilibrium state for the memory; when a measurement is recorded, the system is driven out of equilibrium, eventually going back to the equilibrium state, erasing the memory. Furthermore, being information stored in a physical object, the mechanics of the memories is subjected to a Hamiltonian \( \mathcal{H}_M \). Indeed, the definition of non-equilibrium free energy, Eq. (1.13), can then be extended to the memory system: \( \mathcal{F}_M = \langle \mathcal{H}_M \rangle_{p_m} - T S_{p_m} \). Note that for symmetric memories, i.e. with the same energetic contribution, the free energy change is

\[
\Delta \mathcal{F} = -kT \Delta H(M). \tag{1.17}
\]

Now, to have a complete view, the system needs to account for both the states \( X \) and memories \( M \). Assuming that states and memories interact only from an informational point of view, the Hamiltonian of the super-system \( XM \) is \( \mathcal{H}_{XM} = \mathcal{H}_X + \mathcal{H}_M \), and the
1.3. INFORMATION THEORY AND BIOLOGICAL CHANNELS

non-equilibrium free energy is

\[ F_{XM} = \langle H_{XM} \rangle_{XM} - TS_{XM} \]

\[ = \langle H_X \rangle_X + \langle H_M \rangle_M + kT \sum_x \sum_m p(x, m) \log p(x, m) \]

\[ = \langle H_X \rangle_X + kT \sum_x p(x) \log p(x) + \]

\[ + \langle H_M \rangle_M + kT \sum_m p(m) \log p(m) + kT \sum_x \sum_m p(x, m) \log \frac{p(x, m)}{p(x)p(m)} \]

\[ = F_X + F_M + kT I[X, M]. \] (1.18)

If the system is measured without affecting the state \( X \), hence only the memory has evolved, the free energy change in the super-system is

\[ \Delta F_{XM} = \Delta F_M + kT I[X, M]. \] (1.19)

In a system in which the initial memory has zero entropy, for a symmetrical memory, the free energy change of the memory is \( \Delta F_M = -kT H(M) \) (given by Eq. (1.17)), and for the free-error measurements the mutual information is \( I[X, M] = H(M) \). In this case, the free energy of the super-system \( XM \) (Eq. (1.19)) is zero. Hence, it is possible to acquire information at zero work \( (W \geq \Delta F_{XM} = 0) \). However, to reset the memory, the free energy is \( \Delta F_M^{\text{reset}} = -\Delta F_M^{\text{meas}} \). Note that erasing the memory does not affect the previous measurement. Finally, the total work is \( W = W^{\text{meas}} + W^{\text{reset}} \geq kT I[X, M] \) which is always non-negative. This is the Landauer principle: even if it is possible to acquire information without energy consumption, work is needed to erase information.

Note that the Maxwell demon first measures, then changes the system. Hence, it can use the free energy of the measurement to modify the system, and in this way the average
work can potentially be zero before the erasing the memories.

1.4 Open questions and aims

This thesis presents the first consistent study of how information flows in the *E. coli* chemotaxis pathway, with the aim of linking the information transmission with cell behaviour and energy consumption by the cells. As shown in this chapter, the *E. coli* chemotaxis pathway is experimentally well-characterised and has remarkable properties such as high accuracy of sensing and precise adaptation over a broad range of concentrations. This makes *E. coli* ideal to investigate how the extracellular information is internally encoded and subsequently turned in complex behaviour. To measure information flow, this chapter introduced concepts from information theory. In particular, mutual information has been identified as the key quantity to quantify the ability that cells have to infer the extracellular concentrations. Furthermore, this chapter showed the link between mutual information and thermodynamics. Using these mathematical and physical tools, the thesis aims to address the following fundamental topics:

**Instantaneous information transmission along the chemotaxis pathway.** *E. coli* chemoreceptors are known to work close to the theoretical limit of sensing. Thus, high accuracy is expected to result in high mutual information between the extracellular concentration and the internal response regulator (protein CheY). However, the motor response is ultra-sensitive with an ultrasteep measured dose-response curve. This might result in a bottleneck of information transmission, where the high information transmitted by the receptors is wasted at the motors. Chapter 2, firstly, defines mathematically the optimal achievable mutual information for a biological system subject to evolution.
1.4. OPEN QUESTIONS AND AIMS

Secondly, the theoretical optimal dose-response curves are compared with the experimental data available in literature. Finally, a few insights to solve the paradox of information flow along the whole chemotaxis pathway are suggested.

**Prediction of the typical environment encountered by *E. coli* in the wild.**

Bacteria are normally studied under artificial conditions in a laboratory and very little is known about the conditions they experience in nature. In chapter 3, assuming that the sensory system has evolved to best transmit natural stimuli, information theory, computational simulations and experimental data from collaborators are used to identify the typical distributions of concentrations and gradients experienced by the cell in their natural environment.

**Linking mutual information with chemotactic performance.** Does high information transmission provide any behavioural advantages to the cells? In chapter 3, using simulations of swimming cells, I show how concentrations and gradients selected to transmit optimal mutual information at the receptors lead to a maximal drift velocity up the chemical gradient. This suggests a new design principle: optimal information transmission at the receptors results in a maximal drift velocity of the swimming cell.

**Classification of the behaviours of individual *E. coli* cells.** Although *E. coli* is one of the most studied and well-characterised bacteria, its chemotactic behaviour is complicated even in genetically identical cells and in artificial laboratory environments. In chapter 4, I classify the behaviour in terms of drift velocity. An analytical theory for drift is presented, and alternatives to mutual information (able to capture the time dependencies and memories) are suggested to further support the high-information, high-drift principle.
1.4. OPEN QUESTIONS AND AIMS

Energy dissipation during chemotaxis. The accuracy of sensing and the increase of information transmission are limited by the finite resources of the cells. For instance, a cell has a finite number of downstream proteins and finite energy available. In chapter 5, I present a study of the energy dissipated by a cell while climbing up a gradient. As mutual information is fundamentally linked with the thermodynamic entropy, I expected to observe higher dissipation in regions of higher drift. In contrast, my study shows that the *E. coli* chemotaxis pathway uses much more energy than required for chemotaxis, and, surprisingly, the dissipated energy is lower in regions of higher drift.
2. Instantaneous information transmission in the *Escherichia coli* chemotaxis pathway

2.1 Overview

Cells can be considered as input-output devices which transmit extracellular stimuli (input) via receptors and signalling pathways (communication channel) to internal proteins and ultimately cell behaviour (output). Via their internal pathways, bacteria process environmental cues and make decisions. The organism survival depends on the outcome of such decisions. The higher the amount of environmental information transmitted, the more reliable the response. Hence, one might expect that the responses are designed to optimise the transmitting information under biologically relevant constraints. These constraints might arise from the fact that cells cannot avoid cell-external and -internal noise, which both reduce signal reliability [McMahon et al., 2015], and they often pay an energetic cost to reduce such noise [Lan et al., 2012]. Moreover, cells are often performing multiple tasks at the same time, thus the solution to an individual problem might not be optimal since the cell needs to distribute finite resources to perform all of its duties.
Hence, the debate over the optimality of information transmission in sense and signalling is open.

An early success of the optimality theory relies on the response of the interneurons in a fly’s compound eye to changes in light intensity. Indeed, assuming maximal mutual information between the distribution of light intensity contrast (input) measured from natural scenes and the interneurons electrical responses (output), the measured responses match the theoretical optimal curve [Laughlin, 1981]. This result suggests that sensory systems might be designed to transmit high information. Escherichia coli has a well-characterised sensory system which works close to the optimal limit of sensing [Berg and Purcell, 1977, Mao et al., 2003]. To embed the problem in the original information theory approach, this chapter focuses on the instantaneous transmission of information of the chemotaxis pathway, thus neglecting time dependencies.

The first section reviews the classical methods to optimise mutual information. The second section introduces an innovative method to maximise mutual information aimed at biological systems based on evolution. The third section presents an evolutionary algorithm to achieve such an optimum. Finally, the fourth section compares the measurable and the theoretical optimal responses.

### 2.2 Classical methods to optimise mutual information

Information theory allows us to disregard the details of the channel, however it is important to uniquely define the input, the channel and the output. In biological systems, these definitions are not always clear: outputs might feed back into inputs providing
memories, and the channel generally interacts with the input. Moreover, genetic mutations may affect the channel, which is able to evolve under evolutionary pressure. Here, by focusing on instantaneous transmission of information, the feedbacks and memories are neglected. From the chemotaxis point of view, this is equivalent to measuring the information transmission of a cell adapting its behaviour after a perturbation arises. This setup motivates the use of the Shannon mutual information, which loses its interpretation otherwise [Rivoire and Leibler, 2011].

To be general, in this section I call ‘x’ the input and ‘y’ the output. For the self-consistency, the definition of mutual information for continuous random variable is reported here

\[ I[X; Y] := \int \int dxdyp(y|x)p(x)\log_2 \left( \frac{p(y|x)}{p(y)} \right). \quad (1.8) \]

The input and the output probabilities are connected by the Bayes rule

\[ p(y|x) = p(x|y)p(y), \]

once the conditional probability is known. Furthermore, the conditional probability of observing y at given x, \( p(y|x) \), relates with the average input-output curve and noise. For instance, if the channel is Gaussian,

\[ p(y|x) = \frac{1}{\sqrt{2\pi\sigma_T^2(x)}} e^{-\frac{(y-y'(x))^2}{2\sigma_T^2(x)}}, \quad (2.1) \]

where \( \bar{y}(x) \) is the average input-output curve and \( \sigma_T(x) \) standard deviation from the average. If the channel is not Gaussian other momenta matter. Considering a small-noise Gaussian channel, Eq. (1.8) becomes

\[ I[X; Y] = -\int_{x_{off}}^{x_{on}} dx p(x) \log_2 \left[ \frac{\sqrt{2\pi e \sigma_T(x)}}{\bar{y}(x)} p(x) \right], \quad (2.2) \]
where \( \tilde{y}'(x) = d\tilde{y}/dx \), and \( x_{\text{off}} \) and \( x_{\text{on}} \) set the sensitivity region, i.e. the dynamic range of inputs and I used the conservation of probability \( p(y)dy \approx p(\tilde{y})d\tilde{y} \approx p(x)dx \) valid in small noise approximation only \cite{Detwiler2000, Clausznitzer2014, Tkacik2008b, Tkacik2008a, Brunel1998}. \( \mathcal{I} \) in Eq. (2.2) is a functional of the input distribution \( p(x) \), the average input-output curve \( \tilde{y}(x) \) and the total noise \( \sigma_T(x) \). Moreover, it depends on the boundary conditions \( \tilde{y}(x_{\text{off}}) \) and \( \tilde{y}(x_{\text{on}}) \). The maxima of the functional \( \mathcal{I} \) with respect to the total noise lead to the trivial solution of vanishing noise, which is biologically unreal. Furthermore, if the boundaries are free to vary, the larger the input range (i.e. the higher \( x_{\text{on}} - x_{\text{off}} \)) the higher the mutual information. Hence, using the Euler-Lagrange formalism, where \( \mathcal{L} \) is defined by \( \mathcal{L} = p \log_2 \left[ \sqrt{2\pi e} \sigma_T \tilde{y}' p \right] \), and assuming fixed boundary conditions, there are two ways to maximise mutual information

\[
\max_{p(x)} \mathcal{I}[X,Y] \rightarrow \frac{\partial \mathcal{L}}{\partial p} = 0, \quad (2.3)
\]

\[
\max_{\tilde{y}(x)} \mathcal{I}[X,Y] \rightarrow \frac{\partial \mathcal{L}}{\partial \tilde{y}} - \frac{d}{dx} \frac{\partial \mathcal{L}}{\partial \tilde{y}'} = 0. \quad (2.4)
\]

Note that Eq. (2.3) is the equation for the channel capacity applied to a small-noise Gaussian channel (cf. Eq. (1.6)).

Eqs. (2.3) and (2.4) are the two classical ways to maximise mutual information. The channel capacity is the maximisation with respect to the input distribution at fixed input-output curve. It is normally used for electronic devices where the responses are fixed by construction and only the input distribution can be tuned. The channel capacity is also used as proxy for the effective mutual information in biological systems \cite{Tkacik2008b, Tkacik2008a}. This is widely used, for example, in the study of fly embryo development, where the mother cells might tune the optimal input, while the embryos...
might be provided with the same responses by having the same genetic content [Tkacik et al., 2008b, Tkacik et al., 2008a].

Alternatively, mutual information can be maximised with respect to the input-output curves assuming the input distribution fixed [Detwiler et al., 2000, Clausznitzer et al., 2014]. This method accounts for the evolution of the channel, e.g. genetic or phenotypic changes. However, it assumes that the cells are sensing the same input distribution. When the noise is uniform, \((\sigma_T = \alpha)\) Eqs. (2.3) and (2.4) coincide and the maximisation of the mutual information leads to the same result: \(\bar{y}' \propto p\). In this case, the input-output curve is the cumulative integral of the input distribution [Laughlin, 1981]. However, in general the noise might be considered a function of the input-output curve, \(\sigma_T = \sigma_T(x, \bar{y}, \bar{y}')\), which leads to different solutions (Fig. 2.1). Finally, the channel capacity is linked to the Fisher information [Bernardo, 1979, Brunel and Nadal, 1998] as it is explained in the next section.

**Mutual information and Fisher information**

Let be \(X\) a random variable and \(\theta\) a parameter which describes the likelihood of an event \(x\), in Bayesian statistics \(\theta\) is an outcome of the random variable \(\Theta\). In particular \(p(\theta)\) is called the ‘prior’ distribution, \(p(x|\theta)\) is the ‘likelihood’ and \(p(\theta|x)\) is the ‘posterior’ distribution, which are linked each other through the Bayes’ rule: \(p(x|\theta)p(\theta) = p(\theta|x)p(x)\). Indeed, the prior would affect the posterior inference of \(\theta\). Thus, a major concern is how to choose a prior in order to trust the resulting inference. Naively, one can assume a uniform prior distribution. However, under a change of parametrisation, i.e. a one-to-one map of \(\theta\) to a new variable \(\rho\), the resulting prior distribution of \(\rho\) might be non-uniform. Thus, a priori, there is no any reason to promote \(\theta\) over \(\rho\), since the choice of which parameter
2.2. CLASSICAL METHODS TO OPTIMISE MUTUAL INFORMATION

Figure 2.1: Conventional ways of maximising mutual information. The maximisation is normally done either with respect to the input distribution at given input-output curve and noise (red), or with respect the input-output curve at given input distribution (blue). Noise as a function of the input (top) and input-output Hill function (insets) are designed to be identical for both cases. (A) For small input noise, the two maximisations converge, showing similar input (middle) and output (bottom) distributions. (B) For large input noise, the two approaches predict different input (middle) and output (bottom) distributions.
to use seems equally reasonable. A class of possible priors, called ‘uninformative’ priors, aims to select priors which have the least effect over influencing the posterior distribution. Among them, the ‘reference’ priors, an extension of Jeffreys’ prior [Jeffreys, 1946], provide a formal answer of how a uninformative priors have to be choose [Bernardo, 1979].

Let be \( x = \{x_1, ..., x_n\} \) a set of events occurred, and \( \theta \) a parameter of the model. The reference prior is defined as the prior which maximise the Kullback-Leibler divergence between the posterior and the prior distributions

\[
\max_{p(\theta)} \text{KL}[p(\theta), p(\theta|x)] = \max_{p(\theta)} \int d\theta \ p(\theta|x) \log \frac{p(\theta|x)}{p(\theta)}.
\] (2.5)

Note that the average of such a KL divergence over all the possible outcomes of \( X \) is

\[
\int d^n x \ \text{KL}[p(\theta), p(\theta|x)] \ p(x) = \int d^n x \ p(x) \int d\theta \ p(\theta|x) \log \frac{p(\theta|x)}{p(\theta)} = I[x; \theta].
\] (2.6)

Hence, to find the reference prior, the maximal mutual information with respect to \( p(\theta) \) needs to be evaluated. In other words, the channel capacity considering \( \theta \) as an input and the reference prior are linked. Note that the reference priors provide a general framework that can be easily generalisable to multiple parameters.

Fisher information quantifies the information encoded in the events \( x \) about a parameter of the model \( \theta \). Mathematically,

\[
\mathcal{F}(\theta) = -\int d^n x \ p(x|\theta) \left( \frac{\partial^2}{\partial \theta^2} \log p(x|\theta) \right),
\] (2.7)

where \( p(x|\theta) \) is the likelihood function of \( \theta \) or the probability density of the random variable \( x \) given \( \theta \). Indeed, Fisher information is then the concavity of the log-likelihood
2.2. CLASSICAL METHODS TO OPTIMISE MUTUAL INFORMATION

function. Furthermore, the Fisher information is also the variance of the score, i.e. $\frac{\partial}{\partial \theta} \log p(x|\theta)$. This is straightforward since $\int d^n x p(x|\theta) \frac{\partial}{\partial \theta} \log p(x|\theta) = 0$ guarantees that $F(\theta) = -\int d^n x p(x|\theta) \left( \frac{\partial^2}{\partial \theta^2} \log p(x|\theta) \right) = \int d^n x p(x|\theta) \left( \frac{\partial}{\partial \theta} \log p(x|\theta) \right)^2$.

In general, it is possible to extract the value of $\theta$ from the outcomes of the random variable using an estimator $\hat{\theta} = \hat{\theta}(x)$ without using any inference method. A desirable property of an estimator is to be unbiased, i.e. its expected value is equal to the true value of the parameter ($< \hat{\theta} > = \int d^n x p(x|\theta) \hat{\theta}(x) = \theta$). If an unbiased estimator exists, its variance is linked to the Fisher information via the Cramér-Rao bound

$$\sigma^2_{\theta} = \left< \left( \hat{\theta}(x) - \theta \right)^2 \right>_{\theta} \geq \frac{1}{F(\theta)}.$$  \hspace{1cm} (2.8)

Note that the Fisher information is calculated using $p(\hat{\theta}(x)|\theta)$ instead of $p(x|\theta)$. Hence, the Cramér-Rao bound states that the Fisher information provides a lower bound on the estimate of an unbiased parameter $\theta$. When the equality holds, the estimator is ‘efficient’ and a famous example is the maximum-likelihood estimation, which is asymptotically efficient (for large $n$).

On one hand, mutual information is linked to a way of setting a prior over a parameter $\theta$. On the other hand, there is a way to quantify the information that a random variable carried of a model parameter which is called Fisher information. The question is, are mutual and Fisher information related? To show such a connection, the next paragraph follows the argument of Brunel and Nadal [Brunel and Nadal, 1998].

Let me consider the mutual information between an input $X$ and an output $Y$. The input $X = x$ is a parameter of the model, and that might be estimated with an unbiased estimator from the events $y = \{y_1, ..., y_n\}$. Hence, there is $\hat{x} = \hat{x}(y)$ with mean $x$ and
2.2. CLASSICAL METHODS TO OPTIMISE MUTUAL INFORMATION

variance $1/F(x)$. Then, the mutual information between $x$ and $\hat{x}$ is

$$I[x; \hat{x}] = -\int d\hat{x} p(\hat{x}) \log p(\hat{x}) + \int dx p(x) \int d\hat{x} p(\hat{x} | x) \log p(\hat{x} | x)$$

$$= -\int d\hat{x} p(\hat{x}) \log p(\hat{x}) - \int dx p(x) H(\hat{x} | x)$$

$$\geq -\int d\hat{x} p(\hat{x}) \log p(\hat{x}) - \int dx p(x) \log \sqrt{\frac{2\pi e}{F(x)}}.$$

Where to go from the second to the third line, I used the fact that the entropy of a Gaussian distribution of variance $1/F(x)$ is higher than the entropy of any other distribution with the same variance [Mitra and Stark, 2001, Brunel and Nadal, 1998]. In the limit of small noise $F(x) \gg 1$, and the entropy of the estimator is equal to the entropy of the variable, $\int d\hat{x} p(\hat{x}) \log p(\hat{x}) \approx \int dx p(x) \log p(x)$. Finally, since information can be only lost after further processing

$$I[Y; X] \geq I[X; \hat{X}] \geq -\int dx p(x) \log \sqrt{\frac{2\pi e}{F(x)}} p(x). \quad (2.9)$$

Eq. (2.9) links mutual information with Fisher information through an inequality. When an efficient estimator exists, the equality is expected to hold, and in such a case, maximising the mutual information with respect to the input distribution $p(x)$ results in

$$p(x) \propto \sqrt{F(x)}. \quad (2.10)$$

Note that for a biased estimators $\hat{x} = \hat{x}(y)$, where $\langle \hat{x} \rangle = \int dy p(y | x) \hat{x}(y) = f(x) \neq x$, the Cramér-Rao bound is

$$\sigma_x^2 \geq \left( \frac{df}{dx} \right)^2 \frac{F(x)}{F(x)}. \quad (2.11)$$
2.3. A NOVEL, BIOLOGICALLY INSPIRED METHOD FOR MAXIMISING MUTUAL INFORMATION

Using Eq. (2.11) instead of Eq. (2.8), Eq. (2.9) becomes

\[ I[Y; X] \geq - \int dx \, p(x) \log \sqrt{2\pi e} \sqrt{F(x)} \left( \frac{df}{dx} \right) \, p(x), \]  

(2.12)

which is the equivalent of Eq. (2.2). Additional details and the general case of non-Gaussian-channels can be found in [Brunel and Nadal, 1998].

2.3 A novel, biologically inspired method for maximising mutual information

In the previous section, the two classical ways of maximising mutual information are presented, see Eqs. (2.3) and (2.4). Figure 2.1 shows that in the general case in which the total noise is a complex function, the two maximisations provide two different solutions. However, is there a solution for the maximisation of mutual information with respect to both the input distribution and the input-output curve? To answer this question, the following system

\[
\begin{align*}
\max_{p(x)} I[X, Y] & \quad \Rightarrow \quad \frac{\partial L}{\partial p} = 0, \\
\max_{\bar{y}(x)} I[X, Y] & \quad \Rightarrow \quad \frac{\partial L}{\partial \bar{y}} - \frac{d}{dx} \frac{\partial L}{\partial \bar{y}'} = 0
\end{align*}
\]  

(2.13)

needs to be solved. When the total noise is a function of the input and the input-output curve, \( \sigma_T = \sigma_T(x, \bar{y}, \bar{y}') \), the ‘Lagrangian’, i.e. the integrand of Eq. (2.2), for a Gaussian conditional probability in the small noise approximation becomes

\[ L = -p(x) \log_2 \left[ \frac{\sqrt{2\pi e} \sigma_T(x, \bar{y}(x), \bar{y}'(x))}{\bar{y}'(x)} \right] p(x). \]  

(2.14)
To find the solution of Eq. (2.13), note that

\[
\begin{align*}
\frac{\partial L}{\partial p} &= -\log_2 \left[ \frac{\sqrt{2\pi e} \sigma_T}{\bar{y}'} p \right] - 1, \\
\frac{\partial L}{\partial \bar{y}} &= \frac{p}{\sigma_T} \frac{\partial \sigma_T}{\partial \bar{y}'}, \\
\frac{\partial L}{\partial \bar{y}'} &= \frac{p}{\sigma_T} \frac{\partial \sigma_T}{\partial \bar{y}'} - \frac{p}{\bar{y}'}, \\
\frac{d}{dx} \frac{\partial L}{\partial \bar{y}'} &= -\frac{p'}{\bar{y}'} + \frac{p' y''}{\bar{y}^2} - \frac{p' \partial \sigma_T}{\sigma_T \partial \bar{y}'} + \frac{p \sigma_T' \partial \sigma_T}{\sigma_T^2 \partial \bar{y}'} + \frac{p}{\sigma_T} \frac{d}{dx} \frac{\partial \sigma_T}{\partial \bar{y}'}.
\end{align*}
\]

Thus, Eq. (2.13) becomes

\[
\begin{cases}
p = \frac{\bar{y}'}{Z \sigma_T} \\
2 \frac{y''}{y'} - 2 \frac{\sigma_T'}{\sigma_T} + \frac{d}{dx} \frac{\partial \sigma_T}{\partial \bar{y}'} = -\frac{1}{\bar{y}'} \frac{\partial \sigma_T}{\partial x}.
\end{cases}
\]  

(2.15)

Where \(Z\) is set by the normalization of probability \(p\). By assuming independent input transmitted and output noise \(\sigma_T(x, \bar{y}, \bar{y}') = \sqrt{\sigma_X^2 \bar{y}^2 + \sigma_Y^2}\), the second equation in Eq. (2.15) can be written in terms of \(\sigma_X\) and \(\sigma_Y\)

\[
2 \frac{y''}{y'} - 2 \frac{\sigma_T'}{\sigma_T} + \frac{d}{dx} \frac{\partial \sigma_T}{\partial \bar{y}'} = -\frac{\partial \sigma_X^2}{\partial x} \frac{1}{2 \sigma_X^2}.
\]

After integrating over \(x\),

\[
\log \left( \frac{\bar{y}^3 \sigma_X^3}{\sigma_T^3} \right) = K.
\]
2.3. A NOVEL, BIOLOGICALLY INSPIRED METHOD FOR MAXIMISING MUTUAL INFORMATION

Where \( K \) is an integration constant given by boundary conditions (note that \( K < 0 \)).

Finally, the solution of the system Eq. (2.13) is

\[
\begin{cases}
    p = \frac{\bar{y}'}{Z \sigma_T}, \\
    \sigma_x^2 \bar{y}^2 = Q \sigma_y^2
\end{cases}
\]

where \( Q = e^{2K/3}/(1 - e^{2K/3}) \), and \( K \) is a constant set by the boundary conditions. The maximal mutual information is

\[ I = \log_2 \sqrt{2\pi e Z}, \]

where \( Z = \int dx \frac{\bar{y}'}{\sigma_T} \) depends on the optimal response \( \bar{y} \). The noise scales the optimal value of \( I \). However, a purely multiplying factor over the noise does not influence the optimal input-output curve.

The first equation in the system (2.16) extends the matching relationship found in [Laughlin, 1981] to non-uniform noise. In this general case, the optimal input distribution weighs inputs with low uncertainty more than inputs with high noise [Tkacik et al., 2008a, Brunel and Nadal, 1998, Komorowski et al., 2015]. The second equation in the system (2.16) states that the optimal input-output curve is designed to have transmitted input noise proportional to the output noise.

To test whether a system optimally transmits information, the input and output noise need to be extracted separately. This would allow checking whether the measured input-output curve matches the optimal curve. Although the total noise is measurable [Clausznitzer et al., 2014], experiments at the microscopic level need to be carefully designed to separately measure the input and output noise. Hence, here in absence of such experiments, I will make realistic theoretical assumptions on the biological nature of the input and output noise. As shown in the following sections, the input and output noise depend on parameters \( \alpha \)'s, which are effective, time averaged noise parameters, which eventually need to be extracted experimentally or modelled in more detail. This
chapter shows analytically traceable solutions which can be used to evaluate whether a sensory system is transmitting maximal information. Furthermore, this approach would help future experiments which aim to extract noise parameters.

Uniform noise, input noise only and output noise only

If $\sigma_T(x, \bar{y}(x), \bar{y}'(x)) = \sigma$, i.e. for uniform noise, $\partial L / \partial y = 0$ and Eq. (2.13) becomes:

$$
\begin{cases}
  p = \frac{\bar{y}'}{Z' 
  \frac{d}{dx} \left( \frac{p}{\bar{y}'} \right) = 0.
\end{cases}
$$

(2.17)

In this case, the two equations in the system (2.17) provide the same result and the system has a set of infinite solutions which lay on $p = \bar{y}' / Z'$, where $Z' = 1/\sigma \int dx \bar{y}'(x)$ is a normalisation constant. Thus the two approaches presented in Eq. (2.13) converge to the same result. Furthermore, either the input distribution or the input-output curve need to be known in order to find the other.

If $\sigma_T(x, \bar{y}(x), \bar{y}'(x)) = \sigma_X(x)$, meaning that there is no cell-internal noise (input noise only), $\partial L / \partial y$ is still zero. Thus again Eq. (2.13) becomes Eq. (2.17), and the solution is $p = \bar{y}' / Z'$, where now $Z' = \int dx \frac{\bar{y}'(x)}{\sigma_X(x)}$. If $\sigma_T(x, \bar{y}(x), \bar{y}'(x)) = \sigma_Y(\bar{y})$, meaning that there is no cell-external noise (output noise only), Eq. (2.13) becomes

$$
\begin{cases}
  p = \frac{\bar{y}'}{Z' 
  \frac{1}{\sigma_Y} \frac{d}{dy} \frac{\sigma_X}{\sigma_Y} - \frac{p'}{p} + \frac{\bar{y}''}{\bar{y}'} = 0.
\end{cases}
$$

(2.18)

Again, $p = \bar{y}' / Z'$ solves Eq. (2.18) with $Z' = \int dx \frac{\bar{y}'(x)}{\sigma_T(\bar{y}(x))}$. The three cases presented here, namely uniform noise, input noise only and output noise only formally have an
2.3. A NOVEL, BIOLOGICALLY INSPIRED METHOD FOR MAXIMISING MUTUAL INFORMATION

Figure 2.2: Total noise as a function of the input for optimal information transmission for different types of output noise. Output noise $\sigma^2_Y = \alpha_4$ (blue solid line), $\sigma^2_Y = \alpha_3 \bar{y} + \alpha_4$ (blue dashed line), $\sigma^2_Y = \alpha_2 \bar{y}(1 - \bar{y}) + \alpha_3 \bar{y} + \alpha_4$ (blue dotted line) and Eq. (2.24) (red solid line). Where $\bar{y}$ is the corresponding optimal average input-output curve. The input ‘a la’ Berg-and-Purcell noise $\sigma^2_X = \alpha_1 x$ is the same in all cases. Note that the peak in the total noise appears when the optimal response is sigmoidal (cf. with [Clausznitzer et al., 2014]).

Parameters: $\alpha_1 = 10^{-7}$, $\alpha_2 = 10^{-7}$, $\alpha_3 = 10^{-7}$, $\alpha_4 = 10^{-8}$, $\sigma_1 = 7 \cdot 10^{-7}$.

equivalent solutions. Indeed, when the input noise is small the two solutions converge (Fig. 2.1A).

Independent transmitted input and output noise

Here, different functions $\sigma_X$ and $\sigma_Y$ are considered. The input noise is considered Poissonian and it scales with the amplitude of the input strength (‘a la’ Berg-and-Purcell [Berg and Purcell, 1977]), i.e. $\sigma^2_X = \alpha_1 x$, while there are different models for the output noise. The general solution of Eq. (2.16) holds for any function, however here the following cases are considered

• $\sigma^2_Y = \alpha_4$, with $\alpha_4$ constant, which means uniform output noise (Fig. 2.2, blue solid line). The solution for the optimal input-output curve is then

$$\bar{y} = \pm 2 \sqrt{\frac{Q \alpha_4}{\alpha_1} \sqrt{x} + C}.$$
2.4 An Adaptive Algorithm

- $\sigma_Y^2 = \alpha_3 y + \alpha_4$, with $\alpha_{3,4}$ constants, which means Poissonian and uniform sources of output noise (Fig. 2.2 blue dashed line). The solution for the optimal input-output curve is then

$$\bar{y} = \frac{Q\alpha_3}{\alpha_1} x \pm \alpha_3 C \sqrt{\frac{Q}{\alpha_1}} \sqrt{x} + \left( \frac{\alpha_3 C^2}{4} - \frac{\alpha_4}{\alpha_3} \right).$$

- $\sigma_Y^2 = \alpha_2 \bar{y}(1 - \bar{y}) + \alpha_3 y + \alpha_4$, with $\alpha_{2,3,4}$ constants, which means introducing a switching noise source in addition to Poissonian and uniform sources of output noise (Fig. 2.2 blue dotted line). The solution for the optimal input-output curve is then

$$\bar{y} = \frac{\alpha_2 + \alpha_3 \pm \sqrt{(\alpha_2 + \alpha_3)^2 + 4\alpha_2 \alpha_4 \sin \left[ \sqrt{\frac{\alpha_2}{\alpha_1}} \left( C + \frac{2\sqrt{Qx}}{\alpha_1} \right) \right]}}{2\alpha_2}.$$

(2.19)

$C$ and $Q$ are given by imposing the boundary, $\bar{y}(x_{\text{off}}) = 1$ and $\bar{y}(x_{\text{on}}) = 0$, which also determines the sign where $\pm$ appears.

2.4 An adaptive algorithm

Until this point, this chapter has presented a new way to maximise the mutual information between a generic input and an output which involves the optimisation of both input distribution and input-output curves. However, how can a biological system reach this optimum?

Take for instance the chemotactic bacterium *E. coli*; the microenvironment in which this well-characterized bacterium lives is unknown. As an example, *E. coli* experiences changes in nutrient concentrations and gradients, as well as changes in temperature and pH. Even by restricting to only one stimulus, for example aspartate, the concentration
and gradients encountered by the bacteria are unknown. However, considering a wild type strain living in a defined environment, the input distribution can be defined as the probability distribution of observing a certain concentration of, for example, aspartate \( p(c) \). The receptors respond to aspartate by changing their activity (conformation), and an input-output curve, \( \bar{A}(c) \), is selected by evolution. Assuming the mutual information as a measure of fitness, in a given environment \( p(c) \), the \( \bar{A}(c) \) which maximises the mutual information is selected (mathematically Eq. (2.4), Fig. 2.3A).

The swimming behaviour of \( E. coli \) also influences the perceived input distribution. For instance, by changing the swimming velocity, or the run-and-tumble bias, cells might experience different input distributions, even in the same environment. Assuming that behaviours are selected to increase information transmission, which means that an optimal input distribution can be selected, the system is optimised with respect to the input distribution at given input-output curve (see Eq. (2.3), Fig. 2.3B). Subsequently, genetic mutations and phenotypic switching allow the average input-output curve to change again (Fig. 2.3C). This cycle is repeated multiple times and eventually converges to the mathematical solution in Eq. (2.13).

In this way, without knowing the environment, there are simple predictions for both input-output curves and input distribution. The predicted input-output curves can be compared with the internal dose-response curves measured using FRET, thus providing a way to test whether the system is evolved to maximally transmit information.
Figure 2.3: **Adaptive evolutionary approach.** (A) A given distribution of inputs for the environment in which bacteria live. (B) Information transmission is assumed to be a measure of fitness; it is assumed that mutations which create a phenotype matching the optimal input-output curve is selected by evolution. (C) Considering the input-output curve presented in (B), the optimal input distribution (solid line) generally differs from the initial input distribution (dashed line) presented in (A). The change in behaviour that results in a change in the inputs which increases the information transmission is selected by evolution. (D) In response to the change of the input distribution, the input-output curve adapts to maximise information transmission (full line). With each iteration of these processes, the system comes closer to a joint maximisation of both the input distribution and the input-output curve (Eq. (2.16)).

### 2.5 *Escherichia coli* chemotaxis responses and information processing

*E. coli* senses extracellular concentrations with its chemoreceptors, which act cooperatively and allow detection of small changes of ligands by amplifying the signals. By changing methylation levels, the receptors also maintain their sensitivity over a broad range of concentrations (see Chapter 1). Using the non-metabolisable analogue of aspartate, methylaspartate (MeAsp), as an input, the average receptor response at fixed
methylation level has been quantified, using FRET between CheY_p/CheZ [Sourjik and Berg, 2002a]. The average responses have been explained using the MWC model, and they show a relatively low Hill coefficient when fitted to an Hill function [Endres and Wingreen, 2006, Endres et al., 2008]. In contrast, the average input-output curves at the motors have the highest Hill coefficient ever measure in wild cells [Yuan and Berg, 2013].

This observation questions the idea of maximal information transmission. Indeed, such a switch-like motor response might limit the information transmission (for instance a step function only transmits one bit). Furthermore, the CheY_p level needs to be in the sensitive region of the motor to allow high information transmission. Here, I first apply the new method to maximise mutual information to the chemoreceptor response. Then I consider the whole pathway in order to compare the measurable responses from [Clausznitzer et al., 2014, Yuan and Berg, 2013] with the optimal responses from Eq. (2.16).

<table>
<thead>
<tr>
<th>$CheRcheB$ mutant</th>
<th>$c_{off}$ [$\mu M$]</th>
<th>$c_{on}$ [$\mu M$]</th>
<th>ligand conc. prod. half occupation [$\mu M$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>QEQE</td>
<td>0.017</td>
<td>0.163</td>
<td>0.056</td>
</tr>
<tr>
<td>QEQQ</td>
<td>0.115</td>
<td>0.323</td>
<td>0.189</td>
</tr>
<tr>
<td>QQQQ</td>
<td>0.329</td>
<td>0.801</td>
<td>0.497</td>
</tr>
</tbody>
</table>

Table 2.1: Receptors sensitive regime for the input-output curve of $cheRcheB$ mutants. Mutant modification, initial sensitive concentration, final sensitive concentration, and ligand concentration producing half occupation as columns. Mutants QEQE ($m^* = 4$), QEQQ ($m^* = 6$) and QQQQ ($m^* = 8$) as rows.

### 2.5.1 Information transmission at the chemoreceptors

Here the input is the concentration of extracellular methylaspartate (MeAsp), $c$, and the output is the level of the internal protein CheY_p, $Y$. Note that experimentally, it is...
only possible to measure the FRET activity, i.e. CheY\(_p\)/CheZ and not CheY\(_p\) directly. However, the total levels of CheY and CheZ are considered fixed and the FRET activity can be considered a readout of CheY\(_p\) (in relative units) \cite{Vladimirov2008}. In order to use the method exposed in Sec. 2.3, the noise and the boundary conditions, i.e. the sensitive region, need to be evaluated.

The noise is the independent combination of the transmitted input noise and the output noise, \( \sigma_T(c, \bar{Y}, \bar{Y}') = \sqrt{\sigma_c^2\bar{Y}^2 + \sigma_{\bar{Y}}^2} \), where \( \bar{Y}(c) \) is the average input-output curve. The input noise is considered proportional to the input strength, similar to the Berg-and-Purcell limit, \( \sigma_c^2 = \alpha_1 c \) \cite{Berg1977}. The output noise has three biological inspired contributions and is \( \sigma_{\bar{Y}}^2 = \alpha_2 \bar{Y}(1 - \bar{Y}) + \alpha_3 \bar{Y} + \alpha_4 \) \cite{Clausznitzer2014}. The first term is due to on/off switches of the receptor state \cite{Endres2008}, the second is a Poisson term, to mimic the stochastic (de)phosphorylation of CheY (CheYp), and the third term is a constant background noise. \( \alpha_1-4 \) are parameters, here arbitrarily chosen. Using this noise, the solution to the maximisation of the mutual information between \( c \) and \( \bar{Y} \) is given by Eqs. (2.16) and (2.19).

In the MWC model, the receptor activity is given by

\[ A = \frac{1}{1 + e^F}, \]  

(2.20)

from which CheY\(_p\) can be evaluated (\( Y \propto A \)). \( F \) is the free-energy difference and depends on the concentration of extracellular ligand \( c \) and methylation level \( m \)

\[ F(c, m) = N \left[ \epsilon(m) + \sum_r \log \left( \frac{1 + c/K_{r\text{off}}}{1 + c/K_{r\text{on}}} \right) \right]. \]  

(2.21)

Here, \( N \) and \( \epsilon(m) \) are the number of receptors in a cooperative receptor cluster and the
energy at given methylation level, respectively. $K_{on}^r$ and $K_{off}^r$ are the ligand dissociation constants in the on and off state, respectively, and $r$ denotes the type of receptor in the cluster [Keymer et al., 2006]. Focusing on Tar receptors only, to extract the sensitive regime at fixed methylation level, $m$ is fixed and $N$ and $\epsilon(m)$ are taken from [Endres et al., 2008]. The sensitivity is defined by $\delta A / (\delta c / c)$ [Endres and Wingreen, 2006]. Thus, the sensitive regime is found by solving

$$\frac{\delta A}{\delta c/c} = \frac{\max\left\{ \frac{\delta A}{\delta c/c} \right\}}{2} \bigg|_{m=m^*}$$

(2.22)

for $c$, where the activity $A$ is given by Eqs. (2.20) and (2.21) evaluated at fixed methylation level $m = m^*$ with $m^* = 4, 6, 8$ for QEQE, QEQQ and QQQQ mutants respectively. In this way, $c_{on}$ and $c_{off}$ are selected (see Table 2.1). The response is assumed to span the whole range of responses, thus ranging from $Y(c_{off}) = 1$ to $Y(c_{off}) = 0$ in relative units.

Having extracted the noise and the boundary conditions, the optimal input-output curve is given by

$$\tilde{Y} = \frac{\alpha_2 + \alpha_3 \pm \sqrt{(\alpha_2 + \alpha_3)^2 + 4\alpha_2\alpha_4 \sin \left[ \sqrt{\alpha_2 \left( C + \frac{2\sqrt{Qc}}{\sqrt{m^*}} \right) } \right]}}{2\alpha_2}. \quad (2.19)$$

By implementing numerically the adapted algorithm presented in Sec. 2.4 and assuming mutual information as measure of fitness, the system converges to Eq. (2.19) after a few iterative cycles and the mutual information increases with each step (Fig. 2.4). The comparison between the optimal curve, the experimental results and the MWC model is show in Fig. 2.5. To capture at least the cooperative aspect of the experimental curves, Eq. (2.19) is fitted to an Hill function and the Hill coefficient is considered as the key parameter. The optimal curve selects a Hill coefficient compatible with the experimental
Figure 2.4: **Results of the adaptive algorithm.** (A) Input distributions $p(x)$ (top) and input-output curves $\bar{y}(x)$ (bottom). (Left) Beginning from a Hill function with Hill coefficient 10, the adaptive algorithm (light blue) approaches the analytical result (dark blue). (Right) Comparison between our analytical result (dark blue) and Hill-function constrained input-output curves with adjustable Hill coefficient. (B) The mutual information (between input $x$ and output $y$) as a function of Hill coefficient $n$ for input distributions and input-output curves taken from A. The adaptive algorithm for the unconstrained (dark blue solid line) and Hill-function constrained input-output curves (red solid line) decreases the Hill coefficient (starting in 10 and ending around 2), while the mutual information always increases. For constrained input-output curves, the highest mutual information at a given Hill coefficient is shown by the red dashed line. Noise parameters: $\alpha_1 = 10^{-7}$ and $\alpha_2 = 10^{-8}$. 


2.5. *Escherichia coli* Chemotaxis Responses and Information Processing

measurements from FRET data (Table 2.2), increasing the mutual information by about 1.5 (Table 2.3). However, Eq. (2.19) is unable to capture the sigmoidal shape of the experimental dose-response curves (Fig. 2.5). One of the main reasons for the mismatch, is that the system in Eq. (2.13) is solved only inside the sensitive regime. In other words, the model assumes that the receptors are either fully sensitive or fully insensitive, whereas the transition in the real system is smooth, which might lead to smoother curves similar to a Hill function. Moreover, *E. coli* needs to account for many other constraints and performs other biologically relevant tasks while climbing a gradient of aspartate, thus its sensory system might be suboptimal in transmitting information about a single chemical attractant. However, the optimal solution in Eq. (2.19) comes from a biologically relevant type of noise when maximising the mutual information with respect to both the input distribution and the input-output curve.

<table>
<thead>
<tr>
<th><em>cheRcheB</em> mutant</th>
<th>MWC model</th>
<th>optimal</th>
<th>optimal Hill function</th>
</tr>
</thead>
<tbody>
<tr>
<td>QEQE</td>
<td>6.0 (5.5 - 6.6)</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>QEQQ</td>
<td>8.6 (5.1 - 12.1)</td>
<td>5.0</td>
<td>4.9</td>
</tr>
<tr>
<td>QQQQ</td>
<td>7.2 (4.6 - 9.8)</td>
<td>5.8</td>
<td>5.9</td>
</tr>
</tbody>
</table>

Table 2.2: The Hill coefficients of the input-output curves for different models. For different mutants (first column - QEQE, QEQQ, and QQQQ second, third and last row respectively), the Hill coefficient fitted to the MWC model, with the variability given by the variability in previous fit to MWC model (second column), is compared with the Hill coefficient fitted to the optimal model of Eq. (2.19) (third column), and with the Hill coefficient of the Hill function which has highest mutual information (fourth column). Parameters of the MWC are taken from *Endres et al., 2008*. Parameters for Eq. (2.19): $\alpha_1 = 10^{-7}, \alpha_2 = 10^{-8}, \alpha_3 = 10^{-7}$, and $\alpha_4 = 10^{-8}$. The sensitive regime is taken from Tab. 2.1.
Figure 2.5: Comparison of optimal input-output curves with experimental data from [Clausznitzer et al., 2014]. (A) Receptor activity as function of MeAsp concentration for three different methylation levels of non-adapting mutants: QEQE (blue), QEQQ (red) and QQQQ (green). Comparison of the established MWC model for E. coli chemoreceptors (solid line) [Clausznitzer et al., 2014, Keymer et al., 2006], optimal input-output curves from Eq. (2.19) (dashed line), and FRET data for low-expression of Tar (1.4× native, from [Clausznitzer et al., 2014]) (discs). Parameters: $\alpha_1 = 10^{-7}$, $\alpha_2 = 10^{-8}$, $\alpha_3 = 10^{-7}$, and $\alpha_4 = 10^{-8}$. (B) Comparison of the MWC model (solid line) and optimal input-output curves for different noise parameters (dashed and dotted lines). Parameters: $\alpha_1 = 10^{-7}$, $\alpha_2 = 10^{-8}$, $\alpha_3 = 10^{-7}$ and $\alpha_4 = 10^{-6}$ (dark dashed), $\alpha_4 = 10^{-10}$ (dark dotted) $\alpha_4 = 10^{-8}$ (light dashed). (C) Comparison of the MWC model for E. coli chemoreceptors (solid line), optimal input-output curve (dashed lines), and the Hill-function constrained optimised input-output curves (dotted lines). Parameters: $\alpha_1 = 10^{-7}$, $\alpha_2 = 10^{-8}$, $\alpha_3 = 10^{-7}$, and $\alpha_4 = 10^{-8}$. 

[72x799]2.5. ESCHERICHIA COLI CHEMOTAXIS RESPONSES AND INFORMATION PROCESSING

QEQE
QEQQ
QQQQ
MWC [Monod et al. 1965]
Eq. (2.19)
Exp. Data [Clausznitzer et al. 2014]
c [mM]
0.01 0.05 0.1 1.0
Activity
0.0 0.2 0.4 0.6 0.8 1.0
0.01 0.05 0.10 0.50 1
0.0
0.2
0.4
0.6
0.8
1.0
0.01 0.05 0.10 0.50 1
0.0
0.2
0.4
0.6
0.8
1.0
2.5. *Escherichia coli* chemotaxis responses and information processing

<table>
<thead>
<tr>
<th>Fold increase in $I$ / Mutant</th>
<th>optimal / MWC</th>
<th>optimal Hill function / MWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>QEQE</td>
<td>1.54</td>
<td>1.52</td>
</tr>
<tr>
<td>QEQQ</td>
<td>1.48</td>
<td>1.46</td>
</tr>
<tr>
<td>QQQQ</td>
<td>1.46</td>
<td>1.45</td>
</tr>
</tbody>
</table>

Table 2.3: The fold increase in the transmitted mutual information. For different mutants (first column - QEQE, QEQQ, and QQQQ second, third and last row respectively), the fold increase in mutual information $I$, of Eq. (2.19), relative to the MWC model (second column) is compared with the fold increase in $I$ of Hill function constraint model relative to the MWC model (third column). Parameters of the MWC are taken from [Endres et al., 2008]. Parameters for Eq. (2.19): $\alpha_1 = 10^{-7}$, $\alpha_2 = 10^{-8}$, $\alpha_3 = 10^{-7}$, and $\alpha_4 = 10^{-8}$. The sensitive regime is taken from Tab. 2.1.

Inverse problem: internal noise expression to generate a Hill function as optimal solution

Here, I solve the inverse problem. Given the Berg-and-Purcell input noise $\sigma_X = \sqrt{\alpha_1 c}$, and an Hill function $\bar{Y}(c) = k^n/(c^n + k^n)$ as optimal input-output solution of the novel maximisation introduced in Sec. 2.3, how should the output (or internal) noise be? To answer this question, I need to solve the second equation in system of Eq. (2.16), which becomes

$$Q\sigma_Y^2 = \sigma_1 c \left( -\frac{k^n c^{n-1}}{(k^n + c^n)^2} \right)^2.$$  

Thus, the output noise as a function of the input $c$ is

$$\sigma_Y^2 = \frac{\sigma_1 k^{2n} c^{2n-1}}{Q (k^n + c^n)^4}. \hspace{1cm} (2.23)$$
However, the output noise can be written as a function of the average input-output curve \( \bar{Y} \). Hence, Eq. (2.23) can be rewritten as

\[
\sigma^2_Y = \frac{\sigma_1 n^2 \bar{Y}^4}{Q k} \left( \frac{1 - \bar{Y}}{\bar{Y}} \right)^{\frac{2n-1}{n}}.
\] (2.24)

See Fig. 2.2 for the comparison between the total noise for the different types of noise presented in this chapter.

### 2.5.2 Information transmission along the chemotaxis pathway

At the receptor level, the chemotactic response seems to be designed to maximally transmit information. However, is the whole chemotaxis pathway able to transmit maximal information despite the ultrasensitive response of the motors? To answer this question, a simple model of two concatenated Shannon channels is considered (Fig. 2.6A). The first channel is the receptor channel, describing the sensing by the chemoreceptors. This channel has the extracellular concentration of MeAsp \((c)\) as input and the level of CheY\(_p\) \((Y)\) as output. The second channel is the motor channel, describing the response of the rotary motors. This channel has CheY\(_p\) \((Y)\) as input and the clockwise (CW) motor bias \((t)\) as output. Note that while CheY\(_p\) is both the output of the receptor channel and the input of the motor channel, experimentally CheY\(_p\) is measured in different ways. In particular, at the receptor channel, CheY\(_p\) is measured via FRET, i.e. as ratio of CheY\(_p\)/CheZ, [Sourjik and Berg, 2002a], while for the motor response CheY\(_p\) is measured in mutants without CheZ [Cluzel et al., 2000] or inferred using models [Yuan and Berg, 2013]. My model of two concatenate channels assumes constant levels of total CheY and CheZ.

To connect the theory with experiments, and to simplify the model, the input-output
curves are approximated as Hill functions, where only the Hill coefficient is free to vary in order to maximise the mutual information with respect to both the input distribution and the input-output curve (similarly to Tkačik et al., 2009). The Hill function for receptor sensing has a Hill coefficient $n$, while the Hill coefficient of the motor is $m$.

The dissociation constants are fixed in both cases by the experimental values [Keymer et al., 2006, Yuan and Berg, 2013] (see Table 2.1 for the values of the receptor channel).

By constraining the optimisation of the receptors to Hill functions, the optimal Hill coefficient matches the optimal unconstrained optimised function described in Sec. 2.5.1 (Table 2.2 and Fig. 2.5C), with only a minor decrease in the mutual information (Table 2.3 and Fig. 2.4B). The noise is always increasing downstream with contributions from the transmitted input and the output, $\sigma_{\text{rec}}^T = \sqrt{\alpha_1 c Y^2 + \alpha_2 Y (1 - Y) + \alpha_3 Y + \alpha_4}$ and $\sigma_{\text{mot}}^T = \sqrt{(\sigma_{\text{sen}}^T)^2 t^2 + \beta_2 t (1 - t) + \beta_3 t + \beta_4}$ for the receptors and motors, respectively.

The sensitivity region is extracted from experiments as explained in the previous section.

The numerical implementation of the adaptive algorithm suggests that once considered separately, the receptor and the motor channels evolve to an optimal Hill coefficient (Fig. 2.6B, red areas). However, the total mutual information between the external concentration $c$ and the final output $t$, $I[c; t]$, is limited by the minimum between the mutual information at the receptor channel, $I[c; Y]$, and at the motor channel, $I[Y; t]$ for the data processing inequality, $I[c; t] \leq \min\{I[c; Y], I[Y; t]\}$. Thus, one way to possibly increase the mutual information is to maximise the limiting mutual information. Hence, in the following I focus on the maximisation of the two channel separately and discuss over possible bottlenecks.
2.5. *Escherichia coli* Chemotaxis Responses and Information Processing

**Single motor**

Initially, I consider a single motor, i.e. a single output $t$ for the motor channel. The heat maps of the mutual information as a function of the Hill coefficients $n$ and $m$ for both receptor and motor channels are shown in Fig. 2.6B. Since the noise at the motor is higher than the noise at the receptors, the mutual information at the motor channel always limits the total information transmission (Fig. 2.6D). Hence, single-motor cells would optimize the motor rather than the receptor channel to increase the total mutual information. The resulting optimal values for $n$ and $m$ are relatively low ($\approx 6$; red area in right panel of Fig. 2.6B). In particular, the Hill coefficient at the receptors agrees with the experimental values, which ranges from $6-12$ in Tar-only cells [Endres et al., 2008]. However, the optimal value for the motor channel does not explain the experimental ultra-steep motor response with $m \approx 20$. Hence, a single-motor model for *E. coli* chemotaxis does not seem to transmit maximal mutual information.

**Extension to multiple motors**

The model is then extended to $k$ motors, which are seen as multiple measurements of the internal CheY$_p$ concentration. This way, the motor channel has $Y$ as input and $k$ outputs $t_1, \ldots, t_k$, i.e. the tumble bias of each motor. The chain rule of mutual information allows for the calculation of the mutual information at the motor channel, $\mathcal{I}[Y; t_1, \ldots, t_k] = \sum_{i=1}^{k} \mathcal{I}[t_i; Y|t_{i-1}, \ldots, t_1]$ (see Sec. 1.3.1). This mutual information, in general, is difficult to evaluate. However, by making assumptions on the dependence between motors, some interesting considerations emerge.

For instance, the lower limit is the maximum of the mutual information of the single motor channel, $\mathcal{I}[Y; t_1, \ldots, t_k] \geq \max\{\mathcal{I}[Y, t_1]; \ldots; \mathcal{I}[Y, t_k]\}$. In the case of completely coupled
motors the lower limit is achieved and $\mathcal{I}[Y; t_1, \ldots, t_k] = \mathcal{I}[Y; t_i]$, for any $i \in [1, k]$. In contrast, the upper bound for continuous variables is infinity \cite{Cheong2011}. In addition, for $k$ independent motors, the mutual information at the motor channel scales with the number of motors $\mathcal{I}[Y; t_1, \ldots, t_k] = k \mathcal{I}[Y; t_1]$. Hence, the higher the number of motors the higher the mutual information at the motor channel. However, real motors are not independent, they all depend on the common level of CheY. Furthermore, motors show evidence of a partial coupling: once a flagellum rotates clockwise, it has been observed that the other motors have a propensity to rotate clockwise as well. For instance, \textit{E. coli} has 6-8 flagella and a recent observation suggests that the bacterial movement can be explained by a model of 2-3 conditionally independent motors \cite{Mears2014b}.

By assuming conditional independency of two outputs at given input, i.e. $p(t_1, t_2|Y) = p(t_1|Y)p(t_2|Y)$, and that $\mathcal{I}[t_1; Y]$ and $\mathcal{I}[t_2; Y]$ are both positive, Cheong and co-authors showed that the lower limit is not achievable (motors are either coupled or conditionally independent and only in the former case the equality holds), thus $\mathcal{I}[Y; t_1, t_2] > \max\{\mathcal{I}[Y, t_1], \mathcal{I}[Y, t_1]\}$ \cite{Cheong2011}. By generalising their argument to multiple motors, it is easy to show this for $k$ conditional independent motors with $\mathcal{I}[t_i; Y] > 0 \forall i \in [1, k]$, $\mathcal{I}[Y; t_1, \ldots, t_i] > \mathcal{I}[Y; t_1, \ldots, t_{i-1}]$. In conclusion, this property suggests that for conditionally independent identical motors as well, the mutual information at the motor channel increases with the number of motors. Hence, for a sufficiently high number of motors, the mutual information at the motor channel eventually becomes higher than the mutual information at the receptor channel. However, the number of the motor at which this happens is unknown. Furthermore, we cannot identify the Hill coefficient that maximises the information transmission.

To get additional insights into the multiple motors case, I initially consider two motors.
The mutual information at the motor channel is

\[
I[Y; t_1, t_2] = I[t_1; Y] + I[t_2; Y|t_1]
\]

\[
= I[t_1; Y] + \int dY dt_1 dt_2 p(t_1, t_2, Y) \log \frac{p(t_2|Y)}{p(t_2|t_1)}
\]

\[
= H(t_1) - H(t_1|Y) - H(t_2|Y) + H(t_2|t_1)
\]

(2.25)

Eq. (2.25) is easily generalizable to \(k\) identical motors

\[
I[Y; t_1, \ldots, t_k] = I[t_1; Y] - kH(t_1|Y) + \sum_{i=2}^{k} H(t_k|t_{i-1}, \ldots, t_1).
\]

(2.26)

The last terms, i.e. the conditional entropies between motors, need to be measured experimentally or modelled in more detail to investigate whether the chemotaxis pathway is transmitting maximal information, and this will be the object of my future studies.

However, by further assuming that \(p(t_i|t_{i-1}, \ldots, t_1)\) are Gaussians with constant noise \(\sigma_i\), the conditional entropies are simple scaling factors, \(H(t_i|t_{i-1}, \ldots, t_1) = \log(\sqrt{2\pi e}\sigma_i)\). This allows me to discuss the possible optimal Hill coefficients at the motor channel. Fig. 2.6C shows the mutual information of the motor channel for a small (here a single) number of motors (solid red line), an intermediate (here two) number of motors (dashed red line) and a large (here three) number of motors (dotted red line) compared with the mutual information of the receptor channel (solid blue line).

Although still preliminary, the following scenario seems to emerge: for a small number of motors, the cell has high information transmission at the receptors, which is wasted at the motors (cf. red and blue solid lines in Fig. 2.6C). In contrast, for a large number of motors, the cell maximises information transmission at the receptors but large motor numbers will not bring any further improvement (cf. red dotted and blue solid lines
2.6 DISCUSSION

in Fig. 2.6C). However, in the intermediate case, both receptors and motors equally limit the total mutual information (cf. red dashed and blue solid lines in Fig. 2.6C). To avoid bottlenecks and to optimally allocate resources, the latter case is the most advantageous [Govern and ten Wolde, 2014b]. Hence, the ultra-steep Hill function of the motor ($m \approx 20$) can be explained by this intermediate case (blue arrow in Fig. 2.6D). Note that in addition to the motors high Hill coefficient $m$, there is also a corresponding low $m$ solution (green arrow in Fig. 2.6). However, the latter is not robust to changes in $m$ (a small change in $m$ leads to a drastic reduction of information transmission). This can emerge from varying the number of FliM molecules of the motor [Yuan et al., 2012].

2.6 Discussion

This chapter presented a new approach to maximise the mutual information, which is particularly suited to systems in which both the environmental and internal responses can change. Previously, the focus of most papers was either on the channel capacity, i.e. the maximisation with respect to the input distribution (thus neglecting the changes to the responses, e.g., due to mutation) [Brunel and Nadal, 1998, Tkacik et al., 2008b, Tkacik et al., 2008a], or on the maximisation with respect to the input-output curve (thus neglecting the environmental changes) [Detwiler et al., 2000, Clausznitzer et al., 2014]. The only paper (to my knowledge) that performed the joint maximisation of the input distribution and input-output curve was done assuming that the curve is a Hill function and then optimising the Hill coefficient [Tkacik et al., 2009]. Here, I went beyond the state of the art by optimizing the input-output curves and by proposing analytical solutions. Furthermore, the new approach captures the feedback between mutation and environmental changes, as shown by the evolutionary algorithm.

77
2.6. DISCUSSION

Figure 2.6: **Optimal instantaneous information transmission in E. coli chemotaxis pathway.** (A) A minimal model to describe the E. coli chemotaxis pathway, featuring two concatenated instantaneous Shannon channels. For the receptor channel, the input, $c$, is the extracellular concentration of stimulus and the output, $Y_p$, is the CheY concentration. For the motility channel, the input is $Y_p$, and the output, $t$, is the tumble bias. At the receptor and motor channels, the input-output curves are Hill functions with adjustable Hill coefficients $n$ and $m$, respectively. (B) Mutual-information heat maps as function of Hill coefficients $n$ and $m$ for the receptor $I[c;Y]$ (left) and motor $I[Y; t]$ channels (right), for a single motor. (C) Mutual information as function of Hill coefficient $m$ at $n = 6$. For a single motor, the motor channel’s mutual information (solid red line) is lower than that of the receptor channel (solid blue line). However, for an increasing number of motors (see inset), the mutual information of the motor channel increases (here, for two motors (dashed red line) and three motors (dotted red line)). Predicted $m$ values are indicated by arrows: the optimal $m$ value is indicated by a red arrow for a single motor, and for two motors the equal limiting $m$’s are indicated by blue and green arrows. (D) Heat maps of the mutual information of the sensing channel, $I[c;Y]$, (left), and the motility channel, $I[Y; t]$, for one (middle) and two (right) motors. Parameters: $\alpha_1 = 10^{-7}$, $\alpha_2 = 10^{-8}$, $\alpha_3 = 10^{-8}$, $\alpha_4 = 10^{-5}$, $\beta_2 = 3 \cdot 10^{-4}$, $\beta_3 = 3 \cdot 10^{-4}$, and $\beta_4 = 3 \cdot 10^{-4}$. 

78
This approach is general with a vast range of applications. Here, I used it to predict the optimal input distribution and the optimal input-output curves of the entire *E. coli* chemotaxis pathway. By comparing the optimal and the experimental input-output curves, the idea that such a pathway allows high information transmission is confirmed. In particular, the optimal Hill coefficient selected at the receptor channel agrees with experiments [Sourjik and Berg, 2002a, Endres et al., 2008]. Furthermore, the ultra-sensitive motor response might not restrict information transmission since multiple conditional independent motors can boost the information transmission at the motility channel. It would be interesting to measure the motor response of the uni-flagellated bacterium *Rhodobacter sphaeroides*, which has three different CheY’s [Sourjik and Armitage, 2010]. *R. sphaeroides* can be modelled in a way similar to the one presented in this chapter [Sourjik and Armitage, 2010]. For the symmetry of mutual information, my approach predicts that having three inputs and one output at the motor channel *R. sphaeroides* can sustain an ultra-sensitive motor and maximal information transmission. Finally, this model makes the interesting prediction that single motor bacteria with a chemotaxis pathway similar to *E. coli* should prefer a relatively low Hill coefficient at the motor to transmit high information. Uni-flagellated *Pseudomonas aeruginosa* or monotrichous marine bacteria might allow to test this prediction [Stocker, 2011].

Although the model presented here can be applied to many different systems, it makes a number of assumptions. Firstly, the solution depends on the noise. Different types of noise with biological meanings have been presented. However, all types of noise are based on theoretical assumptions and specifically designed experiments need to be performed to validate the model and extract the parameters. Secondly, the maximisation approach is valid inside a sensitivity regime, which here is extracted from experiments and con-
sidered fixed. In real systems, evolution can act on the sensitive regime and change it. Furthermore, real sensory systems have a continuous drop of sensitivity rather than the abrupt switch from sensitive to insensitive that is used in the model. This latter assumption might be the reason behind the optimal input-output curves not being sigmoidal. Thirdly, any cell-to-cell variability is neglected by the model, since it is assumed that all cells are perceiving the same input distribution. Cell-to-cell variability leads to advantages in terms of bet-hedging strategies, however it is not directly related to information processing. Finally, the time dependencies and histories are neglected, focusing on the instantaneous information transmission only.

In conclusion, this chapter provides a biologically inspired adaptive algorithm for solving complex problems in information transmission in sensory systems. It shows that the *E. coli* chemotaxis pathway is transmitting maximal information at the instantaneous level. However, time-dependencies, including adaptation both at the receptors and motors, need to be accounted for and the high information transmission might be associated with certain behaviours. The next chapter focuses to the link between behaviour and transmission of information.
3. Linking mutual information with chemotactic performance

3.1 Overview: reversing the problem of sensing

Bacteria live in complex, experimentally inaccessible, microenvironments in which the spatiotemporal concentrations of chemicals experienced by the cells are unknown (and often experimentally inaccessible). Can we use our accumulated knowledge about the *E. coli* sensory systems to predict the typical stimuli encountered by wild-type cells? The sensory system has evolved to detect stimuli important to the organism, assuming it performs better under these relevant conditions. Knowing the functioning and design of this sensory system, one should be able to predict typical stimuli, such as concentrations and gradients. Here, using experimental data, information theory, and simulations, some insights into the bacterial microenvironments are achieved. Firstly, using the experimental data from my collaborators, Dr. Silke Neumann and Prof. Victor Sourjik, and the instantaneous mutual information of the sensory system, the typical concentrations of chemicals are identified. Secondly, using simulations of swimming bacteria, the typical gradients are also selected. Interestingly, such gradients are also the gradients in which the bacteria have higher chemotactic drift in the direction of the attractant. This ob-
3.1. OVERVIEW: REVERSING THE PROBLEM OF SENSING

Figure 3.1: The Escherichia coli sensory system is adapted to its chemical environment by evolution. Complex interactions between bacteria in a cellular community, the chemical microenvironment and organisms behaviours (top) shape the typical ligand input distribution experienced by E. coli (left). The bacteria sensory system produces an intracellular response, e.g. the receptor activity, which is the intracellular instantaneous output (bottom). The intracellular output results in the final behaviours of the swimming bacteria (right). Thus, the output behaviours feed back into the environment. Evolution is expected to have selected the optimal shape of the input-output (dose-response) curve to allow for an appropriate response to typical stimuli. Adapted from Micali and Endres, 2016.

observation provides, for the first time, a link between the high information at the sensory system and the final behavioural performance of the cells. The results of this chapter have been published Clausznitzer et al., 2014.
3.2 Identifying the distribution of attractant with maximal information transmission

Bacteria live in large cellular communities in which they interact with bacteria of the same and other species (Fig. 3.1, top left). The chemical concentrations are heterogeneous in space and continuously changing in time, shaped by nutrient supply and depletion (Fig. 3.1, top center). Swimming behaviour allows *E. coli* to follow gradients of chemicals, thus shaping the experienced environmental conditions (Fig. 3.1, top right). Although the ligand distribution (input) experienced by wild-type bacteria is unknown, the sensory system processes these inputs and produces outputs, i.e. behaviours which then feed back into the environment (Fig. 3.1). Therefore, given the receptor dose-responses, i.e. the input-output curves and the noise, the problem can be reversed and it might be possible to infer the typical distribution of input.

Given the mutual information as a measure of fitness [Taylor et al., 2007], i.e. assuming that the receptors have evolved to optimally transmit information, once the dose-response curves are measured the optimal input and output distributions can be predicted. In particular, here it is assumed that the system maximises the mutual information with respect to the input-output curve at fixed input distribution (Eq. (2.4)). Once the dose-response curves are measured, this equation is reversed and solved to provide the predicted input distribution (see Chapter 2 for additional discussion).

The multiple dose-response measurements presented here were performed by my experimental collaborators using FRET on a population of cells expressing only the Tar receptors [Clausznitzer et al., 2014] (Fig. 3.2A). The input is the concentration of MeAsp, c, and the output is the FRET intensity, A. Data in Fig. 3.2A shows large variability in
3.2. IDENTIFYING THE DISTRIBUTION OF ATTRACTANT WITH MAXIMAL INFORMATION TRANSMISSION

the measurements. However, it was found that this variability could be traced back to plasmid copy-number variation, causing a noisy expression in receptor expression which varied day to day [Endres et al., 2008]. To remove this variation from the variance, principal component analysis was performed and the first component associated with the variation in amplitude was removed [Endres et al., 2008, Clausznitzer et al., 2014] (Fig. 3.2B, left column). Furthermore, this principal component decomposition allows a meaningful fit of the experimental data with the MWC model, and hence, to extract the average input-output curves [Clausznitzer et al., 2014]. Note that the single-cell curves are assumed to be the average curves from the populations.

The single-cell noise can be extracted from the population measurements by making some assumptions. First, the noise is assumed to have two independent components, the transmitted input noise, and the output noise. Second, to extract input and output noises for single cells from the variance of the total activity from approximately $n \approx 400$ cells, the input noise is assumed to arise from coherent addition of input fluctuations and hence scales with $n^2$. In contrast, the output noise results from the incoherent addition of cell-internal fluctuations and hence scales with $n$ [Clausznitzer et al., 2014]. Specifically, the input noise is assumed to arise from fluctuations of the ligand concentration in the flow chamber, and hence applies to all the cells equally. To be concrete, these fluctuations are chosen to be Poissonian with the variance proportional to the mean (formally similar to the Berg-and-Purcell limit [Berg and Purcell, 1977]). In contrast, the output noise has two contributions: a Binomial switching contribution due to the two states of each receptor cluster and a Poissonian contribution (see also Sec. 2.3 for additional discussion.
3.2. IDENTIFYING THE DISTRIBUTION OF ATTRACTANT WITH MAXIMAL INFORMATION TRANSMISSION

of the noise terms). Thus, the noise reads,

\[
\sigma_{n,Tot}^2 = n^2 \sigma_c^2 \left( \frac{\partial \bar{A}}{\partial c} \right)^2 + n \sigma_A^2
\]

\[= n^2 \alpha_1 c \left( \frac{\partial \bar{A}}{\partial c} \right)^2 + n \left[ \alpha_2 \bar{A}(1 - \bar{A}) + \alpha_3 \bar{A} \right].\] (3.1)

Where \(\sigma_c^2\) and \(\sigma_A^2\) are the input and output noise respectively, and \(\alpha_{1-3}\) are parameters.

The experimental noise, after the removal of the principal component, is fitted with this equation in order to extract the fitting parameters \(\alpha_{1-3}\) (Fig. 3.2B, left column, and Table 3.1) [Clausznitzer et al., 2014].

Equipped with average single-cell input-output curves and noise and assuming a small Gaussian channel (see Eq. (1.9)), the input distributions which guarantee that the measured average single-cell input-output curves are the curves that maximise the mutual information between \(c\) and \(A\) at fixed input distribution are

\[
p(c) = \frac{\sigma_{1,Tot}^2(c) \partial \bar{A}}{\sigma_A^2(c) Z \frac{\partial \bar{A}}{\partial c}}.\] (3.3)

Where \(Z = \int dc \frac{\partial \bar{A}}{\partial c} \sigma_{1,Tot}^2(c) \frac{\partial \bar{A}}{\partial c} \) is the normalization function. Once solved for \(p(c)\), Eq. (3.3) provides the best transmitted distributions of concentration which are shown in Fig. 3.2B, middle column. The output distributions, \(p(A)\), can be evaluated using \(p(A) = p(c)/\bar{A}(c)\) valid in small noise approximation (Fig. 3.2B, right column), or in alternative using \(p(A) = \int p(A|c)p(c)\) valid in any regime which provide a similar solution (data not shown). The input distributions can be fitted by log-normal distributions with a large standard deviation compare to the uniform noise prediction [Laughlin, 1981] [Clausznitzer et al., 2014] (Fig. 3.2C, fitting parameters are listed in Table 3.2). However, the best transmitted output distributions are bimodal, with large deviation from the uniform
3.2. IDENTIFYING THE DISTRIBUTION OF ATTRACTANT WITH MAXIMAL INFORMATION TRANSMISSION

Figure 3.2: Predicted distributions of chemical inputs and intracellular outputs. (A) Experimental dose-response curves at fixed methylation level for adapting wild-type and non-adapting mutants. Data points for various measurements of the FRET activity at different concentrations of MeAsp (filled symbols), as well as their mean values are shown (open symbols). Wild-type adapted cells at 0.1 mM MeAsp (black triangles) as background concentration, QEQE (green), QEQQ (blue) and QQQQ (orange) mutants. Corresponding solid lines are the fits of the MWC model [Endres et al., 2008]. Data normalized to maximal activity. (B, left) The standard deviation (SD) extracted from the FRET measurements in panel A. Dots are data: open symbol for bare data and filled symbols for data excluding receptor expression noise. Lines are fits to the standard deviation: dashed and solid for fitting bare and excluding receptor noise data, respectively. (Middle) Predicted distributions of attractant concentrations (input). (Right) Predicted distributions of signalling activities (output). (C) Log-normal scaling of the predicted input concentrations. All curves are normalized with area of one. Symbols are the predicted input distributions panel B (QEQE, green; WT 2 (0.1 mM), black; QEQQ, blue; QQQQ, orange) and coloured lines are log-normal fits to the distributions. See Tables 3.1 and 3.2 for fitting parameters. Adapted from Clausznitzer et al., 2014.
### 3.2. IDENTIFYING THE DISTRIBUTION OF ATTRACTANT WITH MAXIMAL INFORMATION TRANSMISSION

<table>
<thead>
<tr>
<th></th>
<th>( \alpha_1 ) (confidence interval)</th>
<th>( \alpha_2 ) (confidence interval)</th>
<th>( \alpha_3 ) (confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WT</strong></td>
<td>( 3.910 \times 10^{-14} )</td>
<td>( 3.344 \times 10^{-8} )</td>
<td>( 2.165 \times 10^{-9} )</td>
</tr>
<tr>
<td></td>
<td>( 0.2398 \times 10^{-13} )</td>
<td>( 1.927 \times 10^{-8}, 4.731 \times 10^{-8} )</td>
<td>( 0.5989 \times 10^{-9} )</td>
</tr>
<tr>
<td><strong>QEQQ</strong></td>
<td>( 8.197 \times 10^{-13} )</td>
<td>( 1.565 \times 10^{-7} )</td>
<td>( 3.923 \times 10^{-9} )</td>
</tr>
<tr>
<td></td>
<td>( 0.4521 \times 10^{-13} )</td>
<td>( 1.200 \times 10^{-7}, 1.928 \times 10^{-7} )</td>
<td>( 0.8051 \times 10^{-9} )</td>
</tr>
<tr>
<td><strong>QQQQ</strong></td>
<td>( 3.494 \times 10^{-12} )</td>
<td>( 1.298 \times 10^{-6} )</td>
<td>( 1.400 \times 10^{-8} )</td>
</tr>
<tr>
<td></td>
<td>( 0.9358 \times 10^{-12} )</td>
<td>( 1.061 \times 10^{-6}, 1.521 \times 10^{-6} )</td>
<td>( 0.4910 \times 10^{-8} )</td>
</tr>
<tr>
<td><strong>WT</strong></td>
<td>( 1.528 \times 10^{-7} )</td>
<td>( 8.241 \times 10^{-8} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( 0.1074 \times 10^{-11} )</td>
<td>( 0.3208 \times 10^{-7} )</td>
<td>( 5.620 \times 10^{-8}, 1.081 \times 10^{-7} )</td>
</tr>
</tbody>
</table>

**Table 3.1: Fit parameters of the variance in FRET activity.** The variance in FRET data fitted to the model \( \sigma_T^2 = n^2 \alpha_1 c \left( \frac{dA}{dx} \right)^2 + n \left[ \alpha_2 \dot{A} + \alpha_3 \dot{A}(1 - \dot{A}) \right] \). (Top) Excluding gene-expression noise [Endres et al., 2008]. (Bottom) Including gene-expression noise. Fitted values and, below in brackets, the corresponding 95% confidence intervals calculated using the profile-likelihood approach. See Supporting Information of [Clausznitzer et al., 2014] for additional details, work done by Diana Clausznitzer.

Distributions predicted for uniform noise [Laughlin, 1981], but similar to other biological systems [Tkacik et al., 2008b] (Fig. 3.2B, right column). Under the assumption that the system is highly transmitting common inputs, the optimal distributions identified in Fig. 3.2B are the predicted distribution of concentrations encountered by the cells during chemotaxis.

---

87
3.3 IDENTIFYING THE GRADIENTS OF ATTRACTANT WITH HIGH INFORMATION TRANSMISSION

Table 3.2: Log-normal fits to predicted input distributions. Estimated parameter values for the fit of predicted input distribution to log-normal distributions

$$p(c) = \frac{1}{c\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log c - \mu)^2}{2\sigma^2}\right)$$

in Fig. 3.2. Fitted values and, below in brackets, the corresponding 95% confidence intervals. (Top) Predictions based on variances including gene-expression noise. (Middle) Predictions based on variances excluding gene-expression noise, work done by Diana Clausznitzer [Clausznitzer et al., 2014].

<table>
<thead>
<tr>
<th>Strain</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Including gene-expression noise</td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>-2.3975</td>
<td>0.2479</td>
</tr>
<tr>
<td></td>
<td>(-2.3985, -2.3964)</td>
<td>(0.2471, 0.2488)</td>
</tr>
<tr>
<td>QEQE</td>
<td>-2.7727</td>
<td>0.2695</td>
</tr>
<tr>
<td></td>
<td>(-2.7736, -2.7717)</td>
<td>(0.2687, 0.2703)</td>
</tr>
<tr>
<td>QEQQ</td>
<td>-1.5675</td>
<td>0.2242</td>
</tr>
<tr>
<td></td>
<td>(-1.5684, -1.5666)</td>
<td>(0.2234, 0.2249)</td>
</tr>
<tr>
<td>QQQQ</td>
<td>-0.4678</td>
<td>0.2492</td>
</tr>
<tr>
<td></td>
<td>(-0.4696, -0.4660)</td>
<td>(0.2477, 0.2507)</td>
</tr>
<tr>
<td></td>
<td>Excluding gene-expression noise</td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>-2.3222</td>
<td>0.2283</td>
</tr>
<tr>
<td></td>
<td>(-2.3234, -2.3209)</td>
<td>(0.2272, 0.2293)</td>
</tr>
<tr>
<td>QEQE</td>
<td>-2.7136</td>
<td>0.2387</td>
</tr>
<tr>
<td></td>
<td>(-2.7148, -2.7124)</td>
<td>(0.2377, 0.2396)</td>
</tr>
<tr>
<td>QEQQ</td>
<td>-1.5350</td>
<td>0.2022</td>
</tr>
<tr>
<td></td>
<td>(-1.5361, -1.5338)</td>
<td>(0.2012, 0.2031)</td>
</tr>
<tr>
<td>QQQQ</td>
<td>-0.5037</td>
<td>0.2560</td>
</tr>
<tr>
<td></td>
<td>(-0.5053, -0.5020)</td>
<td>(0.2546, 0.2574)</td>
</tr>
</tbody>
</table>

3.3 Identifying the gradients of attractant with high information transmission

In the previous section, the distributions of MeAsp which are optimally transmitted by the E. coli sensory system at fixed methylation level have been identified. Here, I identify the linear gradients which produce distributions of concentration sampled by swimming bacteria similar to the optimal distributions of Sec. 3.2. The idea is to simulate identical cells swimming in different linear gradients of MeAsp along the $x$ direction, then...
to collect the distribution of sampled concentration at fixed methylation level, and finally to compare the sampled distributions, denoted \( p(c|m^*) \) (see red trajectories in Fig. 3.3B), with the optimal ones, here denoted with \( q(c|m^*) \). The gradients in which the sampled distributions closely match the optimal distribution of concentration are the gradients which transmit information best.

For this purpose, I simulated swimming bacteria using a modified version of the RapidCell software [Vladimirov et al., 2008] (see Appendix A). Briefly, cells are able to sense and adapt, to run and tumble based on cell-internal signalling, and are affected by rotational diffusion. Note that simulations are necessary, since both the external concentration \( c \) sampled and the internal methylation level \( m \) need to be extracted at the same time from individual trajectories of swimming bacteria, data that are experimentally inaccessible.

At constant extracellular concentration, \( c^* \), in absence of noise, the cells are in the adapted state, \( A^* \), and they are in a unique methylation level \( m^* \) (Fig. 3.3A). To experience different methylation levels, bacteria need to navigate in gradients of nutrient and to experience some noise. Thus, focusing on a predicted input distribution at fixed methylation level \( m^* \) and Tar-only cells, the gradients are calibrated to span concentrations around \( c^* \). The linear gradients are designed to range from very shallow to very steep (here from 0.1 to 1.75 in unit of \( 1/c^* \)). As shown in Fig. 3.3B, for one exemplar linear gradient, 100 s-long trajectories of about 1000 identical cells are simulated for different linear gradients. In a very shallow chemical gradient, cells usually sample a symmetrical range of concentrations around \( c^* \) (see Fig. 3.3C, top panel). When the gradient gets steeper, cells move up the gradient. As a result, cells sample a broader range of concentrations up the gradient than down the gradient, producing an asym-
3.3. IDENTIFYING THE GRADIENTS OF ATTRACTANT WITH HIGH INFORMATION TRANSMISSION

metric distribution of sampled concentrations (see Fig. 3.3C, middle panel). However, when the gradient is very steep, the sampled distribution $p(c|m^*)$ differs from the optimal input distribution $q(c|m^*)$ (see Fig. 3.3C, bottom panel). The gradients which produce a large overlap between $p(c|m^*)$ and $q(c|m^*)$ have been selected as the optimal linear gradients (Fig. 3.4B) where the overlap is calculated using the overlapping coefficient $OVL = \int \min\{p(c|m^*); q(c|m^*)\}$ [Inman and Bradley Jr, 1989].

Linear gradients might be rare in natural environments. In fact, free diffusion of a fixed number of molecules produces a Gaussian gradient [Berg, 1993], diffusion from a point source with constant rate of production results in a hyperbolic gradient [Berg, 1993], or in an exponential gradient if there is homogeneous degradation [Ibanes and Belmonte, 2008]. However, a linear gradient emerges when there is diffusion between points with constant rates of production and absorption [van Haastert and Postma, 2007, Vladimirov et al., 2008]. Furthermore, any type of gradient is linear in a first approximation.

Among all the gradient types, the Weber-Fechner law suggests that exponential gradients, $c(x) = A\lambda^x$, are better sensed by *E. coli*, since the extracellular concentrations are perceived as constant in the receptors sensitive regime (logarithm sensing) [Lazova et al., 2011, Clausznitzer et al., 2014]. Independently of the methylation level of the receptors (and hence of the mutant type), the cells’ preference to swim in a particular exponential gradient with fixed relative gradient $\lambda = c^{-1}\frac{dc}{dx}$ would result in a consistent overlap between the sample and the optimal input distribution. In other words, it is expected that if the overlap is plotted against the relative gradients for my simulations in linear gradients, the curves at different methylation levels should collapse, which is approximately the case (see Fig. 3.4B). The overlap allows for the selection of the optimal $\lambda$. Here, for instance, the relative gradients that achieve an overlap within 20% of maximal overlap
3.4 HIGH INFORMATION TRANSMISSION LEADS TO HIGH DRIFT VELOCITY

Figure 3.3: Caption next page.

have been selected. The distribution of exponential gradients predicted from this selection is remarkably broad (wide range of rate values in Fig. 3.4D), therefore evidencing the sensory systems’ wide dynamic range.

3.4 High information transmission leads to high drift velocity

Until this section, this chapter selected the concentrations and gradients which are best transmitted, in terms of information theory, at the receptor level. If cell sensory system has evolved to best transmit common inputs, these optimal concentrations and gradients
3.4. HIGH INFORMATION TRANSMISSION LEADS TO HIGH DRIFT VELOCITY

Figure 3.3: **Simulations of swimming bacteria in chemical gradients.** (A) Strategy to extract sampled distribution of inputs from simulations. In absence of any gradient and noise, the adapted activity $A^*$ determines the modification level $m^*$ at concentration $c^*$. In the presence of a gradient and noise, cells swim and modify their receptor activity, the methylation level, in accordance with the sample concentration. In typical gradients, cells should sample the distribution of concentrations at fixed methylation level, similarly to what is predicted by information theory (light blue curve). (B) Typical simulation of cells swimming in a rectangular box with linear gradient in $x$ (here, Tar-only cells, relative gradient $0.1 \text{ mm}^{-1}$, $m^* = 6$ for QEQQ mutants). Blue lines correspond to the full trajectories, red lines to partial trajectories in which the receptor methylation level is $m = m^*$. The initial positions are identified by green dots. Note that the red trajectories are in correspondence of $x^* = x^{-1}(c^*)$. (C) Using $m^* = 4$ (QEQQ) as an example, predicted distribution of inputs from the information theory, $q(c|m^* = 4)$ (black lines) and distributions of sampled concentrations, $p(c|m^* = 4)$, (blue lines) obtained for cells swimming in increasing relative linear gradients, $0.1 \text{ mm}^{-1}$ (top), $0.5 \text{ mm}^{-1}$ (middle), and $1.5 \text{ mm}^{-1}$ (bottom) (gradients relative to $c^* = 0.065 \text{ mM}$). To imitate cell-external noise, the base concentration of the gradients was fluctuating every $0.1 \text{s}$ with standard deviation $0.001 \text{ mM}$. To imitate cell-internal noise, modification level was selected from normal distribution with relative standard deviation $\sigma = 0.04$ in line with previous results [Clausznitzer and Endres, 2011]. (inset) The chemotaxis index (CI) defined as the ratio between net movement in the direction of the gradient and the total movement for the trajectories at $m = m^*$. Adapted from [Clausznitzer et al., 2014].

are those which are typically encountered by the cells. Furthermore, it is expected that cells should have behavioural advantages under the common external conditions. Hence, to test for behavioural advantages, I study here the chemotaxis performance of cells under the optimally transmitted concentrations and linear gradients identified in Sec. 3.3.

To gain intuition, I measure the chemotaxis performance in terms of the chemotactic index, $\text{CI} = \sum_i \frac{\delta x_i}{\delta s_i}$, and the drift velocity, $v_d = \sum_i \frac{\delta x_i}{\delta t}$, for the simulations of swimming cells (same simulations shown in Sec. 3.3). Here, $x$ is the direction of the gradient, $\delta s_i = \sqrt{\delta x_i^2 + \delta y_i^2}$ is the distance between two points of a simulated trajectory, $\delta x_i$ and $\delta y_i$ are the changes in the $x$ and $y$ directions respectively, and data points are collected
3.4. HIGH INFORMATION TRANSMISSION LEADS TO HIGH DRIFT VELOCITY

Figure 3.4: Caption next page.
3.4. **HIGH INFORMATION TRANSMISSION LEADS TO HIGH DRIFT VELOCITY**

Figure 3.4: **Reconstruction of distributions of sampled gradients.** (A) Chemotactic index (CI) averaged over all simulated trajectories in the box for different linear gradients, as a function of concentration in the box. Linear gradients measured relatively to $c^* = 0.22$ mM: 0.1 mm$^{-1}$ (blue), 0.25 mm$^{-1}$ (red) and 1.0 mm$^{-1}$ (green). The CI trend suggests an optimal range of gradients that maximise CI at fixed concentration $c^*$. Here, for example, at $c^* \approx 0.22$ mM, the optimal relative linear gradient is 0.25 mm$^{-1}$. Symbols and error bars indicate averages and standard errors of the mean from several trajectories at this concentration and gradient. Lines are interpolations by smooth functions. (B) Overlap between distribution of sampled concentrations and predicted distribution (blue shades), chemotactic index (CI, red shades) and drift velocity (green shades) with modification level $m = m^*$. Symbols indicate modification level: squares, circles and triangles stand for $m^* = 4$ (QEQE, $c^* \approx 0.07$ mM), $m^* = 4.6$ (WT, $c^* \approx 0.09$ mM), and $m^* = 6$ (QEQQ, $c^* \approx 0.22$ mM), respectively. $m^* = 8$ (QQQQ, $c^* \approx 0.63$ mM) is not included, as Tar-only cells do not adapt at high values of $c^*$. Horizontal arrow illustrates range of relative gradients over which the overlap is within 20% of maximal value, on average, between the three modification levels. (C) Sampled distributions from different relative gradients (0.1-1.75 mm$^{-1}$) fit prediction with overlap 90%. (D) Range of exponential gradients predicted to be sensed best (blue area), according to the range indicated by horizontal arrow in (B). Adapted from Clausznitzer et al., 2014.

with a time step of $\delta t$. Note that the CI and the drift velocity are both measures of the net movement of the cells up a chemical gradient. However, while the CI is the ratio between net movement and the total movement, the drift velocity is the ratio between the net movement and the total time spent. Hence, there are only minor differences between the CI and the drift velocity in how the tumble events are accounted (see further discussion in Sec. 4.2). In this Chapter, I focus on the CI. However, all the results are also valid for the drift velocity (see Fig. 3.4B).

The CI averaged over many identical cells are complicated, non-monotonic functions of the concentrations and gradients experienced by the cells (see Chapter 4 for more details). For instance, see the quantification of CI for simulated cells in different linear gradients shown in Fig. 3.4A. A few important observations can be made. Firstly, in linear
3.4. HIGH INFORMATION TRANSMISSION LEADS TO HIGH DRIFT VELOCITY

gradients, the CI decreases for high concentrations, since the relative gradient decreases. Secondly, at fixed concentration, there is an intermediate linear gradient which increases the CI compared to shallow gradients, but also compared to very steep gradients. Take for instance the fixed concentration $c^* = 0.22\text{ mM}$ in Fig. 3.4A; the CI at intermediate relative gradient (red line) is higher than both in shallow gradient (blue line) and steep gradient (green line). As already noticed in Sec. 3.3, this is because in a shallow gradient, the CI is low since cells are sensing small changes in extracellular concentrations and thus maintaining a quasi-adapted behaviour, while at steep gradients, the CI is also low since cells show long runs which are affected by rotational diffusion leading them off course. Put otherwise, a cell must balance long runs with a sufficient number of tumble events to correct alignment and efficiently progress up the gradient [Dufour et al., 2014; Clausznitzer et al., 2014; Long et al., 2016]. The generalisation of this argument to any concentration $c^*$ suggests that the CI can be optimised with respect to the gradient at fixed concentration.

This can be proved by focusing on trajectories at fixed methylation level $m^*$, which identify a fixed $c^*$, Fig. 3.3A). Indeed, the CI and the drift velocity for trajectories in which the cells have fixed methylation level $m^*$ show a maximum when plotted as a function of the relative gradients (Fig. 3.4B). Interestingly, the maximum corresponds to the maximal overlap between simulated and predicted distribution of concentrations, and the curves for different methylation levels collapse (Fig. 3.4B). Unlike the overlap based on instantaneous mutual information at the receptors, the drift velocity is the final output of the whole chemotaxis pathway, encoding time dependencies. Furthermore, the drift velocity is the quantity which most likely is under evolutionary pressure [Skoge et al., 2011].

This result reinforces the idea of optimal information transmission along the whole
chemotaxis pathway shown in Chapter 2. Furthermore, it suggests that optimal transmission of information results in higher drift velocity, linking the theory of optimal transmission of information with the behavioural output of chemotaxis. Finally, since the drift is the result of complex histories and time dependencies, this result calls for further investigation between drift and information processing in time.

3.5 Discussion

Bacteria’s immediate natural environment differs from that found in experimentally-controlled conditions. They are dynamic microenvironments, influenced by highly dense cellular communities which are regulated by subtle spatio-temporal gradients of partially digested food and host secretions, and these minute microenvironments are, to this day, inaccessible to experimental sampling. Before this work, it was predicted that exponential gradients are the typical gradients for *E. coli* chemotaxis. This prediction was based on models accounting for Weber’s law, Weber-Fechner law and fold-change detection, hence able to make predictions about average gradients only \cite{Kalinin2009,Lazova2011,Vladimirov2008}. My work uses data from *in vivo* FRET measurements, simulations of swimming bacteria, and information theory to identify the distribution of concentrations and linear gradients which are best sensed by the *E. coli* sensory system \cite{Clausznitzer2014}. These concentrations and gradients might be those typically encountered by the cells in a natural environment, since the sensory system should have evolved to better sense the most common stimuli \cite{Clausznitzer2014}. The predicted distributions of input concentrations are consistent with bacteria in a wide range of exponential gradients. Furthermore, this model is able to connect phenomenological laws, such as Weber’s law, with information theory at the molecular level \cite{Clausznitzer2014}. 

96
The predictions put forward by this approach are consistent with the exponential gradients predicted by phenomenological laws, and go beyond the average gradients. However, this work is based on several assumptions. Firstly, the results are based on Tar-only cells, while wild-type cells have multiple receptor types. Secondly, the attractant used is MeAsp, which mimics the aspartate response but is not metabolisable by cells. Furthermore, to extract single-cell noise from population measurements, a number of assumptions have been made: I assumed small Gaussian noise and independent input and output noise with certain functional forms of the noise (see Eq. (3.1)). Moreover, the average responses from populations are assumed to be the average single-cell responses. In addition, this approach focuses only on instantaneous receptor signalling, excluding slower downstream dynamics, e.g. the receptors adaptation and the motor dynamics [Tostevin and ten Wolde, 2009]. However, as shown in Chapter 2, it is expected that instantaneous information transmission can be optimised by the whole chemotaxis pathway. Additionally, this work focused on chemosensing, neglecting the effects of temperature and pH [Hu and Tu, 2014]. Finally, this approach is further limited by the fact that chemotactic gene expression is regulated in part by life cycle, nutrient supply, and growth conditions, thus limiting the predictive power of this approach [McCarter, 2006, Jordan et al., 2013].

The results presented in this chapter show, for the first time, a link between the maximal instantaneous information transmission at the receptors and the chemotactic performance of the cells (here measured using both drift velocity and chemotactic index). This important result calls for further investigation. Firstly, how do cells process information in time? The mutual information in time needs to account for full histories and
hence for the full trajectories of inputs and outputs, which make calculation extremely
difficult [Tostevin and ten Wolde, 2009]. Thus, new approaches are needed to under-
stand information processing in time. Furthermore, a model to characterise the drift
velocity in any gradient is needed [Dufour et al., 2014]. Experimental tracking of swim-
ming cells in micro-fabricated devices can help to characterise the drift in well-defined
gradients [van Haastert and Postma, 2007, Masson et al., 2012, Colin et al., 2014]. Fi-
nally, the problem can be approached from a thermodynamic point of view, since entropy
dissipation is theoretically linked with mutual information [Parrondo et al., 2015]. The
principle of high-information, high-drift presented here suggests that chemotaxis and in-
fotaxis describe the same process [Vergassola et al., 2007]. The next chapter provides an
analytical model for drift which has been tested with experimental data, and allows for
the identification of gradients and concentrations of high and low drift.
4. Behaviours of individual *Escherichia coli* cells

4.1 Overview

The *E. coli* chemotaxis pathway transmits maximal instantaneous mutual information (Chapter 2), and the instantaneous mutual information at the receptors has been linked with the ability of *E. coli* to swim up linear gradients of nutrients (Chapter 3). Here, I aim to generalise those results two fold, by accounting both for different types of gradients and for the full-history and time dependencies. However, swimming behaviours are quite complicated and difficult to interpret, requiring a more comprehensive theory to summarise behaviour. This is due to two main reasons. Firstly, by exposing a population of genetically identical cells to a gradient, it has been observed that some cells outperform the majority of the other cells in the population in climbing the gradient. This suggests the presence of a high-performance phenotype [Emonet, 2016]. Different gradients select for different behaviours and such biodiversity in genetically identical cells provides advantages for the population under fluctuating environments [Long et al., 2016] [Frankel et al., 2014]. The observed biodiversity has been explained by gene regulation changes which affect the level of the chemotactic proteins [Frankel et al., 2014] [Dufour et al., 2014].
4.2. BEHAVIOUR IN TERMS OF DRIFT VELOCITY

Secondly, *E. coli* cells grow and divide while chemotacting and have different behaviours during their lifetime [Jordan et al., 2013, Leibler, 2016]. Moreover, the mutual information on trajectories is hard to be calculated. Thus, different ways of quantifying the information gain need to be introduced.

Here, by focusing on single-cell trajectories of simulations of a unique, defined, phenotype, I demonstrate that the behaviour summarised in terms of the drift velocity is complicated, even within phenotypically identical cells. Firstly, I present the limits and variations of drift velocity, which are interpreted within an analytical theory. Secondly, to validate these findings, I present the results of experiments conducted in microfluidic chambers, which I designed and were conducted by my experimental collaborators Dr. Remy Colin and Prof. Victor Sourjik at the Max Planck Institute of Terrestrial Microbiology in Marburg, Germany, and for which I analysed the data. In such experiments, single-cell trajectories of swimming bacteria are recorded in carefully designed chemical gradients. Finally, I further investigate the link between the drift velocity and information gain. I propose two quantities, which can be considered alternatives to mutual information, to evaluate the information gain along trajectories of swimming bacteria in exponential gradients. This allows me to go beyond the instantaneous transmission of information studied in Chapters 2-3.

4.2 Behaviour in terms of drift velocity

Here, I use the drift velocity as a quantity to summarise behaviour. The drift velocity is defined as the component of the velocity in the direction of the attractant gradient $\bar{v}_d = \delta x/\delta t \hat{x}$, when the gradient is in the $\hat{x}$ direction. Alternatively, the chemotactic index, $\text{CI} = \delta x/\delta s$ with $\delta s$ the length of a trajectory during a small time $\delta t$, can also be
4.2. BEHAVIOUR IN TERMS OF DRIFT VELOCITY

<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F = N\left(\epsilon(m) + \nu_a \log \frac{1+c/K_{a\text{off}}}{1+c/K_{a\text{on}}} + \nu_s \log \frac{1+c/K_{s\text{off}}}{1+c/K_{s\text{on}}} \right)$</td>
<td>Receptor free-energy: Keymer et al., 2006</td>
</tr>
<tr>
<td>$A = \frac{1}{1+e^F}$</td>
<td>Receptor activity: Keymer et al., 2006</td>
</tr>
<tr>
<td>$\frac{dm}{dt} = g_R(1 - A) - g_B A^3$</td>
<td>Methylation dynamics: Clausznitzer et al., 2010</td>
</tr>
<tr>
<td>$Y_p = \frac{k_y A Y_T}{k_y A + k_z Z + \gamma Y}$</td>
<td>CheY$_p$: Vladimirov et al., 2008</td>
</tr>
<tr>
<td>$\text{CW}_{bias} = \frac{Y_p^m}{Y_p^m + K}$</td>
<td>Motor bias: Yuan and Berg, 2013</td>
</tr>
</tbody>
</table>

Table 4.1: Model for the *E. coli* chemotaxis pathway. Schematic of the equations involved in the *E. coli* chemotaxis pathway with references. The detailed model is presented in Appendix A for simulations and in B for the analytical theory in linear regime.

used and provides very similar results. However, CI does not account for the duration of the tumbles, since the trajectory lengths are the same whether the bacteria stop or not. Furthermore, both in experiments and simulations, a trajectory appears as a collection of points at discrete time steps. Hence, the length of the recorded trajectory $\delta s_{\text{meas}}$ during $\delta t$ is the length of the straight line between subsequent points, while the real trajectory can bend during $\delta t$. Therefore, the length of the recorded trajectory is an estimation of the real trajectory, which depends on $\delta t$. In particular, the recorded trajectory is a lower bound for the real trajectory (since $\delta s_{\text{meas}} \leq \delta s_{\text{real}}$). The recorded trajectory approaches the real trajectories for $\delta t$ small. In contrast, the drift velocity, $\bar{v}_d$, does account for the time of tumbles and does not require the evaluation of the length of the trajectory, which makes the evaluation of the $\bar{v}_d$ easier than that of CI. For these reasons, the drift velocity is preferred to CI to summarise behaviour of run-and-tumble bacteria, and it is the focus of this chapter.
4.2. BEHAVIOUR IN TERMS OF DRIFT VELOCITY

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperative receptor number</td>
<td>$N$ $\in [5; 16]$</td>
<td>Endres et al., 2008</td>
</tr>
<tr>
<td>Methylation free-energy</td>
<td>$\epsilon(m)$ $\in [-2.6; 0.25]$</td>
<td>Endres et al., 2008</td>
</tr>
<tr>
<td>Fraction of Tar receptors</td>
<td>$\nu_a$ $\simeq 1/3$</td>
<td>Keymer et al., 2006</td>
</tr>
<tr>
<td>Fraction of Tsr receptors</td>
<td>$\nu_a$ $\simeq 2/3$</td>
<td>Keymer et al., 2006</td>
</tr>
<tr>
<td>Active receptors dissociation constant Tar</td>
<td>$K^{on}_{Tar}$ 0.5 mM</td>
<td>Keymer et al., 2006</td>
</tr>
<tr>
<td>Active receptors dissociation constant Tsr</td>
<td>$K^{on}_{Tsr}$ 10$^6$ mM</td>
<td>Keymer et al., 2006</td>
</tr>
<tr>
<td>Inactive receptors dissociation constant Tar</td>
<td>$K^{off}_{Tar}$ 0.02 mM</td>
<td>Keymer et al., 2006</td>
</tr>
<tr>
<td>Inactive receptors dissociation constant Tsr</td>
<td>$K^{off}_{Tsr}$ 100 mM</td>
<td>Keymer et al., 2006</td>
</tr>
<tr>
<td>Methylation rate</td>
<td>$g_R$ 0.0069 s$^{-1}$</td>
<td>Clausznitzer et al., 2010</td>
</tr>
<tr>
<td>De-methylation rate</td>
<td>$g_B$ 0.12 s$^{-1}$</td>
<td>Clausznitzer et al., 2010</td>
</tr>
<tr>
<td>Phosphorylation rate of $Y_p$</td>
<td>$k_Y$ 100 $\mu$M$^{-1}$s$^{-1}$</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>De-phosphorylation rate of $Y_p$</td>
<td>$k_Z$ 30 $Z^{-1}$s$^{-1}$</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>Degradation rate of $Y_p$</td>
<td>$\gamma_Y$ 0.1</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>Adapted activity</td>
<td>$A^*$ $\simeq 0.3$</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>Total concentration of CheY</td>
<td>$Y_T$ $\in [6; 9.7]$ $\mu$M</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>Concentration of CheZ</td>
<td>$Z$</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>Motor Hill coefficient</td>
<td>$m$ $\in [15; 22]$</td>
<td>Yuan and Berg, 2013</td>
</tr>
<tr>
<td>Motor dissociation constant</td>
<td>$K$ $\simeq 3$ $\mu$M</td>
<td>Yuan and Berg, 2013</td>
</tr>
<tr>
<td>Rotational diffusion</td>
<td>$D_s$ $\simeq 0.062$ rad/s</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>Adaptation time</td>
<td>$\tau_m$ $\tau_m(g_R, \gamma_B, \gamma_A^*)$</td>
<td>Vladimirov et al., 2008</td>
</tr>
<tr>
<td>Run velocity</td>
<td>$v_0$ $\simeq 12.5 [\mu$M$]$</td>
<td>Measured from data</td>
</tr>
</tbody>
</table>

Table 4.2: Parameters in *E. coli* chemotaxis. Schematic of the parameters involved in chemotaxis with values from literature.

*All K here are for Me-Asp*

### 4.2.1 An analytical theory for drift in shallow gradients

The *E. coli* chemotaxis pathway is considered one of the best characterised systems in cell biology (see Tabs. 4.1 and 4.2) [Micali and Endres, 2016]. However, from a theoretical point of view, the calculation of drift is generally hampered due to the cellular memory from adaptation by slow receptor methylation. This limits our ability to produce general analytical theories to shallow gradients. In addition, existing approaches linearise the pathway, either around the adapted state or around the steady-state generated by the cells climbing exponential gradients, thus neglecting gradients different than...
4.2. BEHAVIOUR IN TERMS OF DRIFT VELOCITY

Figure 4.1: Contributions to drift velocity. (A) Internal response, \( K(A) \), as a function of the receptor activity, \( A \). Each curve corresponds to a single phenotype. Phenotype 1: \( Y_T = 7.8 \) µM, \( D_s = 0.062 \) rad/s, \( \tau_m = 18 \) s (blue); Phenotype 2: \( Y_T = 8.6 \) µM, \( D_s = 0.062 \) rad/s, \( \tau_m = 18 \) s (red); Phenotype 3: \( Y_T = 7.8 \) µM, \( D_s = 0.062 \) rad/s, \( \tau_m = 30 \) s (green); Phenotype 4: \( Y_T = 7.8 \) µM, \( D_s = 0.042 \) rad/s, \( \tau_m = 18 \) s (purple). The average activity is not the activity which maximises \( K(A) \). (B) Methylation-independent receptor sensitivity, \( \partial F/\partial c \), as a function of external concentration, \( c \). (C) Product of methylation-independent receptor sensitivity and local gradient, \( \partial F/\partial c \nabla c \), for low (dark) and low (light) exponential gradients (blue) and linear gradients (red) as a function of the external concentration \( c \).

From the experimental point of view, drift is generally averaged over the whole population (across the whole observation chamber), removing any spatial dependence. Furthermore, there has been little comparison between theory, simulations and experiments, especially in exponential gradients, since experimentally, gradients are generally linear.

Here, I propose an analytical theory for drift. Although based on linearization, e.g. valid for shallow gradients only, this theory is valid in any regime of the receptors and accounts for all the relevant biological parameters known to be involved in chemotaxis. In order to achieve this, the model of the \( E. coli \) chemotaxis pathway presented in Tab...
4.2. BEHAVIOUR IN TERMS OF DRIFT VELOCITY

is used. Briefly, the receptor free-energy is given by the Monod-Wyman-Changeux model [Keymer et al., 2006], the receptors methylation level evolves following an integral feedback control for all ranges of concentrations [Barkai and Leibler, 1997, Clausznitzer et al., 2010], the internal protein CheY follows the receptor activity [Vladimirov et al., 2008], and the motor switch responds accordingly to the ultrasensitive response measured in [Yuan and Berg, 2013].

The pathway responses are linearised assuming small changes of receptor activity, $A$. When the linear regime holds for times longer than a single run-and-tumble event ($\sim 1s$, in shallow gradients) but shorter than the adaptation time ($\sim 10s$), the drift velocity is

$$\langle \bar{v}_d \rangle = K(A) \frac{\partial F}{\partial c} \nabla c = K(A) c \frac{\partial F}{\partial c} \nabla c,$$

(4.1)

where $K(A)$ is the internal response and it is a function of the average activity (around which the linearization has been performed). It includes the biological parameters (rate constants, adaptation time, sensitivity of receptors and rotational diffusion) (Fig. 4.1A, and Appendix B). Symbol $\nabla c$ denotes the (local) gradient steepness. Importantly, $F$ is the receptor free-energy difference between the active and inactive receptor conformations [Keymer et al., 2006], leading to the methylation independent receptor sensitivity $\partial F/\partial c$ $c = N \sum_{i=(\text{Tar, Tsr})} \nu_i \frac{(K_i^{\text{on}}-K_i^{\text{off}})c}{(c+K_i^{\text{on}})(c+K_i^{\text{on}})}$ with $\nu_i$ the fraction of receptor type $i$ for the most abundant receptor types Tar and Tsr, $K_i^{\text{on}}$ and $K_i^{\text{off}}$ the dissociation constants in the on and off states, and $N$ number of cooperative receptors in a cluster [Endres and Wingreen, 2006] (Fig. 4.1B). Note that once $\langle \bar{v}_d \rangle$ is written as a function of the methylation-independent receptor sensitivity, the local relative gradient, $\nabla c/c$, appears in Eq. (4.1) (right-hand side).

Fig. 4.1 shows the different terms of Eq. (4.1). For a fixed phenotype, the curve
for the internal response $K(A)$ is fixed (Fig. 4.1A). When the receptors are adapted, $A^*$, the corresponding internal response function, $K(A^*)$, is not at its maximum (Fig. 4.1A) [Si et al., 2012; Dufour et al., 2014; Long et al., 2016]. Experiencing an increasing concentration reduces the activity, moving $K(A)$ towards its maximum. Furthermore, for very high stimuli, the maximum can be overtaken. The methylation dynamics would act against the change in $A$, thus resulting in a ‘force’ which pulls $A$ towards its adapted value. The methylation independent receptor sensitivity $c \frac{\partial F}{\partial c}$ is a property of the sensory system and only depends on the concentration experienced (Fig. 4.1B). For a given gradient, $\frac{\partial F}{\partial c} \nabla c$ is also a function of the concentration experienced only (Fig. 4.1C). In exponential gradients (but not in linear gradients), $\frac{\partial F}{\partial c} \nabla c$ has two peaks due to receptor sensitivity to MeAsp from the Tar and Tsr receptors (Fig. 4.1C). Importantly this quantity stays rather constant for exponential gradients, while it declines in linear gradients.

The value of the receptor activity depends both on the concentration experienced and the methylation level, thus it depends on the trajectory of the cell. Hence, to complete the description of behaviour in terms of drift, a second equation that describes the average level of receptor activity is needed [Si et al., 2012; Dufour et al., 2014]. With similar calculations, the average activity in the linear regime can be calculated, obtaining a system for the drift velocity in an external gradient,

\[
\begin{align*}
\langle \bar{v}_d \rangle &= K(\langle A \rangle) \frac{\partial F}{\partial c} \nabla c, \\
\langle A \rangle &= A^* + J \frac{\partial F}{\partial c} \nabla c \langle |v_d| \rangle,
\end{align*}
\]

where $J$ is a constant, which depends on the adaptation time of the receptors and on the
changes of activity due to changes of the free-energy. $A^*$ is the adapted activity. The mathematical details and the full expression of $J$ are shown in Appendix B.

### 4.2.2 The two peaks of drift velocity in exponential gradient

Exponential gradients are those best sensed by *E. coli*, and are most likely encountered in real environments [Clausznitzer et al., 2014]. However, those gradients are difficult to reproduce in laboratory conditions and thus are not often investigated. To close this gap, here the analytical theory of drift velocity and simulations are applied to exponential gradients. As shown in Fig. 4.1C, the methylation-independent response ($\partial F / \partial c \nabla c$) to exponential gradients shows two peaks due to Tar and Tsr receptors. However, the drift velocity also depends on the internal response $K(\langle A \rangle)$, which in turn depends on the history of the cells. Are those peaks in sensitivity reflected in the behaviour of swimming cells in exponential gradients?

When phenotypically identical single-cell are simulated swimming in an exponential gradient of MeAsp $c(x) = Ae^{\lambda(x-x_0)}$, complex behaviour emerges (Fig. 4.2A,B). Indeed, once averaged along the perpendicular direction (y) and plotted as a function of the concentration sampled by the cells (which varies only in the direction along the gradient x), the simulated trajectories show the two characteristic peaks (maxima) of the drift velocity (Fig. 4.2C). Furthermore the two peaks can be explained by the analytical theory, which agrees well with simulations for exponential gradients with a small relative gradient (Fig. 4.2C).
4.2. BEHAVIOUR IN TERMS OF DRIFT VELOCITY

\[ c = A \exp \left[ \lambda (x-x_0) \right] \]

\[ \lambda = \begin{cases} 0.75 \text{ mm}^{-1} \\ 1.0 \text{ mm}^{-1} \\ 1.5 \text{ mm}^{-1} \\ 2.0 \text{ mm}^{-1} \end{cases} \]

Figure 4.2: Caption next page.
4.2. BEHAVIOUR IN TERMS OF DRIFT VELOCITY

Figure 4.2: *Swimming behaviour of identical cells in terms of trajectories and drift.* (A) An exemplar exponential chemical gradient of MeAsp. (B) Trajectories from simulations based on a modified version of Rapid-Cell software (see Appendix A for details) [Vladimirov et al., 2008]. (Inset) Trajectories with rotational diffusion ‘turned off’. (C) Average drift velocity as function of the experienced concentration from simulations (thick lines) and analytical theory (thin lines) for exponential gradients of increasing relative gradient ($\lambda = 0.4 \text{ mm}^{-1}$ orange; $\lambda = 0.75 \text{ mm}^{-1}$ red; $\lambda = 1 \text{ mm}^{-1}$ green; $\lambda = 1.5 \text{ mm}^{-1}$ blue; and $\lambda = 2 \text{ mm}^{-1}$ black). Vertical and horizontal dashed lines indicate the discrepancy between theory and simulations due to finite number of methylation sites (vertical) and very steep gradients (horizontal). Simulations of 100 cells. Parameters listed in Tab. 4.2.

4.2.3 Limits of the theory

For large relative gradients, the theory fails to match the simulations. Although this is expected, since the theory is valid only in a linear approximation and thus for shallow gradients, here I discuss what we can learn from the limits of the analytical theory.

The theory identifies three important time scales: the average run duration set by the tumble bias $\tau_b$, the adaptation time set by the methylation dynamics of CheR and CheB $\tau_m$, and finally the time after which a cell forgets its initial direction due to rotational diffusion, $\tau_r$, [Long et al., 2016]. To allow the linearisation, the theory assumes $\tau_b \ll \tau_m, \tau_r$. However, in a steep gradient, a large change in concentration results in a large change in receptor activity and in the methylation level, potentially leading to $\tau_r > \tau_m$ (see also Fig. 4.3 for phenotypes with different time scales).

Interestingly, the drift from simulations never becomes larger than a threshold value of about $3.3 \mu\text{m/s}$ for the parameters from Fig. 4.2C. This is because very steep gradients (large $\lambda$) lead to complete receptor inhibition and non-stop running, and the rotational diffusion brings cells off course. In presence of rotational diffusion, tumbles are required to restart chemotaxis, thus there is a particular run length which optimises the drift.
4.3 EXPERIMENTAL PROOF OF THE FIRST PEAK

velocity [Dufour et al., 2014, Long et al., 2016]: higher tumbling probability reduces the drift, and a lower tumbling probability does not increases the drift. However, when the rotational diffusion is removed, the horizontal threshold disappears (Fig. 4.3A).

Moreover, beyond a concentration of about $10^3$ mM, the drift declines sharply. This is attributed to the limited number of methylation sites available on each receptor. While my theory assumes an infinite number of methylation sites, in the simulations, the number of methylations is assumed to be 8, in accordance with the data. In this condition, the adaptation is lost beyond about $10^3$ mM. This concentration is seen on Fig. 4.2B at approximately $x = 22$ mm, and corresponds to a semi-permeable boundary. Here, cells, which are moving from a lower to a higher concentration reach this boundary, are then pushed into a state of permanent running. When accounting for rotational diffusion, cells can hazard back to an area where the concentration is around $10^3$ mM, and can then regain sensitivity and eventually moving up the gradient. However, if the rotational diffusion is removed from the simulations, the semi-permeable boundary disappears (Fig. 4.2B, inset).

The drift velocity for different phenotypes is shown in Fig. 4.3 in particular for different initial run biases (due to changes in the total level of CheY, $Y^r$) and for different adaptation times. Note that changing the total level of CheY has an effects both on the amplitude and the broadness of the peaks (Fig. 4.3B), while the adaptation time strongly influences the amplitude of the peak and its horizontal threshold value (Fig. 4.3C).

4.3 Experimental proof of the first peak

The complex behaviour in terms of drift velocity of swimming bacteria in exponential gradients allow to study the correspondence between drift velocity and information trans-
4.3. EXPERIMENTAL PROOF OF THE FIRST PEAK

Figure 4.3: Drift velocity in different solutions and for different phenotypes. Average drift velocity as function of the experienced concentration from simulations (thick lines) and analytical theory (thin lines) in exponential gradient of $\lambda = 0.75 \text{ mm}^{-1}$. (A) Changes in swimming behaviour due to different rotational diffusion experienced by cells: usual value $D_s = 0.062 \text{ rad/s}$ (green) \cite{Vladimirov2008}, double of the usual value (blue), and in absence of rotational diffusion (red). (B) Changes in swimming behaviour of different phenotypes which have different maximal level of CheY ($Y^T$, and thus different adapted motor bias), usual value $Y^T = 7.6 \mu M$ (green), $Y^T = 6.0 \mu M$ (red), and $Y^T = 8.0 \mu M$ (blue). (C) Changes in swimming behaviour of different phenotypes which have different adaptation time ($\tau_m$), $\tau_m = 15.5 \text{ s}$ (green), $\tau_m = 114 \text{ s}$ (red), and $\tau_m = 2.15 \text{ s}$ (blue). Simulations of 100 cells. Not described parameters are listed in Tab. 4.2.

mission. However, most experimentally determined drift velocities are based on averaging over whole field of cells, removing any spatial structure in cell behavior \cite{Colin2014}. Hence, before proceeding, it is necessary to experimentally test the existence of the two
peaks. For this reason, I contacted our experimental collaborators (Victor Sourjik lab at Max Planck Institute, Marnburg) suggesting to prove my predictions.

Experiments were performed in a microfluidic chamber, where a gradient of MeAsp is generated by maintaining a fixed concentration on a side of the chamber, and zero on the other side (Fig. 4.5A). This way, a stable linear gradient is generated after about 1h 30min. Swimming bacteria, *E. coli* strain MG1655, were introduced from the side of the chamber at zero concentration, while a camera imaged a quasi-2d focal plane in the middle of the chamber with resolution of 0.7 µm and frame rate of 30 fps. A tracking package designed by the collaborators was able to track the particles appearing in the focal plane.

I contributed to the experiments in two ways. Firstly, I designed the linear gradients and I checked whether they match the predictions in exponential gradients using simulations. Limited by experimental constraints, this was done for ligand concentrations suitable to sample only the first peak predicted by simulations in exponential gradients. Secondly, I analyzed the cell trajectories in term of drift velocity in order to check the predictions from theory and simulations.
4.3. EXPERIMENTAL PROOF OF THE FIRST PEAK

\[ \lambda = 1.00 \text{ [mm}^{-1}\text{]; } Z = 0.1 \text{ [mM]; } y_0 = 5 \text{ [mm]} \]

\[ \lambda = 0.75 \text{ [mm}^{-1}\text{]; } Z = 0.1 \text{ [mM]; } y_0 = 5 \text{ [mm]} \]

\[ \lambda = 1.25 \text{ [mm}^{-1}\text{]; } Z = 0.1 \text{ [mM]; } y_0 = 5 \text{ [mm]} \]

**Exp.** \[ c = Z e^{\lambda (y - y_0)} \]

**Linear** \[ c = m y + c_0 \]

\[ m = 10^{-4} \text{ [mM/mm]} \]

\[ c_0 = 4 \times 10^{-5} \text{ [mM]} \]

\[ m = 10^{-3} \text{ [mM/mm]} \]

\[ c_0 = 4 \times 10^{-4} \text{ [mM]} \]

\[ m = 10^{-2} \text{ [mM/mm]} \]

\[ c_0 = 4 \times 10^{-3} \text{ [mM]} \]

\[ m = 10^{-1} \text{ [mM/mm]} \]

\[ c_0 = 4 \times 10^{-2} \text{ [mM]} \]

\[ m = 1 \text{ [mM/mm]} \]

\[ c_0 = 4 \times 10^{-1} \text{ [mM]} \]

\[ m = 10 \text{ [mM/mm]} \]

\[ c_0 = 4 \text{ [mM]} \]

Measured at \( y^* = 0.6 \text{ mm} \)

**Figure 4.4:** Caption next page.
4.3. EXPERIMENTAL PROOF OF THE FIRST PEAK

Figure 4.4: Design of the experiments. (A) Simulations in linear gradients (left), \( c = my + c_0 \), with gradient \( m \) chosen to match the gradient conditions of simulations in exponential gradients \( c = Ze^{\lambda(y-y_0)} \) (right). The matching relationship for the gradients is \( m = \lambda c \), where \( c \) are chosen to span the first peak in the average drift velocity from simulations in exponential gradients (for fixed \( \lambda = 0.75 \) mm\(^{-1} \)). Finally, the position in space \( y^* \) for the linear gradient simulations (dashed colour lines) in which the matching relationship is fulfilled is kept fixed. (B) The average activity (left) and drift (right) as a function of the concentration from simulations of 100 cells. Both the average activity and the drift velocity in linear gradients match the respective average activity and drift in exponential gradients at \( c^*(y^*) \). (C) Legend of panel B.

4.3.1 Designing the experiments

The system of equations in Eq. (4.2) predicts that the average drift of a population of bacteria swimming in a (shallow) gradient has two peaks of high drift (Fig 4.2). However, it is experimentally hard to go to concentrations higher than 10 mM without affecting other physical parameters such as the pH and the viscosity of the solution. For this reason, the experiments focussed on the first peak only. Furthermore, to make reliable suggestions to the experimental collaborators, it is important to question whether experiments in linear gradients provide insights about the swimming behaviour in exponential gradients. Obviously, a linear gradient can be seen as a local approximation of an exponential gradient. However, determining the drift would not be trivial, since the memory component also plays a part. To answer the question, simulations of the experimental setup in a linear gradient were conducted. As shown in Fig. 4.4, experiments in carefully designed linear gradients match the expectation of simulations in exponential gradients, giving us the confidence to carry out the experiments.

In particular, at fixed concentration \( c^* \), the condition to have the same local relative gradient \( \nabla c|_{c^*}/c^* \) between a linear, \( c = my + c_0 \), and an exponential, \( c = Ze^{\lambda(y-y_0)} \),
4.3. EXPERIMENTAL PROOF OF THE FIRST PEAK

The gradient in the y direction is \( m = \lambda c^* \). The simulations show a good agreement between linear and exponential gradients for both the average activity and the drift velocity at concentration \( c^* \). Thus, setting the target concentrations \( c^* \) in order to span the first peak of drift and the target \( \lambda \), the \( m \) and \( c_0 \) that need to be used in the experiments were identified (schematic in Fig. 4.4).

4.3.2 Analysing the experimental trajectories

At least four replicas of experiments in linear gradients were conducted with maximal concentrations of \( c^{\text{max}} = \{2 \times 10^{-3}, 2 \times 10^{-2}, 2 \times 10^{-1}, 2, 20\} \) [mM] at the top of the channel, resulting in the target concentrations of about \( c^* \approx \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10\} \) [mM] in the middle of the channel. For each experiment, three sets of data were acquired: after 2, 3 and 4 hours. Calibration of gradients show a stable gradient from 1h30 to over 4h; thus the gradient was measured only after the 4h [Colin et al., 2014]. The trajectories were recorded by a camera of size 716.8 \times 716.8 \mu M (Fig. 4.5A).

In simulations and from the theory in linear gradients, the drift velocity varies at different concentrations, for instance decreasing at large concentrations due to adaptation and low \( \partial F/\partial c \nabla c \) (Figs. 4.4 and 3.4A). In contrast, in the experiments, the drift velocity is noisy and constant in the camera field, suggesting a constant trend for drift at different concentrations (Fig. 4.5B). This might be due to the combination of different effects. Firstly, the overall drift velocity is low, secondly the real cells have higher variability than the simulated ones, and finally the chamber might be too small to see the chemotactic velocity slow down. As a result, the drift velocity is extracted by averaging over all the camera views. The relative gradient thus ranges from about 0.4 to about 0.5 mm\(^{-1}\), and the concentrations are not precisely \( c^* \) (Fig. 4.5C).
4.3. EXPERIMENTAL PROOF OF THE FIRST PEAK

![Diagram](image)

Figure 4.5: **Experimental setup and preliminary data analysis.** (A) Schematic of the experimental setup. (B) Average drift velocity as a function of the location in the particles camera for all the experiments (gray lines). A particular experiment is shown in black line to highlight the constant noisy trend. (C) The local relative gradients as functions of the local concentration for each experiment, from the data recorded by the gradient camera (grey lines). The start and the end of the particle cameras are shown in red and blue dots, respectively. Two experiments are below the target relative gradient and are disregarded (green lines). (D) Typical distribution of the recorded particles velocity for one experiment. The distribution of all the particles (blue line and dots) is bimodal, allowing the distinction between non-swimming particles (green line and dots) and swimming particles (red line and dots).
4.4 ADDITIONAL INSIGHTS ON HIGH-DRIFT, HIGH-INFORMATION PRINCIPLE

To calculate the drift velocity, it is important to discriminate between swimming bacteria and non-swimming particles, such as bacteria that have lost their flagella’s or Brownian particles. I considered only trajectories longer than 1 s, which ensured the observation of at least one run (note that the tumble average time is 0.1 s). Furthermore, the velocity distribution of all particles is bimodal, which allows for the discrimination between swimmers and non-swimmers (Fig. 4.5D). Finally, in order to account for an eventual flow in the chamber, the drift velocity of non-swimmers (supposedly zero, in absence of flow) has been subtracted from the drift velocity of the swimmers, $v_{d}^{\text{Exp}} = v_{d}^{\text{SW}} - v_{d}^{\text{NSW}}$.

The experiments verify the Tar peak at the expected location (Fig. 4.6). However, there are minor discrepancies that might be due to the unknown sensitive regime of the used strain. Alternatively, different phenotypes in the experiments could perform differently under the same conditions, thus resulting in a different drift to the theoretical expectation, which is based on a single phenotype [Dufour et al., 2016].

4.4 Additional insights on high-drift, high-information principle

The analytical theory and experiments presented here explain some of the fine structure shown in Fig. 4.2B. However, can the regions of high and low drift be explained in terms of information gain? Previous application of information theory to E. coli chemotaxis indicated that maximal information transmission (mutual information) leads to the highest drift (Chapter 3) [Clausznitzer et al., 2014]. This suggests a potential new design principle in the bacterial sensory systems: maximizing information transmission.
Figure 4.6: **Experimental verification of peak using microfluidics.** Average drift velocity as a function of the experienced concentration from theory (blue lines and shade), simulations (green lines and shade), and experimental data (red dots). Shades span the range of exponential gradients from $\lambda = 0.4 \text{mm}^{-1}$ to $\lambda = 0.4 \text{mm}^{-1}$. For experimental data, the vertical error bars correspond to the standard deviation, while the horizontal error bars show the maximal and the minimal concentrations experienced by the cells in the trajectory cameras. Simulations of 100 cells. Parameters for theory and simulations listed in Tab. 4.2.
optimizes cell behaviour. However, that theory had a number of limitations, including
the restriction to instantaneous responses, thus neglecting any history dependency and
memory effects [Micali and Endres, 2016]. Calculating the mutual information based on
trajectories is difficult because of the high dimensional space of trajectories [Tostevin
and ten Wolde, 2009]. Here, I illustrate alternative ways to gain intuition about the
connection between drift, information gain, and sensing accuracy.

Firstly, to quantify information gain, the Kullback-Leibler divergence (KL) is pro-
posed. In general, the KL-divergence between two probabilities of observing an event $s$
is defined by $\text{KL}(p_s, q_s) := \sum_s p_s \log_2 p_s / q_s$. Note that the KL-divergence is defined only
if for all the events $s^*$ with probability $q$ equal zero, $(q_{s^*} = 0)$ probability $p$ is also equal
to zero $(p_{s^*} = 0)$.

In the simulations, at a given level of CheY_p, there is a certain probability that a
motor will switch its rotation from CW to CCW or from CCW to CW. The probability
of a run or a tumble depends on the actual state of its 3 motors, and thus needs to
be evaluated from simulations. Here, in order to evaluate the information gain along
a trajectory, the events are the run and tumble of the cell, $s \in \{\text{run, tumble}\}$, $p_{s,t}$ are
the actual probabilities of run and tumble at a particular time $t$ along the trajectory
climbing the gradient, and $q_s$ are the average probabilities of run and tumble of a cell in
the adapted state. Thus, mathematically the KL-divergence along a trajectory at time $t$
is

$$\text{KL}_t(p_{s,t}, q_s) = \sum_{s=\{\text{run, tumble}\}} p_{s,t} \log_2 \frac{p_{s,t}}{q_s}. \quad (4.3)$$

KL_t(p_{s,t}, q_s) estimates the information lost once the probability of runs and tumbles in
absence of chemotaxis is used to explain the behaviour of the cells in a particular gradient.
Indeed, when averaged over many trajectories at fixed concentrations, KL_t(p_{s,t}, q_s) peaks
4.4. ADDITIONAL INSIGHTS ON HIGH-DRIFT, HIGH-INFORMATION PRINCIPLE

at the concentration where the drift peaks (Fig. 4.7B, blue solid line). Note that in Eq. (4.3), while $p_{s,t}$ are evaluated on a trajectory, $q_s$ are averaged over time and over many trajectories (from simulations in no gradient).

Secondly, to quantify the sensing accuracy, an approach inspired by the Fisher information from estimation theory is proposed (see Sec. 2.2 for detailed introduction of Fisher information) [Brunel and Nadal, 1998]. The Fisher information focused on the likelihood, given by the conditional probability of observing an outcome given a certain parameter. Indeed, this definition can be interpreted as the curvature of the log-likelihood function and as a result the Fisher information cannot be estimated from a single event. Furthermore, it estimates a lower bound of the uncertainty of sensing, via the Cramér-Rao bound

$$\frac{\delta c}{c} \geq \frac{1}{\sqrt{c\mathcal{F}(c)}}.$$  (4.4)

In case of swimming bacteria performing chemotaxis, I am interested in a quantity similar to Fisher information that can be defined along a swimming trajectory. In this case, the likelihood is given by the conditional probability of the cell to run or tumble given the external concentration and the cell’s history, i.e. its trajectory. Thus for a trajectory, the stringent definition of Fisher information is lost. However, I can still computationally evaluate a Fisher-like quantity which estimates the information of the concentration sensed by a cell swimming up a chemical gradient over individual trajectories. In particular, this Fisher-like information along a trajectory at time $t$ is defined by

$$F_t(c) := \sum_{s=\{run,tumble\}} p_{s,t} \left( \frac{\log(p_{s,t+\delta t}) - \log(p_{s,t})}{c_{t+\delta t} - c_t} \right)^2.$$  (4.5)

Further extending the parallelism between Fisher information, $\mathcal{F}$, and Fisher-like infor-
4.4. ADDITIONAL INSIGHTS ON HIGH-DRIFT, HIGH-INFORMATION PRINCIPLE

Figure 4.7: **Connection between behavior and information processing.** (a) Average drift velocity. (b) Kullback-Leibler divergence as a measure of information gain (blue) and uncertainty of sensing from Cramér-Rao bound (red).

For information, $F$, I aim to evaluate the equivalent of the Cramér-Rao bound in Eq. (4.4) for $F$. Interestingly, when averaged over many trajectories at a fixed concentration, the lower bound in Eq. (4.4) dips at concentrations at which the drift and the KL divergence peak (Fig. 4.7B, red line), showing that in principle, the cells bias their movement in regions where they have higher accuracy of sensing. Note that to reach the Cramér-Rao bound even for the Fisher information, an efficient estimator is required. Here, the Fisher-like information is calculated over two data points, and thus the lower bound may not be reached.
4.5 Discussion

Even in the well-characterised bacterium *E. coli*, behaviour is remarkably complex. In this chapter, I summarised the behaviour in terms of drift velocity by proposing an analytical theory for drift which goes beyond existing theories [Celani and Vergassola, 2010; Jiang et al., 2010; Dufour et al., 2014; Colin et al., 2014]. The main improvement provided by this theory is that it allows for the description of all concentrations and types of gradients, without restriction to the sensitive regime of the chemoreceptors and exponential gradients. However, the theory presented here applies only to shallow gradients, as the other existing theories. As a result, the theory agrees with simulations of swimming bacteria only for shallow gradients.

Furthermore, I designed experiments in linear gradients in order to prove the predictions provided by the analytical theory and simulations. When swimming in exponential gradients from low (\(\sim 10^{-4} \text{ mM}\)) to high (\(\sim 10^4 \text{ mM}\)) concentrations of MeAsp, the average drift velocity shows two peaks of high drift in correspondence to the sensitive regimes of the most abundant receptor types. Here, I designed and analysed experiments performed by my experimental collaborators. The result is the first proof of the first peak of drift. However, the experimental data shows that the peak is slightly shifted to higher concentrations than expected, which may be due to the specific sensitivity of the bacterial strain used and to the fact that theory and simulations do not account for the variety of phenotypes which are present in real cells.

Aiming to link information gain with bacterial behaviour and guided by the discovered design principle I also presented approaches different from instantaneous mutual information. The goal was to extend my results including time dependences. However,
this goal turns out to be extremely complicated: the trajectory space is extremely high dimensional, making the mutual information over trajectories impossible to calculate in the general case [Tostevin and ten Wolde, 2009]. Thus, I proposed to adapt the concepts of Kullback-Leibler divergence and Fisher information to the trajectories of swimming bacteria. The results indicate that the regions of high drift correspond to high information gain and high Fisher information. As a consequence of the Cramér-Rao bound (if the bacteria can reach the limits imposed by Fisher information), the higher the drift the higher the accuracy of sensing. These interesting results call for further proofs. For instance, analytical theories for the KL divergence and the Fisher information are missing. Those would help to further clarify the link between high information and high drift, and would inspire further round of experiments to prove this connection. Furthermore, KL divergence is not a measure of the information processing of the environmental stimulus and the Cramér-Rao bound may not be reached for small samples. Finally, the definition of the Fisher information for trajectories is problematic and so is the related Cramér-Rao bound. A more fundamental study is required to evaluate the predictive power of this approach. In the next chapter, the link between information transmission and energy dissipation will be explored.
5. Energy dissipation during chemotaxis

5.1 Overview

The work in this thesis develops the idea that sensory systems and signalling pathways are designed to maximise information transmission for evolutionary advantages. However, organisms need to account for the energy consumed for operating the signalling pathway and, as suggested by several works in neuroscience, nature may search for efficient ways of transmitting information [Laughlin, 1981, Laughlin et al., 1998, Laughlin, 2001]. Mutual information is based on Shannon information which is mathematically liked to thermodynamics (see Sec. 1.3.4). Furthermore, recent experiments at the nanometre scale confirmed that to perform the logically irreversible step of erasing information, any computational device (alive or not) needs to dissipate energy [Bérut et al., 2012], providing a first experimental proof of the Landauer principle [Landauer, 1992]. Here, by looking at *E. coli* chemotaxis, I aim to build a bridge between information gain, behaviour and energy consumption. Is thermodynamics setting limits on the information a cell can gain about the environment? What is the energetic cost of chemotactic drift?

In the literature, both the problems of identifying the fundamental limit of sensing
and the mechanism that leads to precise adaptation have been studied from a thermodynamic point of view (reviewed in [Micali and Endres, 2016]). Since the seminal work of Berg and Purcell in the ‘70s [Berg and Purcell, 1977], scientists have tried to identify the fundamental limit of sensing, meaning the limit that a sensory system has in inferring the ligand concentrations from inherent stochastic measurements (for example see [Berg and Purcell, 1977, Endres and Wingreen, 2009, Micali et al., 2015, Aquino et al., 2016]). However, to achieve such a limit, cells need to average over time, which may require downstream proteins to store the information (memories). Both the production of downstream proteins and time averaging are in general energetically costly processes [Dekel and Alon, 2005, Skoge et al., 2013]. Thus, whether a cell works close to the fundamental limit of sensing is still an open question.

Theoretical studies suggest that the energetic costs limit the accuracy of sensing [Mehta and Schwab, 2012, Govern and ten Wolde, 2014b]. In particular, for a simple system of independent receptors and readout proteins inspired by the *E. coli* sensory system, an analytical theory identifies three independent constraints that limits the accuracy of sensing: receptors and their averaging time, the number of downstream proteins, and fuel consumption [Govern and ten Wolde, 2014b]. Preliminary experimental data provides an insight into *E. coli* resource allocation, suggesting that the three constraints are equally limiting factors to the sensing accuracy, and that this bacterium distributes resources optimally in order to avoid bottlenecks [Govern and ten Wolde, 2014b]. Further experimental data are needed to confirm these observations.

Furthermore, those studies do not account for the environmental fluctuations, which may provide work that a cell can use to enhance information gain without internal energetic costs. When simple models of sensing are embedded in fluctuating environments,
5.1. OVERVIEW

several conclusions can be drawn. For instance, the mutual information between an internal process and an external stimulus is not bounded by the thermodynamic entropy production, i.e. the cost of maintaining the network [Barato et al., 2013]. Instead, it is the conditional Shannon entropy, called learning rate, which is bounded by the thermodynamic entropy production [Barato et al., 2014]. This learning rate represents the rate at which the uncertainty about the external environment decreases due to the dynamics of the internal process. Furthermore, it has been proved that the learning rate can exceed the thermodynamic entropy produced by the internal processes if work is done by the external environment [Barato et al., 2014].

Although alternative pathway architectures that do not consume fuel have been theoretically identified, the *E. coli* chemotaxis pathway consumes energy to maintain the dissipative steady state, and the cost of responding to a step-change in ligand concentration is only 10% higher than the steady state cost [Sartori et al., 2014, De Palo and Endres, 2013]. Indeed, there are several advantages to maintaining such dissipative pathways. For instance, the higher the energy consumption, the faster and the more accurate the adaptation [De Palo and Endres, 2013, Lan et al., 2012, Sartori and Tu, 2015]. Furthermore, the energy can be used to reduce the correlations between cell-external and cell-internal noise, which ultimately results in a higher accuracy of sensing by time averaging [Govern and ten Wolde, 2014a].

However, those works focus on detailed models of receptors and downstream proteins, neglecting the motors and the actual movement of cells. Here, I aim to estimate the entropy production of an individual cell along its swimming trajectory. The goal is to compare the entropy production along a trajectory with the information gain and drift velocity.
5.2 Entropy production during chemotaxis in exponential gradients

Here, we aim to estimate the energy consumption by looking at the entropy changes in time. Note that the time derivative of the entropy $dS/dt$ has two contributions: the entropy production rate, which is the free-energy dissipation rate, and the entropy exchange rate due to exchange with the environment (entropy flow rate) [Ge and Qian, 2010]. Note, assuming matter within quasi-steady state means that the two contributions to the entropy change are the same, but with opposite sign. There are two ways to estimate the entropy production of a cell along its swimming trajectory in an exponential gradient. Here, I briefly introduce the methods and show how I applied those to the simulations shown in Chapter 4.

Entropy production from chemical reactions

Given a simple chemical reaction, by which a species A is converted to species B with rate $k^+$ and reverse reaction rate $k^-$, in an isolated system at equilibrium

$$\text{A} \xrightleftharpoons[k^-]{k^+} \text{B},$$

the relative number of molecules of A and B is given by

$$\frac{[B]_{\text{eq}}}{[A]_{\text{eq}}} = \frac{k^+}{k^-} = e^{-\frac{\mu^0_B - \mu^0_A}{k_B T}},$$

where $k_B$ is the Boltzmann constant, $T$ is the temperature, and $\mu^0$ is the the chemical potential at equilibrium. Note that from the internal energy of a system $U$, the funda-
5.2. ENTROPY PRODUCTION DURING CHEMOTAXIS IN EXPONENTIAL GRADIENTS

mental relation of thermodynamics $dU = TdS - pdV + \sum_i \mu_i dN_i$ provides a definition of the chemical potential

$$\mu_i := \left( \frac{\partial U}{\partial N_i} \right)_{S,V,N_j \neq i} = -T \left( \frac{\partial S}{\partial N_i} \right)_{U,V,N_j \neq i},$$  \hspace{1cm} (5.1)$$

where $S$ is the entropy, $N_i$ is the number of particles of species $i$, $V$ is the volume, and $T$ is the temperature. Note that the chemical potential of species $i$ is normally defined as the partial derivative of the internal energy with respect of the number of particles of the species $i$ at fixed entropy, volume and other species number. However, by inverting the fundamental relation $(dS = dU/T + p/TdV - \sum_i \mu_i/TdN_i)$, the chemical potential can be express as the partial derivative of the entropy with respect to the number of particles of the species $i$, at fixed internal energy, volume and other species number. In an out-of-equilibrium system, in which the number of particles of species A and B are $[A]$ and $[B]$ respectively, a flux $J$ appears, which is given by

$$J = J^+ - J^- = k^+ [A] - k^- [B].$$

The chemical potentials will then be $\mu_A = \mu_A^0 + k_B T \log [A]$ and $\mu_B = \mu_B^0 + k_B T \log [B]$, and the difference in chemical potential is

$$\Delta \mu = \mu_B - \mu_A = k_B T \log \frac{J^-}{J^+},$$  \hspace{1cm} (5.2)$$

[Qian and Beard, 2005] [Qian, 2007]. Thus, the entropy production for a biochemical (BC) reaction is

$$\frac{dS_{BC}}{dt} = k_B J \log \frac{J^+}{J^-}. \hspace{1cm} (5.3)$$
5.2. ENTROPY PRODUCTION DURING CHEMOTAXIS IN EXPONENTIAL GRADIENTS

Here, assuming that the motors consume the same energy to rotate CW and CCW, i.e. the motor entropy production is the same for the run and the tumble state [Gabel and Berg, 2003], the biochemical reactions considered are the methylation of receptors and phosphorylation of the response regulator CheY. The phosphorylation of CheY is governed by the reaction

\[ Y + ATP \xrightarrow{A k_y^+} Y_p + ADP \xrightarrow{Z k_z^+} Y + ADP + P_i, \]

where \( A \) is the receptor activity, \( Z \) is the CheZ concentration, \([ATP]\), \([ADP]\), \([Pi]\) are the concentrations of ATP, ADP and phosphate, respectively; \( Y_p \) and \( Y = Y^T - Y_p \) are the concentrations of CheY_p and CheY, respectively. The dynamical equation is

\[ \frac{dY_p}{dt} = \tilde{k}_y^+ A (Y^T - Y_p) - \tilde{k}_y^- AY_p - \tilde{k}_z^+ Y_p + \tilde{k}_z^- Y, \]

where effective parameters have been used [Vladimirov et al., 2008]. In particular, \( \tilde{k}_y^+ = k_y^+[ATP] = 100 \mu M^{-1}s^{-1}, \tilde{k}_y^- = k_y^-[ADP] = \nu_1 k_y^+ \) with \( \nu_1 \ll 1, \tilde{k}_z^+ = Z k_z^+ = 30 \text{ s}^{-1}, \)

and \( \tilde{k}_z^- = Z k_z^-[Pi] = \nu_2 k_z^+ \), with \( \nu_2 \ll 1 \). Using Eq. (5.3), the entropy production of the phosphorylation dynamics of CheY_p is

\[ \frac{dS_{BC}^Y}{dt} = k_B \left[ \tilde{k}_y^+(Y^T - Y_p) - \tilde{k}_y^- Y_p \right] A \log \left[ \frac{\tilde{k}_y^+(Y^T - Y_p)}{\tilde{k}_y^- Y_p} \right] + k_B \left[ \tilde{k}_z^+ Y_p - \tilde{k}_z^- (Y^T - Y_p) \right] \log \left[ \frac{\tilde{k}_z^+ Y_p}{\tilde{k}_z^- (Y^T - Y_p)} \right]. \] (5.4)
Similarly, the methylation and demethylation dynamics are governed by the reactions

\[
[m]_0 + \text{SAM} \xrightleftharpoons[k_R^-]{k_R^+} [m + 1]_0 + \text{SAH}
\]

\[
[m + 1]_1 + \text{H}_2\text{O} \xrightleftharpoons[k_B^-]{k_B^+} [m]_1 + \text{CH}_3\text{OH},
\]

where only the inactive sites \([m]_0\) are methylated and only the active site \([m]_1\) are demethylated. The effective dynamical equation is given by \cite{Clausznitzer2010}

\[
\frac{dm}{dt} = \left(\tilde{k}_R^+ - \tilde{k}_R^-\right)(1 - A) - \left(\tilde{k}_B^+ - \tilde{k}_B^-\right)A^3,
\]

where effective parameter are \(\tilde{k}_R^+ = k_R^+[\text{SAM}] = 0.0069\ \text{s}^{-1}\), \(\tilde{k}_R^- = k_R^-[\text{SAH}] = \nu_3\tilde{k}_R^+\), \(\tilde{k}_B^+ = k_B^+[\text{H}_2\text{O}] = 0.12\ \text{s}^{-1}\), \(\tilde{k}_B^- = k_B^-[\text{CH}_3\text{OH}] = \nu_4\tilde{k}_B^+\), and \([m]_T = 10^4\) is the total number of methylation sites \cite{Clausznitzer2010,Li2004}. Hence, using Eq. (5.3), the entropy produced by the methylation and demethylation dynamics is

\[
\frac{dS_{\text{BC}}^m}{dt} = k_B\left(\tilde{k}_R^+ - \tilde{k}_R^-\right)(1 - A)[m]_T \log\left(\frac{\tilde{k}_R^+}{\tilde{k}_R^-}\right) +
\]

\[
+ k_B\left(\tilde{k}_B^+ - \tilde{k}_B^-\right)A^3[m]_T \log\left(\frac{\tilde{k}_B^+}{\tilde{k}_B^-}\right).
\] (5.5)

Surprisingly, when \(dS_{\text{BC}}/dt = dS_{\text{Y}}/dt + dS_{\text{BC}}^m/dt\) is calculated along the trajectories of swimming bacteria in an exponential gradient from simulations, the entropy production decreases in the region of high drift (Fig. 5.1, blue line). However, this is simply because a higher ligand concentration inhibits receptor activity and phosphorylation. In other bacteria, such as \textit{Bacillus subtilis}, the opposite behaviour would be seen, i.e. an increased rate of entropy production during chemotaxis \cite{Rao2008}. 

129
5.2. ENTROPY PRODUCTION DURING CHEMOTAXIS IN EXPONENTIAL GRADIENTS

Lower limit of entropy production from fluctuation theorem

To study a large number of particles in a box, one can describe the dynamics of each particle from a mechanical point of view, or use thermodynamics to describe the state of the system. The great success of thermodynamics is due to the fact that it allows us to describe a system of a large number of particles with a few physical observables. Such observables can be of two types, either extensive (e.g. entropy $S$, volume $V$ and the number of particles $N$), or intensive (e.g. temperature $T$, pressure $p$, and chemical potential $\mu$), depending on whether they scale with the size of the system (extensive) or not (intensive). Of particular interest is the paradox that emerges from the formulation of the second law of thermodynamics. This law states that the entropy of an isolated system always increases, providing the system with a temporal direction. However, the laws of physics that govern the mechanics are symmetric under time reversal. This paradox is called Loshmidt’s paradox [Sevick et al., 2008, Seifert, 2012]. Furthermore, other limitations of the classical formulation of thermodynamics concern the fact that the thermodynamical laws are only valid for large systems (thermodynamical limit) at equilibrium (or in its proximity). Recently, new tools from stochastic thermodynamics such as fluctuation theorems have provided an explanation of Loshmidt’s paradox, and have expanded the systems that can be understood in thermodynamic terms [Crooks, 1999, Evans and Searles, 2002, Sevick et al., 2008, Seifert, 2012]. Here, I briefly introduce the Evans-Searles fluctuation theorem and show how it can be applied to the trajectories of swimming bacteria in exponential gradients.

The Evans-Searles fluctuation theorem (FT) bridges the microscopic and macroscopic description of a system of any size, arbitrarily far away from equilibrium. The focus of FT is probability $p(\Omega_t)$ of observing a functional $\Omega_t = \Omega[x(t)]$ along a trajectory of length
5.2. ENTROPY PRODUCTION DURING CHEMOTAXIS IN EXPONENTIAL GRADIENTS

Such a functional is normally the dimensionless dissipated energy. The theorem states

\[ \frac{p(\Omega_t = \sigma)}{p(\Omega_t = -\sigma)} = e^\sigma, \]  

(5.6)

thus expressing the ratio of the probabilities of observing trajectories of duration \( t \) characterised by the dissipation function \( \Omega_t \) taking values \( \sigma \) and \(-\sigma\), respectively. Although the dissipated energy would depend on the system, under time-reversible mechanics, any trajectory with \( \Omega_t = \sigma \) has a time-reversed trajectory ('anti-trajectory') with \( \Omega_t = -\sigma \) [Sevick et al., 2008]. This means that Eq. (5.6) can be interpreted as a relation between trajectories (\( \Gamma \)) and anti-trajectories (\( \bar{\Gamma} \)), \( p_\Gamma/p_{\bar{\Gamma}} = \exp(\delta S_{FT}) \).

This theorem solved Loshmidt’s paradox and extends the limits of thermodynamics. A key result is that the second law of thermodynamics is violated since there is a non-zero probability of observing trajectories in which the entropy decreases. However, observable \( \Omega_t \) is extensive, and thus as the system size increases (or the observation time becomes longer), antitrajectories become increasingly rare. In such a case, it becomes extremely likely that the system looks time irreversible, in agreement with the second law of thermodynamics. Additionally, it follows from Eq. (5.6) that the ensemble-averaged dissipation function is positive for any system size and for any \( t \) [Sevick et al., 2008, Seifert, 2012].

The proof is simple and provided here. From Eq. (5.6), \( \langle e^{-\sigma} \rangle = 1 \). Using Jensen’s inequality \( \langle e^{-\sigma} \rangle \geq e^{-\langle \sigma \rangle} \), and thus \( \langle \sigma \rangle \geq 0 \).

To apply FT to the trajectories of swimming bacteria in an exponential gradient of attractant, I need to define and simulate anti-trajectories, \( \bar{\Gamma} \). The forward 2d trajectory \( \Gamma \) of a cell swimming in a gradient is defined by the pathway \([x(t), y(t), c(t), m(t), s(t)]\), where \( x \) and \( y \) are the spatial coordinates, starting in \( (x_0, y_0) \) at time \( t_0 = 0 \), and finishing in \( (x_f, y_f) \) at \( t_f = T \), \( c \) is the concentration experienced by the swimming cell, \( m \) is the
5.3. DISCUSSION

receptor methylation level of the cell, and s is the state of the cell, either run or tumble. The corresponding $\bar{\Gamma}$ is defined by the pathway $[x(T-t), y(T-t), c(T-t), m^*(t), s(T-t)]$. Thus, given a trajectory, the coordinates, the concentration experienced and the state of the corresponding anti-trajectory are also known, while the receptor methylation dynamics of the anti-trajectory, $m^*$, need to be extracted by reversing the simulation and solving the corresponding differential equations (see Appendix A). Given $\Gamma$ and $\bar{\Gamma}$, the probabilities of the cell to run in a time $\delta t$ can be calculated from the model of the chemotaxis pathway for both the trajectory and anti-trajectory ($p_{\Gamma_t}$ and $p_{\bar{\Gamma}_t}$, respectively) (see Appendix A). Thus, I used

$$\frac{\delta S_{FT}}{\delta t} = \log \left( \frac{p_{\Gamma_t}}{p_{\bar{\Gamma}_t}} \right)$$

(5.7)

to evaluate the entropy production along trajectories, in unit of $k_B$. Note that in Eq. (5.7), $p^i$ is the probability of keeping running if the cell is in the run state or the probability of keeping tumbling if the cell is in the tumble state (see Appendix B). Interestingly, the lower bound of entropy production at a given concentration, once averaged over multiple trajectories, again shows a peak in the region of high drift as expected (Fig. 5.1, red line). This also shows that the actual entropy production based on biochemistry is orders of magnitude higher, and it is interesting to wonder if all this energy consumption is needed for efficient chemotaxis.

5.3 Discussion

In this chapter, I presented a preliminary study about the thermodynamics of chemotaxis. Previous results and existing theories suggested that the E. coli chemotaxis path-
Figure 5.1: Connection between behavior and energy dissipation. (A) Average drift velocity. (B) Actual entropy production from biochemistry (blue) and lower bound from fluctuation theorem (red).

way consumes energy even in the adapted state [Lan et al., 2012, De Palo and Endres, 2013, Sartori et al., 2014]. Furthermore, the energetic cost to respond to extracellular changes of the chemoreceptor is only 10% higher than the cost to maintain the adapted state [Sartori et al., 2014]. Information theory and thermodynamic entropy are conceptually linked [Parrondo et al., 2015]. Hence, I expected to observe a higher entropy production in regions of high information gain and high drift velocity.

Indeed, the lower limit of entropy production estimated from the Evans-Searles fluctuation theorem agrees with the predictions. However, the actual entropy estimated from the biochemical reactions shows the opposite behaviour, suggesting that cells produce
more entropy in regions of low drift and hence low information gain. This is a characteristic of the *E. coli* chemotaxis pathway; the dips of entropy production in regions of high drift velocity are due to the phosphorylation of CheY, which happens with higher rate in the adapted state. Finally, the entropy estimated from the biochemical reactions is three orders of magnitude higher than the lower limit predicted by the fluctuation theorem. Overall, the analysis presented here shows that the thermodynamic cost required to increase the drift is negligible compared to the cost of maintaining the pathway. Why has evolution selected such an energetically costly pathway? Why is the thermodynamic cost higher in the adapted state than in regions of high drift? To answer these questions further studies are needed, in particular into the role of energy dissipation in reducing fluctuations in molecular processes.
Conclusions

Chemotaxis of *E. coli* is considered a relatively simple and well-characterised biological process, which has inspired many theories and design principles that might be used by more complex organisms, just like the hydrogen atom in physics had aided the development of quantum mechanics and theory of complex matter [Sourjik and Wingreen, 2012; Tu, 2013; Micali and Endres, 2016]. Here, I viewed the sensing and signalling of *E. coli* chemotaxis from an information-theoretic point of view. In particular, I focused on mutual information, a quantity which describes the statistical dependencies between inputs and outputs of a general communication channel. Mutual information provides a way of evaluating the ability to infer the input after measuring the output. Being extremely general, it can be used to study information transmission in any biological system [Bowsher and Swain, 2014; Mc Mahon et al., 2014; Levchenko and Nemenman, 2014]. However, mutual information becomes problematic when the time dependency needs to be encoded, since the evaluation of mutual information over trajectory requires the probability distributions of a much larger space, making the calculation extremely difficult in general [Tostevin and ten Wolde, 2009; Rivoire and Leibler, 2011]. Thus, I proposed other quantities to measure the information gain in order to account for the slow methylation/de-methylation dynamics of the receptors. Overall, the main result of this thesis is a new design principle which links the information transmission along the
pathways with behaviour in terms of the drift velocity up a chemical gradient.

In chapter 2, the distribution of inputs and the input-output curves which maximise the instantaneous mutual information are identified analytically. This was done in the very general, biologically inspired case of a system in which both inputs and input-output curves are under evolutionary pressure. Although limited by the lack of experimental data and detailed mathematical models that explain the cell-external and cell-internal noise contributions, I argued that the maximisation proposed here is an important methodological development in the field.

I used this maximisation to compare the optimal input-output curves with the measured dose-response curves of *E. coli* chemotaxis, both at the receptor and motor levels. The conclusion is that the receptor transmits information optimally, while the ultrasteep response of a single motor limits information transmission. However, I argued that for an increasing number of conditionally independent motors, the measured responses are within the optimal information transmission theory. This study makes the important prediction that single-motor bacteria with one response regulator protein should not have an ultrasteep motor response.

In chapter 3, assuming that the sensory system is designed to optimally transmit the natural inputs, after the receptor dose-response curves have been measured by my experimental collaborators, I was able to make predictions about the input distribution of the chemical attractant aspartate encountered by cells in their natural environment. Using simulations, I went beyond the identification of concentration by identifying the gradients which are transmitted best. This result has a number of limitations since the experiments were perform over Tar-only mutants and using the non-metabolisable analogue of aspartate, MeAsp. However, the method presented in this chapter is generally
applicable to any attractant and even wild-type cells with different receptor types.

Interestingly, the optimal linear gradients and concentrations identified in chapter 3 are also the ones in which the cells have the highest drift velocity, meaning that they climb the gradient with the highest velocity. This suggests a new design principle which links optimal information transmission at the receptors with maximal drift velocity in linear gradients. However, there are two main limitations to my approach. Firstly, information transmission is evaluated instantaneously using mutual information, thus neglecting the histories of the cells. Secondly, simulations were performed in linear gradients only.

In chapter 4, I presented an analytical theory for drift velocity of cells in any type of gradient, corroborated by experimental data from my collaborators and computational simulations. The behaviour is quite complex even for phenotypically identical cells. For example, in exponential gradients, concentrations of high drift are separated by concentrations of low drift. I discussed the limits of the theory and the behaviours in details. Importantly, I designed experiments which were performed by my collaborators to experimentally prove the predicted regions of high and low drift. To extend the high-information, high-drift principle, I suggested two alternatives to mutual information (which is in general impossible to calculate for trajectories without severe approximations [Tostevin and ten Wolde, 2009]): the KL-divergence and the Fisher information to estimate the information gain. The results reinforce the principle nicely. However, an analytical theory for information gain is missing and both the KL-divergence and the Fisher information have some limitations. The KL-divergence is a comparison between the behaviour in a uniform environment and in a gradient, and thus is not a direct measure of the environmental information. The definition of Fisher information over trajectories is problematic. However, I defined a Fisher-like quantity for trajectories. I used this
quantity to evaluate the Cramér-Rao bound, which is a lower bound on the accuracy of sensing when related to Fisher information. Although this lower bound on the accuracy is higher in the regions of high drift, further investigation into the Fisher-like information is required and the lower bound may not be achieved.

Finally, in chapter 5, I focused on the thermodynamics of chemotaxis. The fuel consumption was previously identified as a limiting quantity to the accuracy of sensing [Govern and ten Wolde, 2014b], and mutual information is conceptually linked with thermodynamics [Parrondo et al., 2015]. My work shows that the energy required to bias the movement and perform chemotaxis is three orders of magnitude lower than the actual energy consumed to maintain the pathway, which I estimated from the biochemical reactions involved in chemotaxis. This suggests that the energy is not a limiting factor in the high-information, high-drift principle, and it calls for further investigations in order to understand why evolution selected for such an energetically costly pathway. Noise reduction might be one advantage [Lan et al., 2012, Sartori and Tu, 2015].

Overall, the study presented here is a concise study of how a relatively simple pathway processes environmental information. A great achievement is the heuristic connection between drift, information gain and energy consumption. However, further studies are required to strengthen those connections. For instance, I aim to build detailed models for the cell-external and cell-internal noise to provide a final proof of optimal information transmission along the *E. coli* chemotaxis pathway despite the ultrasteep response of the motors. In addition, one could apply the same information transmission approach to *Rhodobacter sphaeroides* that has a single flagellated motor but three response regulators (three CheY’s). The symmetry of mutual information naively suggests that this bacterium might be able to transmit optimal information transmission with a different
architecture of the pathway. Furthermore, I aim to provide an analytical theory for the KL-divergence, the Fisher information, and the energy dissipation. For example, to calculate the Fisher information, I aim use maximum-likelihood estimation over a single trajectory in a shallow linear gradient. This would potentially allow me to identify connections and limits of the heuristic connections between information and drift that I made here.

My work does not account for the phenotypic variability observed in clonal populations. This variability was shown to increase the chances of survival of the species [Balanban et al., 2004] [Kussell and Leibler, 2005] [Frankel et al., 2014]. In my opinion, one of the most fascinating questions is to understand the sources and the levels of noise that are detrimental or beneficial for the cell. In particular, I believe that the future studies on *E. coli* chemotaxis need to account for conditions closer to the ones experienced in nature. For instance, models and experiments should focus on decision making in chemicals sensing, such as nutrient, autoinducer, temperature, and pH gradients simultaneously [Kalinin et al., 2010] [Krembel et al., 2015] [Laganenka et al., 2016] [Oleksiuk et al., 2011] [Yang and Sourjik, 2012]. Furthermore, the role of the cell division in bacterial chemotaxis is often neglected since experiments are done over non dividing cells [Jordan et al., 2013] [Leibler, 2016].

Other improvements can be achieved by using stochastic thermodynamics. The energy consumption and the limits that it imposes over the cellular process are poorly understood, and when models are proposed, the experimental proof is missing. However, experiments that nowadays have been performed on colloidal particles at nanometre scale might become available to a larger scale in the near future [Béruit et al., 2012]. Then, the simple and well-characterised *E. coli* chemotaxis pathway will be used, once again, as a
model to extrapolate design principles.
Bibliography


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


CONCLUSIONS


Appendix A: Description of the computational simulations

The computational simulations are performed in Java and analysed in R. The Java code is based on RapidCell 1.4.2 published in Vladimirov et al., 2008. This software has the advantage of simulating large number of cells with relatively low computational time. This is achieved assuming the internal pathway to be at quasi-steady state, thus avoiding to solve differential equations for the internal dynamics of CheY. The simulations do not account for the interactions between cells.

Setting of the initial condition. There are several parameters and conditions that need to be set before starting simulations.

- Conditions on the box. The simulations are normally produced inside a rectangular box of dimension $2 \times 3$ cm with periodic boundary conditions on the borders perpendicular to the gradient direction and adsorbing boundary conditions on the borders in the direction parallel to the gradient.

- Conditions on the concentrations inside the box. The concentrations and gradients inside the box can be arbitrary and are set for the purpose of the simulation, $c(x, y)$, e.g. linear and exponential gradients, with or without noise. Normally the direction
CONCLUSIONS

of the gradient is along the y-axis.

- Biochemical parameters and rotational diffusion. All parameters appearing in Tab 4.2 need to be set, including the receptor number and their sensitivity, the reaction rates for CheY phosphorylation and for the methylation/de-methylation dynamics, and the motor switching constant rates.

- Cells parameters. The number of cells, the number of rotary motors, the run velocity (constant) and the initial state of each cell need to be specified. Normally, simulations involve from 30 to 100 cells, the run velocity \( v_r \) is extracted from data or set to 20 \( \mu m/s \), and each cells has three rotary motors. The state of each cell is specified by an array which includes the time \( t \), the position \( x \) and \( y \), the orientation \( \theta \), the concentration experienced by the cell \( c \), the receptor activity \( A \), the methylation level of the receptors \( m \), the CCW bias \( \text{CCWb} \), the number of motors moving CW \( n_{CW} \) and the run or tumble state of the cell \( \text{RoT} \). A typical initial state of the cell is

\[
\begin{align*}
(t_0 &= 0 \text{ s}, \ x_0 = 1 \text{ cm}, \ y_0 = 0.25 \text{ cm}, \ \theta = \text{random} \in [0, 2\pi], \ c_0 = c(x_0, y_0), \\
A_0 &= A^* , \ m_0 : A(c_0, m_0) = A_0, \ \text{CCWb}_0 = \text{CCWb}^*, \ n_{CW_0} = 0, \ \text{RoT}_0 = \text{run})
\end{align*}
\]

where the asterisk identifies the adapted quantities.

Description of the code for simulations. At each time step in the simulation, the state of the cells evolves as follow:

- The run or tumble state of the cell is evaluated based on the state of the motors.

I introduced the veto model in which one motor in the CW state is enough for a
CONCLUSIONS

• Based on the cell state, the position and the orientation are updated. If the RoT state is in a ‘tumble’, the position does not change and the orientation $\theta$ is randomly assigned, $\theta \in [0, 2\pi]$. If the RoT state is in a ‘run’, the position is updated: $x_t = x_{t-\delta t} + v_r \delta t \cos(\theta_{t-\delta t})$, $y_t = y_{t-\delta t} + v_r \delta t \sin(\theta_{t-\delta t})$. The orientation is then updated in order to account for rotational diffusion: $\theta_t = \theta_{t-\delta t} + \delta \theta$, where $\delta \theta$ is Gaussian distributed with mean given by the actual orientation and standard deviation $D_s$.

• The concentration experienced is updated given the new position in the box, $c_t = c(x_t, y_t)$. In case the simulation accounts for the noise in extracellular concentration, a noisy contribution is added on top of the deterministic gradient. $c_t = c(x_t, y_t) + \eta$, with $\eta$ random number.

• The receptor activity is updated given the concentration and the methylation level,

$$A_t = \left\{ 1 + \exp\left[ N \left( \epsilon(m_{t-\delta t}) + \sum_{i \in \{\text{Tar, Tsr}\}} \nu_i \log \frac{1 + c_t/k_i^{\text{off}}}{1 + c_t/k_i^{\text{on}}} \right) \right] \right\}^{-1}, \tag{A.1}$$

where $\epsilon(m) = 1 - 0.5m$, $N$ is the number of receptor, $\nu_i$ is the fraction of Tar and Tsr receptors, and $k_i^{\text{on}}$ and $k_i^{\text{off}}$ define the region of sensitivity of the receptors.

• The methylation level is updated following the ODE given in Clausznitzer et al., 2010,

$$m_t = m_{t-\delta t} + g_R (1 - A_t) \delta t - g_B A_t^3 \delta t. \tag{A.2}$$

• The rate of switching from CW to CCW and from CW to CCW are updated
following Cluzel et al., 2000:

\[
\begin{align*}
    k_{\text{CCW} \rightarrow \text{CW}}^t &= 1.3 \exp \left\{ - \left( 20 - 40 \frac{Y^p_t}{Y^p_t + 3.06} \right) \right\}, \\
    k_{\text{CW} \rightarrow \text{CCW}}^t &= 1.3 \exp \left\{ + \left( 20 - 40 \frac{Y^p_t}{Y^p_t + 3.06} \right) \right\},
\end{align*}
\]

where \( Y^p_t = \frac{k_Y A_t Y_T^T}{k_Y A_t + k_Z X^T + \gamma} \), and thus

\[
\text{CCWb}_t = \left( \frac{Y^p_t}{K^m + Y^p_t} \right)^m. \tag{A.3}
\]

For the definitions of the rates and parameters involved see Tab. 4.2.

- The motor state is updated based on the probability given by the rate of switching multiplied by the time step.

- The time is updated \( t = t + \delta t \).

**Simulation output file.** The simulation produces an output file in which for each cell, at each time, a row represents the state of the cell, normally given by (cell number, \( t, x, y, \theta, c, A, m, \text{CCWb}, \text{nCW}, \text{RoT} \)) or by a subset of interest. However, the output does not include the state for each time of the simulation, to avoid large files. Normally, the simulation time step \( \delta t = 0.001 \text{ s} \) and the output lines are produced each 100 time steps. Note, however, that for the evaluation of the fluctuation theorem, the output is produced at each time step.

**Evaluation of the CI and the drift velocity.** The output file is normally analysed using R. Each trajectory is divided in pieces depending either on the methylation level of the receptors (Chapter 3) or depending on the position in the box (Chapter 4). The
chemotaxis index (CI) and the drift velocity are calculated for each piece of trajectory following

$$CI = \sum_i \frac{\delta y_i}{\sqrt{\delta x_i^2 + \delta y_i^2}},$$
$$v_d = \sum_i \frac{\delta y_i}{\delta t},$$

where $y$ is the direction of the gradient and $\delta$ are the steps in the output file.

**The anti-trajectories.** The anti-trajectories are computed by reversing the trajectory and the dynamics. Given a trajectory of a cell $i$,

$$\Gamma = \Gamma(i, t, x_t, y_t, \theta_t, c_t, A_t, m_t, CCWb_t, nCW_t, RoT_t),$$

the anti-trajectory is

$$\bar{\Gamma} = \bar{\Gamma}(i, t^r = t, x^r_t = x_{T-t}, y^r_t = y_{T-t}, \theta^r_t = \theta_{T-t} - \pi, c^r_t = c_{T-t},$$

$$A^r, m^r, CCWb^r, nCW^r_t = nCW_{T-t}^r, RoT^r_t = RoT_{T-t}),$$

where $T$ is the time at the end of the simulation. $A^r, m^r, CCWb^r$ are calculated by assuming the initial state to be $\bar{\Gamma}_0 = \bar{\Gamma}_0(i, 0, x_T, y_T, \theta_T - \pi, c_T, A_T, m_T, CCWb_T, nCW_T, RoT_T)$, and inverting the dynamics, i.e. applying Eqs. 5.8-5.10 to the reverse quantities $A^r, m^r$, \ldots
Note, that the probability $p_{\Gamma_{i}}^t$ in Eq. (5.7) is given by

$$p_{\Gamma_{i}}^t = \begin{cases} (1 - k_{\text{CCW} \rightarrow \text{CW}}^t \delta t)^3 & \text{if } n_{\text{CW}} = 0 \\ 1 - (1 - k_{\text{CCW} \rightarrow \text{CW}}^t \delta t)^2 (k_{\text{CW} \rightarrow \text{CCW}}^t \delta t) & \text{if } n_{\text{CW}} = 1 \\ 1 - (1 - k_{\text{CCW} \rightarrow \text{CW}}^t \delta t) (k_{\text{CW} \rightarrow \text{CCW}}^t \delta t)^2 & \text{if } n_{\text{CW}} = 2 \\ 1 - (k_{\text{CW} \rightarrow \text{CCW}}^t \delta t)^3 & \text{if } n_{\text{CW}} = 3 \end{cases}$$

(A.4)

and $p_{\Gamma_{i}}^t$ is given by Eq. ([A.4]) for the reverse trajectory.
Appendix B: The linear theory for drift velocity

The theory for the drift velocity of a cell swimming in a given concentration $c$ and gradient $\nabla c$ of MeAsp is based on the linearisation of the *E. coli* pathway. Here, I present the details of calculations that lead to Eqs. (4.1) and (4.2). The calculations were performed ‘old-style’, on paper, and checked with MATHEMATICA 8.

The receptor activity, $A$, is a function of the receptor free-energy, $F$, that in turn depends on methylation level $m$ and concentration $c$,

$$A(m, c) = \left[1 + e^{F(m, c)}\right]^{-1},$$

$$F(m, c) = N \left[\epsilon(m) + \sum_{i \in \{\text{Tar}, \text{Tsr}\}} \nu_i \log \frac{1 + c/k_{i\text{off}}}{1 + c/k_{i\text{on}}} \right]$$

where $\epsilon(m) = 1 - \gamma m$, with $\gamma$ constant, $N$ is the number of receptors in a cluster, $\nu_i$ is the fraction of Tar and Tsr receptors, and $k_{i\text{on}}$ and $k_{i\text{off}}$ are the dissociation constants for the active and inactive state of receptors, respectively. Linearising the activity around the initial state with $A_0$ in position $(x_0, y_0)$ for small changes in time provides

$$\delta A = \left. \frac{\partial A}{\partial F} \right|_0 \left( \left. \frac{\partial F}{\partial m} \right|_0 \delta m + \left. \frac{\partial F}{\partial c} \right|_0 \delta c \right).$$

(B.1)
CONCLUSIONS

Note that

\[
\frac{\partial A}{\partial F} \bigg|_0 = -\frac{e^{F_0}}{(1 + e^{F_0})^2} = A_0(1 - A_0),
\]

\[
\frac{\partial F}{\partial m} \bigg|_0 = N \frac{d\epsilon}{dm} = -N \gamma.
\]

The change in concentration \(\delta c\) can be easily expressed as a change in time

\[
\delta c = \nabla c|_0 \cdot \bar{v}_r \delta t,
\]

where \(\bar{v}_r\) is the vector representing the instantaneous run velocity of the cell, which is assumed to have constant module for all the runs. The change of methylation level in \(\delta t\) is calculated from

\[
\frac{dm}{dt} = \gamma_R [1 - A(t)] - \gamma_B A(t)^3,
\]

where \(\gamma_R\) and \(\gamma_B\) are the methylation and de-methylation rates, respectively. Thus \(\delta m\) is obtained by solving

\[
\frac{d(\delta m)}{dt} = -\left[ \gamma_R + 3(A^*)^2 \gamma_B \right] \delta A = -\zeta \delta A,
\]

where \(A^*\) is the adapted activity, and \(\zeta := (\gamma_R + 3(A^*)^2 \gamma_B)^{-1}\). Inserting Eqs. \[(B.1)\] and \[(B.2)\] in Eq. \[(B.3)\] gives

\[
\frac{d(\delta m)}{dt} = -\zeta \frac{\partial A}{\partial F} \bigg|_0 \left( \frac{\partial F}{\partial m} \bigg|_0 \delta m + \frac{\partial F}{\partial c} \bigg|_0 \nabla c|_0 \cdot \bar{v}_r \delta t \right)
\]

\[
= -\zeta (1 - A_0) A_0 \left( N \gamma_m - \frac{\partial F}{\partial c} \bigg|_0 \nabla c|_0 \cdot \bar{v}_r \delta t \right).
\]
This equation is an ODE for $\delta m$, which can be solved to provide an expression for $\delta m$,

$$
\delta m = \left( \frac{\partial F}{\partial c} \bigg|_0 \nabla c \bigg|_0 \cdot \bar{v}_r \right) \left( \delta t - \tau_m(A_0) + \tau_m(A_0) e^{-\frac{\delta t}{\tau_m(A_0)}} \right),
$$

where the adaptation time is defined by $\tau_m(A_0) := (\zeta A_0 (1 - A_0) N \gamma)^{-1}$. The average CheY$_p$ concentration, assuming quasi-steady state, is given by

$$
Y^p = \frac{k_Y A Y^T}{k_Y A + k_Z Z + \gamma_Y},
$$

where $Y^T$ and $Z$ are the total concentration of CheY and CheZ, respectively, and $k_Y$, $k_Z$, and $\gamma_Y$ are the phosphorylation, de-phosphorylation and degradation rates of CheY, respectively. Hence, the variation in $Y^p$ is

$$
\delta Y^p = \frac{\partial Y^p}{\partial A} \bigg|_0 \delta A = \frac{\partial Y^p}{\partial A} \bigg|_0 (1 - A_0) A_0 \tau_m(A_0) \left( \frac{\partial F}{\partial c} \bigg|_0 \nabla c \bigg|_0 \cdot \bar{v}_r \right) \left( 1 - e^{-\frac{\delta t}{\tau_m(A_0)}} \right) = \frac{k_Y (\gamma_Y + k_Z Z)}{(\gamma_Y + A_0 k_Y + k_Z Z)^2} Y^T (1 - A_0) A_0 \tau_m(A_0) \left( \frac{\partial F}{\partial c} \bigg|_0 \nabla c \bigg|_0 \cdot \bar{v}_r \right) \left( 1 - e^{-\frac{\delta t}{\tau_m(A_0)}} \right),
$$

where Eq. (B.2) and (B.5) are used in Eq. (B.1). For simplicity, Eq. (B.6) can be written

$$
\delta Y^p = K'(A_0) \left( 1 - e^{-\frac{\delta t}{\tau_m(A_0)}} \right),
$$

where $K'$ is defined by dividing the right-hand side of Eq. (B.6) by $\left( 1 - e^{-\frac{\delta t}{\tau_m(A_0)}} \right)$. For a cell with $M$ motors in which one motor rotating CW is enough to trigger a tumble, the rate of switching from run to tumble state is $M$ times the rate of switching from CCW
CONCLUSIONS

to CW. Thus

\[ k_{t \rightarrow r} = (d - 1)D_s + 1.3Me^{-\left(\frac{20-40}{\sqrt{r+30}}\right)}, \]

where the term \((d - 1)D_s\) accounts for rotational diffusion \((d\) number of dimension). In contrast, all motors need to rotate CCW to trigger a run. Thus, the expression for the rate of switching from a tumble to a run is a function of the state of the motors. However, assuming that most of the time only one motor would be rotating CW during a tumble, the rate of switching from a tumble to a run is equal to the rate of switching from CW to CCW,

\[ k_{t \rightarrow r} = 1.3e^{\left(\frac{20-40}{\sqrt{r+30}}\right)}. \]

The change of rate \(k_{r \rightarrow t}\) in a small time \(\delta t\) is given by

\[ \delta k_{r \rightarrow t} = \frac{\partial k_{r \rightarrow t}}{\partial Y_p} \bigg|_0 \delta Y_p = \frac{\partial k_{r \rightarrow t}}{\partial Y_p} \bigg|_0 K'(A_0) \left(1 - e^{-\frac{\delta t}{\tau_m(A_0)}}\right). \tag{B.7} \]

A similar equation for \(\delta k_{t \rightarrow r}\) exists. The total time of a run, \(\tau_r\), is given by

\[ \tau_r = \int_0^\infty dt \ e^{-\int_0^t dt' k_{r \rightarrow t}}. \tag{B.8} \]

Note, the total time of a tumble, \(\tau_t\), is given similarly using \(k_{t \rightarrow r}\). In the linear regime, the integral in \(t'\) gives

\[ \int_0^t dt' k_{r \rightarrow t} = k_{r \rightarrow t} \bigg|_0 t + \frac{\partial k_{r \rightarrow t}}{\partial Y_p} \bigg|_0 K'(A_0) \left(t - \tau_m(A_0) + \tau_m(A_0)e^{-\frac{t}{\tau_m(A_0)}}\right). \]
Thus, Eq. (B.8) becomes

\[
\tau_r \simeq \int_0^\infty dt \ e^{-k_r \rightarrow t |}_0 \left[ 1 + \frac{\partial k_r \rightarrow t}{\partial \tilde{Y}_p} \bigg|_0 K'(A_0) \left( t - \tau_m(A_0) + \tau_m(A_0)e^{-\tau_m(A_0)} \right) \right]
\]

\[
= \frac{1}{k_r \rightarrow t |}_0 + \frac{\partial k_r \rightarrow t}{\partial \tilde{Y}_p} \bigg|_0 K'(A_0)
\]

\[
= \frac{1}{k_r \rightarrow t |}_0 + \frac{K''(A_0) \nabla c|_0 \cdot \bar{v}_r}{k_r \rightarrow t |}_0 \frac{K'(A_0)}{1 + \tau_m(A_0) k_r \rightarrow t |}_0 ,
\]

(B.9)

valid in the linear regime only with

\[
K''(A_0) := \frac{\partial k_r \rightarrow t}{\partial \tilde{Y}_p} \bigg|_0 \frac{K'(A_0)}{\nabla c|_0 \cdot \bar{v}_r}.
\]

Note that the term \( \nabla c|_0 \cdot \bar{v}_r \) is the scalar product between the gradient and the run velocity, i.e. \( \nabla c|_0 \cdot \bar{v}_r = \nabla c|_0 v_r \cos \theta \), where \( \theta \) is the angle between the gradient and the direction of running of the cell.

The drift velocity in 2d is then given by

\[
\langle \tilde{v}_d \rangle = \frac{\int_0^{2\pi} d\theta \ \tau_{\|}(\theta) \cos(\theta)}{\int_0^{2\pi} d\theta \ \tau_{\perp}(\theta) + \int_0^{2\pi} d\theta \ \tau_{\perp}(\theta) v_0}.
\]

(B.10)

At the numerator, the zero-order terms, which are independent of \( \theta \), vanish, and the full equation for the drift becomes Eq. (4.1), i.e.

\[
\langle \tilde{v}_d \rangle = K(A_0) \frac{\partial F}{\partial c} \bigg|_0 \nabla c|_0 ,
\]
where the full expression of $K(A_0)$,

$$
K(A_0) = \frac{k_{t \rightarrow r | 0} K''(A_0) |v_r|}{d \ k_{r \rightarrow t | 0} (1 + \tau_m(A_0) k_{r \rightarrow t | 0}) \left( (k_{r \rightarrow t | 0} - (d - 1)D_s) + k_{t \rightarrow r | 0} \right)}
$$

$$= \frac{k_{t \rightarrow r | 0} \partial k_{r \rightarrow t | 0} \left( \frac{k_Y(Y + k_Z)}{(Y + A_0 k_Y + k_Z)^2} \right) Y^T (1 - A_0) A_0 \tau_m(A_0) |v_r|^2}{d \ k_{r \rightarrow t | 0} (1 + \tau_m(A_0) k_{r \rightarrow t | 0}) \left( (k_{r \rightarrow t | 0} - (d - 1)D_s) + k_{t \rightarrow r | 0} \right)},
$$

(B.11)

where $d$ is the number of dimensions (see [Lovely and Dahlquist, 1975]).

The activity $A_0$ around which the drift velocity is calculated is normally the adapted activity $A^*$ [Colin et al., 2014]. However, the theory presented here is general and can apply to any $A_0$. Furthermore, knowing the drift velocity in the linear regime, the average activity $\langle A \rangle$ can be written as

$$
\langle A \rangle = A_0 + \delta A
$$

$$\simeq A_0 + (1 - A_0) A_0 \tau_m(A_0) \frac{\partial F}{\partial c} \bigg|_0 \nabla c |_0 \langle \bar{v}_d \rangle.
$$

(B.12)

This equation appears in the system in Eq. (4.2) where the initial activity is considered the adapted activity $A^*$. I checked, using simulations, that Eq. (B.12) matches the average activity of swimming cells. However, when the linear regime is broken, the average activity can be computed using simulations and then used in Eq. (B.11) to calculate the average drift in a semi-analytical model.