Comparing Information Content of T-Cell Receptor Models as Markov Chain Communication Channels

Kevin Hui

A thesis submitted to

the Faculty of Graduate Studies

in partial fulfillment of the requirements

for the Degree of Master of Science

Graduate Program in and Electrical Engineering and Computer Science

York University

Toronto, Ontario

July 2026

© Kevin Hui, 2026

Abstract

T-Cells are an important component in the human immune system as they recognize and target specific types of intruders, while avoiding "self" proteins. The main avenue of the T-cell's activation is through a surface protein called the T-Cell Receptor (TCR). The exact mechanism of triggering TCR activation is currently unknown. In this thesis, we attempt to answer the question: Between two leading models describing TCR activation, known as classic kinetic proofreading (cKP) and kinetic segregation with kinetic proofreading (KS-KP), which one closer describes reality? To answer the question, we will use information theory, as well as techniques in stochastic simulation, parallel computing, and discrete and continuous time dynamics; our information-theoretic analysis indicates that KS-KP is more consistent with the observed performance of non-self ligands. Throughout the thesis we also devise new methods of processing the mathematical models of receptor proteins and discovered new theorems regarding the information capacity of receptor channels.

Table of Contents

Abstract **ii**

Table of Contents **iii**

List of Tables **v**

List of Figures **vi**

1 Introduction **1**

1.1 Introduction 1

1.2 Literature Review 3

1.2.1 Intro to TCR mechanism literature 5

1.2.2 Surveying the Intersection of Biology and Information theory 8

1.2.3 Previous Work Relating TCR with information theory 9

1.2.4 Models of Main Focus 11

1.3 Main Objectives of this Thesis 12

1.4 Our Contributions 13

2 Technical Background **16**

2.1 Markov Chains 16

2.2 Signal transduction 18

2.3 Stochastic processes in Biochemistry 20

2.3.1 Mass Action and Reaction Rates 20

2.3.2 Biochemical Reactions as Stochastic Processes 22
2.4 Information theory 26
2.4.1 Entropy Rate and Mutual Information Rate 29
2.5 Information Theory of Signal Transduction 30

3 Simulation Methodology and Validation 34
3.1 Introduction to Methodology 34
3.1.1 Example: Channelrhodopsin (ChR2) 35
3.1.2 Belief Propagation 37
3.1.3 Code Implementation 39
3.2 The effect of Δt on Receptor Channels 40
3.2.1 Acetylcholine receptor (ACh) 41
3.2.2 Modified ACh 41

4 Information Transmission in Signal Transduction Channels with Hidden Sensitive Transitions 44
4.1 Model and Definitions 45
4.2 Main result 48

5 Mutual Information Rate Maximization Distinguishes Between Two Models of TCR Activation 54
5.1 Implementing the TCR Models 54
5.2 Markov Distribution for Input and Upgrade to Continuous Time Simulation 57
5.2.1 Defining a Markov Matrix by Specifications 57
5.2.2 Simulation Results with Markov Input Distribution 59
5.2.3 Creation of Continuous Time Monte Carlo Simulations 61
5.2.4 Continuous Time Belief Propagation Algorithm for Mutual Information 63
5.3 Mutual Information Rate Shows KS-KP to be More Realistic 67
5.3.1 Mutual Information Landscape over Activating Ligand's Chemical Kinetic Constants 68

5.3.2 The Mutual Information Maximizing Value of k^- and its Internal Maxima . . . 71
5.3.3 Experimentally described TCR activating ligands follows the trend of kinetic parameter pairs that result in optimal mutual information for the KS-KP model . . . 71

6 Conclusion **74**
6.1 Contributions of This Thesis . . . 74
6.2 Future Direction . . . 75

Bibliography **77**

Appendices **81**

A Validation of Program Code **82**
A.1 Distribution of Log Probability Differences of models . . . 82
A.1.1 Acetylcholine receptor (ACh) . . . 82
A.1.2 TCR KS-KP . . . 85
A.2 Mutual Information Rate as $\Delta t \to 0$ of receptors . . . 88
A.2.1 Calmodulin receptor (CaM), i.i.d. input . . . 88
A.2.2 TCR KS-KP . . . 90

B Calculating Matrix Exponential Efficiently **95**

List of Tables

5.1 Specific constants used in our model, based on [1–4]. 70

List of Figures

1.1 The two leading theories of T-cell activation mechanism. The possible states a TCR can assume are represented by the icons with the little hooks. Yellow blob on the hooks are the ligand bound to the TCR and small orange triangles attached to the tail are phosphorylation applied to the tails of the TCR. Arrows between states represent possible transition of states. 4

1.2 The structure of a TCR sitting on top of the membrane of a T-Cell. pMHC on other cells carries ligands for the TCR to recognize and the ζ tails are locations for phorphorylation. 5

2.1 Binary Symmetric Channel (BSC) 28

2.2 State transition graph of Channelrhodopsin (ChR2) from [5]. Dashed boxes group together states, and bold arrow is a transition whose rate is affected by input. 29

3.1 State transition graph of Channelrhodopsin (ChR2). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition. . 36

3.2 State transition graph of Acetylcholine receptor (ACh). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition. . 41

3.3 Mutual information rate of ACh varied by probability of low input p_L and time step length Δt in i.i.d. input. 42

3.4 State transition graph of ACh but with different grouping (blue dashed boxes). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition. 43
3.5 Mutual information rate of ACh varied by probability of low input p_L and time step length Δt in i.i.d. input. Simulated with the alternate grouping where there are no sensitive edge within a group. 43

5.1 State transition graph of TCR cKP and KS-KP. It have phosphorylation threshold $n = 3$, each vertex is a possible state under a model and edges are possible transitions. Yellow is non-activating ligand, Red is activating ligand, blue arrows represents possible state transition, while yellow triangle is phosphate. Under each model, states within the red box form a macrostate group where the receptor is activated, all other states outside the red box are together the non-activated macrostate. 55
5.2 Rate of mutual information of ACh under Markov distributed input, discrete time step system, simulated with $\Delta t = 5 \cdot 10^{-5}$. 59
5.3 Each surface are of slightly different setup, such as different Δt or total simulation time. $\Delta t \in \{0.0002, 0.0005, 0.001, 0.005\}$ 60
5.4 Mutual information per activating ligand of cKP as function of k_a^- on log scale. Different curves represent different k_a^+ values on log scale from 10^{-4} to 10^2. 69
5.5 Mutual information per activating ligand of KS-KP as function of k_a^- on log scale. Different curves represent different k_a^+ values on log scale from 10^{-2} to 10^4. 70
5.6 Internal maxima in the k_a^- axis for each k_a^+ in mutual information per activating ligand, over cKP and KS-KP with different phosphorylation threshold n. Both axes are on log scale. 72

A.1 State transition graph of Acetylcholine receptor (ACh). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition. . 83

A.2 Effect of simulation length on the distribution of Conditional Probability value 84
A.3 Histogram of the simulation output of two models of TCR, 200000 time steps 86
A.4 Probability plot of the simulation output of two models of TCR, 200000 time steps 86
A.5 Histogram of the simulation output of two models of TCR, 200000 time steps 88
A.6 State transition graph of Calmodulin receptor (CaM). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition. . 89
A.7 Mutual information rate of CaM varied by probability of low input p_L and time step length Δt in i.i.d. input. 89
A.8 Mutual info rate of TCR KS-KP, given different Δt 90
A.9 These ones are ran with higher ligand concentration: $M = 5000$. 92
A.10 These ones also have less realistic parameters: $M = 5000$, $k_p = 6$, $k_{dp} = 12$. . 93

Chapter 1

Introduction

1.1 Introduction

Information theory is the study of the quantification, transmission, and storage of information. It is originally derived from and mostly applied to telecommunication, where it is meant to answer fundamental questions such as, "what is the ultimate data compression", and "what is the ultimate transmission rate of communication" (in the words of [6]). However, it has since found use in physics, economics, and biology. Its application to biology is an open field of research and can answer many unsolved questions by seeing them through the lens of communication. Particularly, information theory techniques can analyze the communications of cells and reactions within cells as abstract 'channels' in the same vein as communication in electrical and computer networks can be abstracted to be similar classes of mathematical models.

Among problems in biology, information theory can be used to study the immune system [7]. In this context, T Lymphocytes, or T-cells, are a crucial component in the human immune system. Each T-cell is unique and designed to fight a specific type of intruder. Once activated, the T-cell create many copies of itself to target and defeat the intruder. To start the process, the T-cell is activated by binding with some foreign protein fragments, or ligands, brought to it by Antigen Presenting Cells (APC) [8]. The detection of ligands is

performed with the T-cell Receptor (TCR), a protein on the membrane of the T-cells. This thesis will attempt to explain the exact mechanism by which the T-cell receptors are able to separate ligands from foreign cells to ally cells from the body. TCRs need to discriminate foreign ligands (non-self ligands) from ligands from ally protein produced by the human body (self ligands) which the TCR can also bind to, and occur in much higher amounts in the usual environment that T-cell exists in. Thus, the TCR need to combine high sensitivity with high specificity [1].

Currently two leading theories attempting to explain TCR activation mechanism are Classical Kinetic Proofreading (cKP) and Kinetic Proofreading through Kinetic Segregation (KS-KP). Briefly, cKP posits that the TCR is activated by only the longest binding ligands using time and energy consuming steps that resets immediately upon the ligand unbinding, while KS-KP is based on cKP but includes the assumption that the resets are not instantaneous. It is unclear whether cKP or KS-KP is the most descriptive model for the TCR, or even whether the true mechanism might be modeled by a combination of both systems [1]. Therefore, the main contribution of this thesis is to model the two leading models of TCR activation as communication channels in information theory and comparing with experimental data to demonstrate which or what combination of the two best describe the real mechanism for TCR activation.

The central theme of this thesis is the question: Using techniques in information theory and computational biochemistry simulation, can the exact mechanism of TCR activation be clarified, by finding out which of these two models more adequately describe the true mechanism of TCR? To answer this question, we look to the intersection of biology and information theory, where simulations can be programmed and yield results that can be measured using information theory metrics such as mutual information given variations in parameters. Then, the parameter landscape of different models can be compared to understand which matches experimental results most closely.

1.2 Literature Review

Previous literature has suggested two models for describing the mechanism of TCR activation, namely, the classical Kinetic Proofreading (cKP) model, and the Kinetic Proofreading through Kinetic Segregation (KS-KP) model. Both of these leading models are based on a concept called Kinetic Proofreading (KP), under which a ligand binding to a receptor does not immediately activate it, but instead requiring extra steps of confirmation, usually in the form of extra changes to the receptor protein due to phosphorylation, before being activated. This allow the system to filter non desired ligands falsely activating the receptor as they much weakly binding to the receptor and thus are more likely to detach before the receptor finishes 'confirming' [9]. Likewise, Kinetic Proofreading through Kinetic Segregation (KS-KP) posits that the TCR is constantly being phosphorylated and dephosphorylated, but once coming into close contact with antigen presenting cells which allow for binding to a ligand, the close contact between the two cell membranes will block out enzymes that attempt to dephosphorylate the TCR, since they also have a portion outside the cell with large molecular size and thus can no longer get close to the TCR, thus allowing phosphorylation [4, 10].

Meanwhile, more work has been put into studying biochemical kinetics and systems biology using the perspective of information theory. A few studies of channel capacities and mutual information of biological system have been carried out and showed the mathematics of information theory is greatly applicable to biological systems. [1, 5, 11–13] In particular, the Shannon capacity of signal transduction receptors has been calculated [5], which is the maximum amount of information per channel usage or time interval, that can be conveyed using this channel. Using a finite state Markov chain model, the possible configurations of a receptor protein are organized into a set of states, and transitions among those states are governed by inputs such as light intensity or ligand concentration over time. From these models, information theoretic parameters can be extracted or calculated, which can explain certain phenomena in the system. As the TCR is a signal transduction receptor which operates according to these same principles, this technique will serve as the inspiration this thesis will build upon.

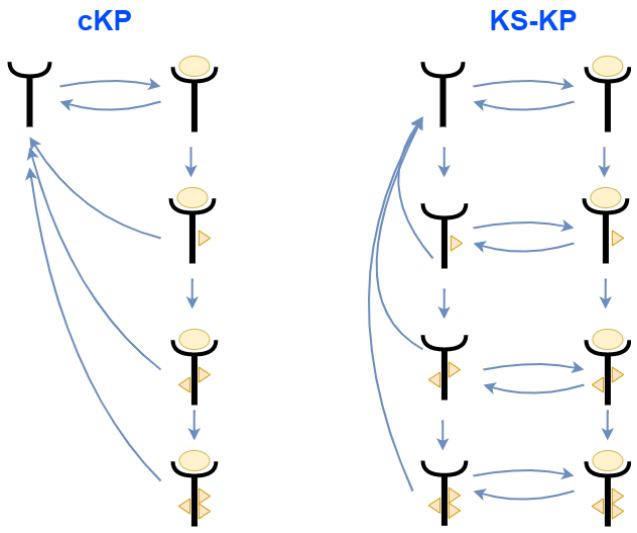


Figure 1.1: The two leading theories of T-cell activation mechanism. The possible states a TCR can assume are represented by the icons with the little hooks. Yellow blob on the hooks are the ligand bound to the TCR and small orange triangles attached to the tail are phosphorylation applied to the tails of the TCR. Arrows between states represent possible transition of states.

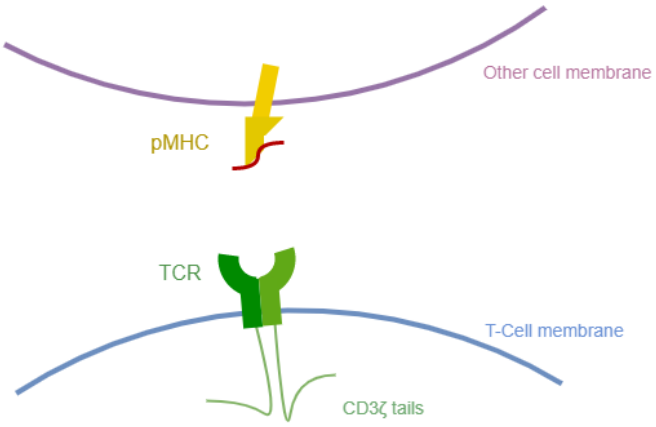


Figure 1.2: The structure of a TCR sitting on top of the membrane of a T-Cell. pMHC on other cells carries ligands for the TCR to recognize and the ζ tails are locations for phorphorylation.

1.2.1 Intro to TCR mechanism literature

Over the years, science has gained better understanding of the mechanism of the chemical kinetics of many biological molecules. For example, it is now possible to describe how biomolecules come together to form metabolic pathways or genetic circuits. Yet, the exact mechanism of the activation of the T-cell receptor remains elusive. Our inquiry into the underlying mechanism of the process that is TCR and T-Cell activation could greatly expand knowledge of the human immune system, as the T-Cell is a key line of response and defense against pathogens. Gaining a full understanding of its molecular mechanism will pave the way for the development of new methods of treatment for infections and cancer immunotherapy in the future.

The T-cell receptor is a protein that sits on the surfaces of T-cells. It consist of a few sub-units assembled and held together by inter-molecular force, including two subunits that faces outside the cell and is where ligands bind to, and two long tails called CD3ζ on the side inside the cell. The sub-units facing outside that come together each can have different variations, where the mRNA is changed to different codes to be translated to become different protein chains, thus allowing for a diverse set of ligand recognition for different intruders [10, 14, 15].

The TCR will bind to ligands attached to some proteins on the outside of a different cell, a structure called peptides bound with Major Histocompatibility Complex (pMHC), which is found on the membranes of many cells in the human body [16]. The CD3ζ tails inside the cell have specific spots called Immunoreceptor Tyrosine-based Activation Motifs (ITAMs), locations at which phosphate ions can be added by other enzymes. This changes the shape of the entire TCR, in a process called phosphorylation. It is currently understood that the phosphorylation of the ITAMs is a essential step to trigger TCR signalling. However, other changes have also been reported, with some being independent to the ITAM phosphorylation while others observed to be related [14]. Here in this thesis, the process of phosphorylation of the tails of the TCR and the steps leading up to it, is highly focused on. The reason being that when the ITAMs in the CD3ζ chain provides a site for a downstream enzyme to bind to, which is a key step to triggering all subsequent reactions of the T-cell being activated. Thus, this is the reason why both cKP and KS-KP ultimately centers around the path towards achieving sufficent amount of phosphorylation to trigger activation.

An overview of the challenges in studying the mechanism as well as the current competing models is given in [14]. Due to the low abundance of intruder cells which carries agonist pMHC ligands, TCR need the sensitivity to react to intruder cells from biding to as little as 1 ligand, while also having the specificity to bind only to that activating ligand out of as many as 10^6 ligands [1, 3, 14].

Currently, a few competing models has been proposed in attempt to fully explain the underlying mechanism of TCR activation. In general, models of TCR triggering generally invoke one or more of the first three basic mechanisms [14]. More recent work has added a fourth basic mechanism called mechanosensing [10]. They can be explained as follows:

1. Conformational change - Conformational change hypothesizes that by binding to a pMHC, the TCR goes through a conformational change, which makes parts of the TCR on the end inside the cell more accessible to kinases that phosphorylates it [14].

2. Segregation or redistribution - Segregation and redistribution proposed that there exists both enzymes that promotes phosphorylation to the TCR-CD3 complex and en-

zymes that promotes dephosphorylation to the TCR-CD3 complex (such as CD45 and CD148) inside the T-cell. This creates a sort of equilibrium that will be disturbed by non-self ligands, which cause reactions that promotes more phosphorylation to the ITAM, which triggers signaling. Interestingly, literature states that when the phosphatases that dephosphorylates TCRs are inhibited, the TCR can trigger signaling and activate the T-cell even when ligands are completely absent [14]. This suggests that the phosphorylation of the TCR through the disruption of such equilibrium must be a sufficient condition for triggering TCR signaling.

Both of the models focused on by this thesis, cKP and KS-KP, are based on the above two ideas. With cKP belonging purely to the conformational change category and KS-KP combining it with segregation.

3. Aggregation - Aggregation or Oligomerization suggests that TCRs that are bound to pMHCs can combine together to form an oligomer, which can then activate the T-Cell [17]. Under it are two subclasses. Co-receptor heterodimerization, which suggests that co-receptors called CD4 and CD8 binds to the same agonist (non-self) pMHC complex as the TCR, thus recruit the associated LCK to phosphorylate the TCR. Meanwhile, the Pseudodimer model suggest that the co-receptors are brought to non-self ligand pMHC bound TCR by another self ligand pMHC bound TCR, which are much more abundant on cell surfaces [14].

4. Mechanosensing - Mechanosensing states that T-cells bind and crawl over APCs to scan their surface for cognate pMHC ligands. The sliding will then generate tensile forces in the molecules similar to stretching an elastic band, which can be leveraged by the TCR to drive biochemical signaling after engaging with pMHCs [10].

Given the above, this thesis will focus on the Conformational change and Segregation categories. In particular, we will attempt to differentiate between classical Kinetic Proofreading (cKP) and Kinetic Proofreading through Kinetic Segregation (KS-KP). They both utilize the conformational change paradigm in a process called Kinetic Proofreading, where

time and energy consuming steps are applied to a TCR bound to a ligand as confirmation steps. These steps will be reset upon the ligand unbinding and serve to filter out weaker binding ligands not intended to be activating sources [9]. Meanwhile, KS-KP further combines this with Segregation, where the membrane of the other cell carrying the ligand can block out enzymes performing the reset to the steps which takes some time after unbinding, as opposed to the reset being spontaneous upon unbinding.

1.2.2 Surveying the Intersection of Biology and Information theory

The application of information theory to fields outside of engineering has gained prominence, especially to biology [18]. Most famously, Rivoire and Leibler [19] popularized the intersection of information theory and population dynamics. Moreover, information theory has also been applied to researching signal transduction. Tkačik et al. first used information theory to model the change in spatial distribution of gene transcription factors during the development of a Drosophila (fruit fly) embryo, and how it can lead to correct separation of head and abdomen [11]. Cheong et al. modelled the interaction of tumor necrosis factor (TNF) signaling pathway to two transcription factors downstream in the reaction chain in cells using channel capacity [20].

Rhee, Cheong, and Levchenko were also among the first to provide a framework for biologists to understand signal and information transduction using information theory. The paper has noted that, "for many biological signaling channels, $p(r|s)$ can be readily experimentally measured, whereas $p(s)$ cannot be easily estimated," where s is the event of the signal source/"sender", and r is the "receiver", meaning, the distribution of response given a controlled signal can be easily measured, but the distribution of the signal itself happening is difficult [21]. The authors noted that an information theoretic framework can be easily implemented in absence of knowledge in the internal mechanism of the biochemical system. Thus, one can apply such techniques and frameworks to study biochemical systems without any understanding or needed references to the underlying molecular dynamics and enzyme

kinetics.

Meanwhile, Uda and Kuroda also gave an important analysis of research work relating signal transduction to topics in information theory and especially channel capacity [22]. They cited and analyzed the aforementioned works [11, 20] as well as the team's own previous work [23], among others to show that information theory is a useful tool for understanding the mechanisms behind signal transduction.

Moreover, Tsuruyama utilized Kullback–Leibler Divergence in communication channels, another related topic in channel coding within information theory, to model the thermodynamics of cell signal transduction cascades, providing an alternative perspective to the usual use of mutual information in this intersection [24]. Likewise, Ito and Sagawa studied the thermodynamics of signal transduction from an information theoretic approach, establishing a solid numerical connection between usual information theory and the information-thermodynamics of biochemical signal transduction [25].

Thomas and Eckford created a more abstract model of generic biochemical signal transduction by modelling the ligand and receptor interaction as a discrete time, finite state Markov chain, with arbitary finite alphabets [26]. Their work showed that the input distribution that will achieve the highest channel capacity will be input signal that are independent and identically distributed (i.i.d.). Such a model will be of great use in the future for studying the mathematical foundations of signal transduction.

1.2.3 Previous Work Relating TCR with information theory

Work has been done relating information theory and molecular biology, and especially with T-cell receptors, which is the main topic of this thesis.

Stewart et al. [27] measured the distribution of amino acid in ligands binding to TCR and Immunoglobulin using Shannon's entropy, and how they related to the physical structure of the bound TCR complex, such as how the nth amino acid in the ligand that strongly bound should statistically be most likely to be certain type of amino acid.

Henderson et al. [28] explored the amount of information the varying of the TCR amino

acid sequence can contribute to its specificity. During T-cell development, the RNA that will be transcripted to the proteins of T-cell components are changed to create variation in structure, which allow different T-cells to bind to and recognize a variety of ligands. The paper provides a framework to model how TCR variations give rise to the recognition of the enormous space of possible ligands, and concludes that variation in the β chain provides the TCR the most amount of its specifity.

Meanwhile, Bene et al. [29] worked on the lipid raft model of TCR activation and combined it with stochastic resonance, which is a phenomenon that describe how noise added to a system can enhance the detection and processing of signals. The paper modelled the process of TCR activation using techniques from signal processing and showed that TCR operates as a threshold detector, like the ones found in analog circuits.

François and Altan-Bonnet [30] touched up on the topic of explaining TCR mechanism with information theory using a system of simulated evolution. The paper simulated a population of cells that can mutate their biological networks, and used mutual information as a fitness metric for the evolutionary selection. The fitness metric is calculated as the mutual information between the distribution of ligand binding time, and the distribution of the output of the biological network, or in other words the probability of an immune response being activated resulting from this ligand with this binding time, under this cell's system. According to the authors, the evolution converged into a process called adaptive sorting, which is a one step kinetic proofreading process that have the phosphorylation step regulated by bound but not phosphorylated receptors. The paper then goes to show that a multi-step adaptive sorting process is likely to be realistic for human TCR mechanism.

Then, more recently, Tserunyan and Finley [31] also used an information theoretic approach to analyze chimeric antigen receptor (CAR) T-cells, a type of modified T-cells involved in an immunotherapy against cancer. They modelled the signal pathway with a network of ordinary differential equations and demonstrated the channel capacity of the process.

1.2.4 Models of Main Focus

For this thesis, Kinetic Proofreading, called classical Kinetic Proofreading (cKP) in this document, and Kinetic Segregation through Kinetic Proofreading (KS-KP), will be the models of main focus, as they are deemed the most likely to be the true mechanism for TCR activation by experimental biochemistry collaborators of our lab [4, 32].

As mentioned, Kinetic Proofreading (KP) is a process where time and energy consuming steps are applied to a TCR bound to a ligand which serves to filter out the weaker binding ligands such that only the ligands with the longest lasting binding time will trigger the activation of the TCR.

Classical Kinetic Proofreading (cKP) suggests that the accuracy and specificity of the discrimination by the TCR is achieved through a multi-step KP process that induce modification to the receptor. Most commonly and also as the case here, each of the steps involve an additional phosphorylation of the receptor, with the downstream signal only being triggered when a sufficient number of steps have been completed. This is reflected in the TCR as the phosphorylation of the ITAMs of the CD3 complex. If the ligand unbinds at any step of the process, the modifications will be lost, thus, only ligands with long binding lifetimes will likely trigger the activation. Since the TCR binds more strongly to nonself ligands and less strongly to ligands from one's own body, thus cKP introduces selectivity to the T-cell activation process, by requiring the ligand to stay bound for a long time [1, 9].

On the other hand, Kinetic Segregation (KS), a subclass of the segregation model,posits that the activation of the TCR is determined by its time spent in 'close contact zones', where Antigen presenting cells (APC) and their pMHC comes into contact with the T-Cell. When the membrane of the two cells come into contact, that allows for the TCR and pMHC to bind together. Moreover, due to the tightness of the cell membranes, a zone that can exclude large molecules is created, a region called the *close contact zone* or merely *close contact*. This causes the enzymes that dephosphorylate the TCR complex, such as CD45, which have a large ectodomain (i.e. the side that extends outside the cell) to be excluded from this close contact zone. Thus, the phosphorylation of the TCR will be favoured [1, 10, 14].

Then, Kinetic Segregation through Kinetic Proofreading (KS-KP) is the combination of the KS and KP model, where ligands bound to the TCR create close contact zones that reduce dephosphorylation but also have their validity reinforced by additional phosphorylation steps, once again only triggering downstream signals when sufficient number of steps have been completed [1]. In other words, unlike cKP, phosphorylation can proceed even if the TCR is not bound to a ligand; this is the key difference between the cKP and KS-KP models.

There is a longstanding debate as to whether cKP or KS-KP is the more accurate model to explain the TCR's selectivity [1, 4, 30, 33].However, a combination of the two models can also possibly be the true mechanism of TCR activation [1]. Thus, we must also identify the contexts in which each model might work better to provide an greater understanding to the subject matter. This debate and set of contexts for each model form the scientific background for this thesis.

In turn, the main focus of this thesis will be to test and clarify wether cKP or KS-KP is the more realistic model to describe the mechanism of TCR activation. In particular, we will do so by evaluating the two models using computational simulations and techniques in information theory to see which model provide metrics more consistent to experimental data. As such the relevant backgrounds will be introduced in the following sections.

1.3 Main Objectives of this Thesis

The main objective of this thesis will be to apply the model of signal transduction systems using information theoretic channel with finite state Markov chains proposed in [5, 26], to study the molecular interaction of the TCR and its downstreams in the biological chain reaction sequence.

Given all of the above citations, one natural step forward occurs: can one measure the process of TCR signaling, as a communication channel, abstracted in the information theory sense? Referring to [5, 26], a model and guideline is given in which future inquiries of the same category can be modelled after. Moreover, in [1] the authors provided even more relevant details to build a sufficient model for studying TCR mechanism as Markov chain

states and differential equations. Likewise, [30] has given important pieces to fill into this puzzle, where a more complete and realistic model can be constructed by including crucial information from them.

However, building such a model to fully describe the mechanism of TCR activation is not a simple task. Firstly, the input alphabet space is not a simple monotonic sliding scale of input intensity, but instead at least equivalent to a 2 element vector, with one component for concentration of self ligands, and the other for concentration of non self ligands. This can potentially be further complicated if the granularity of the model regarding species of ligand molecule or types of TCR variation comes into play, as the different on or off rate of individual specie or class of ligands to different TCR variations must be encapsulated. Secondly, as the full molecular mechanism of the TCR activation process has not been completely elucidated, we had to rely on differential experimentation in our simulations where the action and outcome of different models are compared. Then, the information rate per time of the TCR activation as a channel, and well as the behaviour of the different parameters in the biochemical pathway, simulated under different models, can be established, which can then be compared to experimental data to infer which models are more realistic.

1.4 Our Contributions

In our work, we revisited the receptors described in [5], and recreated the method for calculating the mutual information of a Markov channel receptor model, which after several upgrades and revisions, is then applied to TCR cKP and TCR KS-KP. This will allow us to distinguish between cKP and KS-KP using mutual information and infer which model describes reality more closely. The discoveries along the process of reaching our ultimate result is organized and submitted as 2 papers to conferences in the field of communications and information theory.

Firstly, we program an advanced version of the previous discrete time step simulation and belief propagation algorithms in the form of a massively parallelized system utilizing CUDA kernel API which can be executed efficiently on Nvidia GPU devices. This code is applied to

the receptor systems in [5] and different interpretations of the TCR from [1] as hidden state Markov models. This enable us to uncover phenomena existing in the channel as hidden non stationary Markov chain model that are previously unnoticed due to insufficient exploration limited by computational efficiency.

The upgraded version of simulation and mutual information estimation system then give way to the discovery that receptor channels with a specific structure in its relation between states will actually communicate zero mutual information when time step Δt approaches the limit of 0 and the receptor is given i.i.d. input event distribution over time. Using these observations, we provide a mathematical proof of this phenomenon, which we dubbed the Theory of Hidden Sensitive Channels. This becomes one of our two submitted conference papers. As TCR cKP and TCR KS-KP both belong to this class of models, this prompts us to update our simulation to a system where the input event is persistent, meaning its distribution over time follows a stationary Markov chain or Markov chain like model.

We then observe the that our discrete time model program still require large amount of computational time in terms of how we need to iteratively shrink Δt to observe convergence to obtain the canonical mutual information when $\Delta t \to 0$. Thus, we further upgrade our simulation and belief propagation algorithm to be based on sojourns of states with amorphous time step lengths. This provide us a way to calculate the receptor system's accumulated mutual information over time in a continuous time manner like a continuous integral over infinitesimal steps where Δt by default is 0.

This finally enables us to model and measure the mutual information as a function of chemical kinetic constants' parameter space between TCR cKP and KS-KP. Comparing and contrasting between the two models' mutual information function landscape gives us an idea of what parameters give the maximum mutual information in the system model. We present different ways of comparing the two models based on different search domains within the parameter space and utilizing optimization methods for noisy functions. As such we explored topics such as finding optimal close contact time and occurrence ratios given specific species of ligand molecules given in experimental literature with their kinetic constants. Finally, we

figure out the currently most appropriate way to construct the mutual information landscape, measurement method, and ways to interpret what does the optimal channel capacity entails. Ultimately, by comparing to experimental data from diverse set of literature, we conclude that KS-KP is the more realistic model of the two to describe the true mechanism of TCR activation. This ultimately becomes the second submitted conference paper derived from this thesis.

The remainder of this thesis is organized as follows. In chapter 2 we will give a literature review for an introduction to the biological and mathematical knowledge. In chapter 3 we explain and demonstrate the technical background and setup of the simulation program, as well as provide the preliminary results of our computations and explore the model with expanded parameters and possible configurations. In chapter 4 we present a newly discovered theory built on analyzing observations from computational simulations. This changes the result from previous papers to which we propose a paradigm change that will prove useful in the context of this thesis. Since both TCR cKP and KS-KP belong to cases where this theory is applicable, our research follow this paradigm shift. In chapter 5 we continue to build on this new paradigm and devise novel methods for computational simulations which further improve computational speed. We then present our observations and discoveries utilizing this upgrade, culminating in the argument supported by our simulation result that KS-KP is the more realistic model. Finally, in chapter 6 we give a conclusion to our results and discoveries.

The contributions presented in this thesis has been organized into 2 papers:

- K. Hui, P. J. Thomas, and A. W. Eckford, "Information Transmission in Signal Transduction Channels with Hidden Sensitive Transitions," in International Symposium on Information Theory and Its Applications 2026, Submitted.

- K. Hui, K. A. Ganzinger, and A. W. Eckford, "Mutual Information Rate Maximization Distinguishes Between Two Models of T-Cell Receptor Activation," in ACM NanoCom 2026, Forthcoming.

Chapter 2

Technical Background

In this thesis, cKP and KS-KP will be simulated as non-stationary hidden state Markov chains, whose states changes responding to the rate of transition rules, and such rules are dictated by an input control parameter. The event of the input parameter changing over time along with the observed part of hidden state that is considered the output, will be evaluated as a communication channel. Measuring the maximum capacity of the channels will show us the parameters that forms the optimal environments for the TCR model and provide insight about which is closer to reality by demonstrating consistency with experimental data.

2.1 Markov Chains

Various mathematical tools will be utilized to study the properties of two TCR activation models using computational simulation. Therefore, technical context of topics will be given to better introduce these concepts relevant to our research.

A Markov chain is a discrete stochastic process, or simply a sequence of discrete random events $\{X_n\} = \{x_1, x_2, ..., x_n\}$, where the probability of each event depends only on the

previous outcome. This is called the Markov property [34]. For example:

$$\Pr(x_k|x_{k-1}, x_{k-2}, ..., x_1) = \Pr(x_k|x_{k-1}) \tag{2.1}$$

$$\Pr(x_k, x_{k-1}, x_{k-2}, ..., x_1) = \prod_{j=2}^{k} \Pr(x_j|x_{j-1}) \tag{2.2}$$

A Markov chain can also describe the state of a system changing over time. Thus, in the context of this thesis, the states of a receptor can be viewed as a Markov chain, where the "state" of the receptor might represent a functional configuration of the receptor. For example, ligands, kinases, phosphatases, and other molecules might attach to or remove something from the receptor, thus changing its shape and function; each such configuration is known as the state of the receptor.

We now describe some relevant properties of Markov chains. If it is possible to go from any state to any other state in finite steps, the Markov chain is called irreducible. If the largest common factor of the lengths of different paths from a state to itself is 1, the Markov chain is said to aperiodic [6]. A Markov chain with dead-end states with no way to get out of is called absorbing. We will deal only with irreducible non-absorbing Markov chains in this thesis.

Every Markov chain can be described by a square matrix representing the transition probability between the states, i.e. A matrix P where $P_{ij} = \Pr(X_{n+1} = j|X_n = i)$, where X_n is the state at the nth step. From this, we can also see that each column of a transition probability matrix must add up to 1. Such a matrix is also called a column stochastic matrix.

For any irreducible and aperiodic Markov chains, there exist a stationary distribution, also called steady state distribution, μ, where the the distribution does not change over subsequent steps, meaning $\mu = P\mu$. From any starting distribution $p(x_1)$, the distribution at infinitely many steps $n \to \infty$ will converge to the stationary distribution

$$p(x_{n+1}) = P^n p(x_1) \tag{2.3}$$

$$\lim_{n\to\infty} p(x_{n+1}) = \lim_{n\to\infty} P^n p(x_1) = \mu \tag{2.4}$$

From such definitions, we can also see that μ is a left side eigenvector of P with eigenvalue of 1. In fact, for any Markov matrices P, P will have a largest eigenvalue of 1 and its

corresponding eigenvector will be the stationary state distribution μ. For matrices in column stochastic format, the right side eigenvector should be solved for instead.

For example, we can have a two state Markov chain where each step state 1 have probability a to transition to state 2, and state 2 have probability b to transition to state 1. It can be described by the following matrix:

$$P = \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix}. \tag{2.5}$$

This P have the steady state distribution vector $\mu = [\frac{b}{a+b}, \frac{a}{a+b}]$.

Moreover, for processes that can evolve in continuous time, such the state of the molecules in a biochemical reaction over time where the proteins can occupy different states, such processes can also be 'discretized' into a discrete time Markov chain where each step is the distribution of the state of the protein after every unit time step Δt. For such a system, a matrix R, with elements representing the rate of change from each state to every other, can be constructed. This is referred to as the *rate matrix*. For example, the element $R_{i,j}$ in R denotes the average rate of change from state i to state j per time unit. Then, a Markov transition matrix A can be derived based on the matrix exponential of R:

$$A = e^{\Delta t R} \tag{2.6}$$

where matrix exponential is defined as:

$$e^X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k \tag{2.7}$$

2.2 Signal transduction

In biochemistry and systems biology, signal transduction is the process where information of some event is transmitted through one or more cells using a series of molecular mechanisms, typically resulting in change of cell behaviour. Most commonly, signals exist as concentration of molecules found in the fluid around cells, but can also involve other physical events such

as light [5, 35]. In such systems, proteins that receive and transduce signals are called receptors [36], and molecules that represent a signal and binds to receptors are called ligands [37]. Initially, a cell may detect a chemical signal by encountering a ligand that binds to a receptor protein on the surface of the cell to form a complex which will assume a different shape compared to the idle receptor. Then, the information of this signal will travel through a pathway of a few steps, each with their own signaling molecule and protein that interacts with them. Finally, the signal will lead to a response in the cell, which could take the form of a transcription factor that changes what genes will be printed out as mRNA to be assembled into new proteins, thus changing the behaviour of the cell [37]. Most often, the steps of a signal pathway involve adding or removing a phosphate group to a protein, which changes the protein's shape and thus behaviour [37]. Enzymes that catalyze the adding of a phosphate group are called kinases and ones that catalyzes the removal of phosphate groups are called phosphatases [37].

As a relevant example, one can look into the pathway of the TCR. The TCR is mainly phosphorylated by a kinase called Lck and dephosphorylated by a phosphatase called CD45 [38], in a process that determines whether the ligand the TCR bound to is foreign or ally. This initial step forms the system of TCR activation we study in this thesis. Once the receptor's tail inside the cell is sufficiently phosphorylated, where current literature implies to be 3 or 4 phosphates [1], it is considered 'activated'. An enzyme called ZAP70 can bind to the activated tail, which will then phosphorylate downstream proteins that goes into the nucleus to transcript genes and ultimately lead to the T-cell performing an immune response [38]. However, signaling pathways are rarely a single linear chain of command, and there are a lot more types of proteins involved in the insulin signaling process which is not mentioned in the simple example. These forms complex biological networks that is the main subject of study of thousands of system biologists.

Often, these pathway are modelled as differential equations, where the rate of change of the concentration of each molecule subject to a reaction rate expression, which depends on the concentration of the molecules themselves, forming a first order differential equation

system. The state a receptor protein can be in can also be modelled as a time dependent and continuous time Markov chain process, where the probability of the state of a receptor will assume in the next time step depend only on the current state of the receptor as well as certain extra parameters in the system in current time. Moreover, sometimes these systems falls into a specific class of linear differential equation called master equations, which forms a significant part of the basis of this thesis. Thus, the methods to solve them will be explained below.

2.3 Stochastic processes in Biochemistry

Stochastic processes is widely used in biochemistry to model and simulate interactions between different types of molecules of a system. Between smooth differentiable functions and differential equations to discrete per molecular interaction methods, all are relevant to simulating different aspect of different biochemical processes, which this thesis will utilize for the goal of inferring the underlying mechanisms and parameters of the TCR.

2.3.1 Mass Action and Reaction Rates

In molecular biology and biochemistry, many systems can be described and modeled as stochastic processes. The law of Mass Action states that, given a system of chemical reaction with chemical equation such as:

$$a\mathrm{A} + b\mathrm{B} \rightarrow c\mathrm{C} + d\mathrm{D}. \tag{2.8}$$

If such a system is at equilibrium at a given temperature, then there must be a constant K_{eq} such that

$$K_{eq} = \frac{[\mathrm{C}]^c[\mathrm{D}]^d}{[\mathrm{A}]^a[\mathrm{B}]^b} \tag{2.9}$$

where [C] stands for the concentration of molecule C [39]. When a system is not in equilibrium, K_{eq} is denoted as Q and is called the reaction quotient. As time goes on, the system will tend to become equilibrated, in other words, converge to a state of equilibrium, thus

$Q \to K_{eq}$ [40]. Note that Q is a transient value that changes over time as the reaction in its isolated system reaches equilibrium, while K_{eq} is an asymptotic convergent value that Q will take in an isolated system at infinite time, which can also be defined by the forward and backward rate constants.

Moreover, for a system that is not yet in equilibrium, the rate of reaction is given by:

$$r = k[\mathrm{A}]^m[\mathrm{B}]^n \tag{2.10}$$

with some constant of reaction rate k, and the rate of change of the concentrations of the molecules is thus

$$\frac{d[\mathrm{A}]}{dt} = -a \cdot r, \frac{d[\mathrm{B}]}{dt} = -b \cdot r, \tag{2.11}$$

$$\frac{d[\mathrm{C}]}{dt} = c \cdot r, \frac{d[\mathrm{D}]}{dt} = d \cdot r. \tag{2.12}$$

If the reaction is an elementary reaction, meaning it occurs in a single step of particle collision, then $m = a$ and $n = b$ [40].

These form the basis of differential equation models for simulating the progression of a biochemical process at a high level. Using the appropriate numerical methods, such system can be solved relatively quickly and accurately. However, it implies the system must be perfectly continuous and deterministic, and thus ignore the detail that biochemical process arise from discrete particles of molecules in lower quantity (in the scale of thousands as opposed to mols $(6 \cdot 10^{23})$) and thus is inherently stochastic.

As a relevant example, the reaction of ligand binding to a TCR and creating a bound complex is described by the chemical equation

$$\mathrm{T} + \mathrm{L} \to \mathrm{TL}. \tag{2.13}$$

In this thesis we focus on the perspective of a single TCR, or in the macro scale, we regard a constant amount of TCRs, so the concentration of T is ignored. Thus, the rate of a ligand binding to TCR k^+ is measured in units of $mol^{-1} \cdot s^{-1}$ with the reaction rate expressed as

$$r_{bind} = k^+[\mathrm{L}]. \tag{2.14}$$

The concentration of ligands [L] is measured in mol. Thus, when they are multiplied, one obtain a term with units of a per time rate s^{-1}. This is the rate at which the TCR in context will bind to a ligand when it is present in the environment with this concentration. This $mol^{-1} \cdot s^{-1}$ unit can be seen in all literature describing the binding rate of a TCR ligand.

2.3.2 Biochemical Reactions as Stochastic Processes

We can also simulate many biochemical reactions as stochastic processes. Székely and Burrage [41] gave an overview for applying discrete state stochastic simulations to problems in systems biology and biochemistry. In their review paper, given the assumption of the system being in thermal equilibrium and approximately chemically homogeneous regarding spatial distribution of each type of molecules, they laid out 5 main classes of stochastic simulation methods. They are explained as follows:

Master equation A Master equation is a system of ordinary differential equations that gives an exact description of the behaviour of the Markov process through time. It is of the form:

$$\frac{d\vec{p}}{dt} = A(t)\vec{p} \tag{2.15}$$

Where the matrix A can change over time with values of its elements depending on some independent 'control' parameters. If that is the case, then the system is considered a time varying Markov chain system. Then, simulating a numerical approximation of such equations will give the approximation for the probabilities of all states at any time after starting from given initial conditions. Such a differential equation system can be solved numerically using common methods such as Euler's, implicit Euler's, Runge-Kutta, etc., with relative ease.

Stochastic simulation algorithm (SSA) Stochastic simulation algorithm (SSA), also called Gillespie algorithm after its inventor, follows the same logic as Master equation but

views the system as discrete amount of molecules. As a direct quote from [41]: "In order to simulate one stochastic realisation, we must step along in time, choosing when the next reaction will happen, and what that reaction will be." The process of SSA is as follows [42] : Firstly, define the possible reactions between the molecules and their rate laws, and initial state vector $\vec{p}(0)$ that is the quantity of each type of molecule as integers. Starting from $t = 0$, the beginning of the simulation, for each unit event step that will exist in some time $t \geq 0$, with the state being $\vec{p}(t)$, calculate the rate of each reaction event happening $\{a_1(\vec{p}), a_2(\vec{p}), ..., a_m(\vec{p})\}$. Then, the sum of the rates $r = \sum_{i=1}^{m} a_i(\vec{p})$ is the rate of any event at all happening during this time. Next, we can simulate that the time to the next event Δt happening is drawn from the exponential probability distribution

$$\Delta t \sim f(x; \lambda) = \lambda e^{-\lambda x}; x \geq 0 \tag{2.16}$$

with $\lambda = r$, and the type of reaction event happening will be drawn from the weighted distribution, thus $p(i) = a_i(\vec{p})/r$. Progress the state $\vec{p}(t)$ by the stoichiometry of the reaction where 1 single reaction of that rule had happened to get $\vec{p}(t + \Delta t)$. Likewise, increment time to $t + \Delta t$ and record it. These steps are repeated until maximum iteration count or end time has reached.

Tau-leap Tau-leap is a further level up and is a generalization of SSA without losing its discreteness. Formulated by Gillespie as well [43], its steps are as follows. Define a fixed time interval for steps τ and start with $t = 0$ and initial state of discrete amount of each molecule type $\vec{p}(0)$. At any time t, calculate the rate of each reaction from the current state $\{a_1(\vec{p}), a_2(\vec{p}), ..., a_m(\vec{p})\}$. Then, for each reaction event generate the integer number of instance of that reaction happening in the τ time step, $K_i \sim \text{Pois}(a_i\tau)$ drawn from the Poisson distribution. Then, modify the state vector by

$$\vec{p}(t + \tau) = \vec{p}(t) + \sum_i K_i c_i \tag{2.17}$$

Where c_i is the unit stoichiometric change of the ith reaction. This is repeated until the finishing time is reached. The base Tau-leap method can thus be seen as the discrete stochastic

version of using Euler's method to solve a differential equation, and bridges the gap between SSA and differential equations. Since it can simulate and group many individual reaction events in each advancement of computational step, it is also easier to predict the execution time of this type of simulations.

Higher Order Tau-leap What a higher order tau-leap method is to the basic tau-leap method, is basically the equivalent of higher order numerical methods for differential equations is to the basic Euler's method. Notice that given a differential system

$$\frac{d\vec{p}(t)}{dt} = A(\vec{p}(t)) \tag{2.18}$$

that can be approximated by Euler's method as:

$$\vec{p}(t + \Delta t) = \vec{p}(t) + \Delta t A(\vec{p}(t)) \tag{2.19}$$

the tau-leap method stochastically emulates a discrete version of it by approximating the derivative at time t:

$$\Delta t \cdot A(\vec{p}(t)) = \tau \cdot A(\vec{p}(t)) \approx \sum_i [\sim \text{Pois}(a_i \tau)] \cdot c_i \tag{2.20}$$

to get a stochastic discrete approximation of the short period change.

Given this, a higher order tau-leap method is the translation of higher order numerical methods for DEs, such as mid point and Runge-Kutta, into discrete versions of the same numerical system by converting one or more of the sub steps of the high order method adhering to this principle.

For example, the first such method higher order than basic tau-leap is the midpoint tau-leap method. It is also described in the original tau-leap paper [43], its steps are as follows: Given a time step size τ, current state $\vec{p}(t)$ at time t. First compute the expected change to the state at half of the time step $\vec{q} = \frac{\tau}{2} \sum_i a_i(\vec{p}(t)) c_i$, and the expected state at half of the time step:

$$\vec{p'} = \vec{p}(t + \tau/2) \approx \vec{p}(t) + \vec{q}. \tag{2.21}$$

Then, generate the quantities of each reaction as discrete amount of instances according to the halfway state.

$$L_i \sim \text{Pois}(a_i(\vec{p}') \cdot \tau) \tag{2.22}$$

Finally, multiply them to the stoichiometric unit change and apply this as the overall change to the state over the entire time step

$$\vec{p}(t+\tau) = \vec{p}(t) + \sum_i L_i c_i. \tag{2.23}$$

Multiscale Methods Finally, multiscale methods are as the name suggests, methods that combine different basic methods each suitable for the scale of a portion of the system. In many practical cases, a biochemical dynamics system have both fast and relatively high order reactions (also called "stiff") and slower reactions involving lesser amount of individual molecules. Simulating the progression of both under the same methods and time step resolutions will be computationally wasteful, as the fast changing part of the fast reaction can require many small time steps and settles quickly in between the appropriate length of steps for the slower reaction, before it will even have any substantial change. Oftentimes, the system is separated into two or more parts. The fast and stiff reactions are often modelled with deterministic ordinary differential equations, often with implicit numerical methods for numerical stability, to prevent negative concentration values. Then, the slower reactions are modelled with more accurate granular methods such as tau-leap or SSA. Each parts are simulated semi-independently, and work in their own time resolution scale, hence the name multiscale methods. Multiscale methods allows for efficient use of computational power, but comes at a cost of some accuracy as errors can be introduced when integrating the layers [41].

Usage of the Methods The methods listed here will have applications as the implementations of the finite state chemical kinetics models. Simulating the overall distributions of the states overtime will utilize a non-stationary Master equation. If we want to simulate individual observation sequences of the states over discrete time steps, the master equation will be transformed to a discrete step Markov chain and a system similar to Tau-leap is used.

And later we will see a method based on the Gillespie algorithm can be used to simulate an improved version of the discrete time steps system where steps are measured in sojourns in states of amorphous time lengths.

2.4 Information theory

Now, we look at the basics of information theory, which will be utilized in this thesis to model and describe our biochemical systems.

The most basic quantity in information theory is entropy, which quantifies how much information is needed to describe the outcomes of a random event (random variable) X, denoted $H(X)$ and given as follows:

$$H(X) = -\sum_{x \in X} p(x) \log_2(p(x)). \tag{2.24}$$

It is the weighted average of the logarithm of each outcome's probability, and is most commonly measured in bits with the logarithm being in base 2. The random variable can represent anything that can have multiple outcomes or states. For example, an event can be the result of a coin flip $\{H, T\}$, the analog voltage on a wire, which is a continuous variable, or, in the context of molecular communication and this thesis, the state of a protein $\{0, 1, 2, 3\}$, where the number represent how many phosphate groups are attached to it, where each possibility will cause the protein to assume a distinct shape and thus a slightly different function.

Then, the next step is to quantify the mutual information between two random events X and Y that are related, denoted $I(X;Y)$:

$$I(X;Y) = \sum_{x,y \in X,Y} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{2.25}$$

$$= H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{2.26}$$

This describes the amount of information one would attain about outcomes of Y by knowing the outcomes of X, or vice versa. If X and Y are independent, then $p(x,y) =$

$p(x)p(y)$, thus

$$I(X;Y) = \sum_{x,y \in X,Y} p(x,y) \log_2 \left(\frac{p(x)p(y)}{p(x)p(y)} \right) \tag{2.27}$$

$$= \sum_{x,y \in X,Y} p(x,y) \log_2(1) \tag{2.28}$$

$$= 0. \tag{2.29}$$

This means one cannot infer anything about the outcomes of Y by knowing the outcomes of X, or vice versa.

Then, a channel of communication can be described as two random variables, X and Y for the input and output of the channel. The communication channel can have any set of alphabets, and does not have to be binary or even be the same set. One use of the channel involves sender sending some letter in the input set $x \in X$, and the receiver receiving some letter in the output set $y \in Y$. The two events are related but not necessarily the same, as in basically all practical cases, there is noise in the communication and not every message letter are necessarily perfectly transmitted.

An example of a communication channel is the Binary Symmetric Channel, which stem from an electrical engineering context and can be described as follows. In the Binary Symmetric Channel (BSC), the input and output alphabet are both binary digits, thus $\{0, 1\}$. Over the medium of communication, there is a probability p with $0 < p < \frac{1}{2}$ where the bit is flipped from 1 to 0 or 0 to 1, and with probability $1 - p$ the bit stays unchanged across the channel.

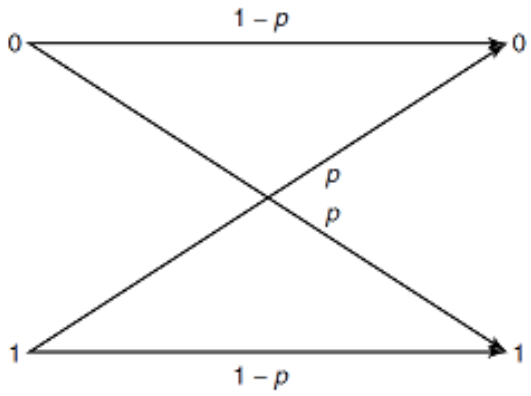


Figure 2.1: Binary Symmetric Channel (BSC)

Thus, mathematically we can express the channel with input event X and output event Y, where they are related by the following expressions using conditional probabilities:

$$\Pr(Y = 0|X = 0) = 1 - p \tag{2.30}$$

$$\Pr(Y = 1|X = 0) = p \tag{2.31}$$

$$\Pr(Y = 0|X = 1) = p \tag{2.32}$$

$$\Pr(Y = 1|X = 1) = 1 - p \tag{2.33}$$

$$\tag{2.34}$$

Or in a different sense:

$$\Pr(Y = X) = 1 - p \tag{2.35}$$

$$\Pr(Y \neq X) = p \tag{2.36}$$

$$\tag{2.37}$$

Then, given any model of communication channels such as the example above, the maximum capacity of the channel C, is given by:

$$C = \max_{p(x)} I(X;Y) \tag{2.38}$$

In bits of information per channel use. Detailed proof of this can be found in most information theory textbooks.

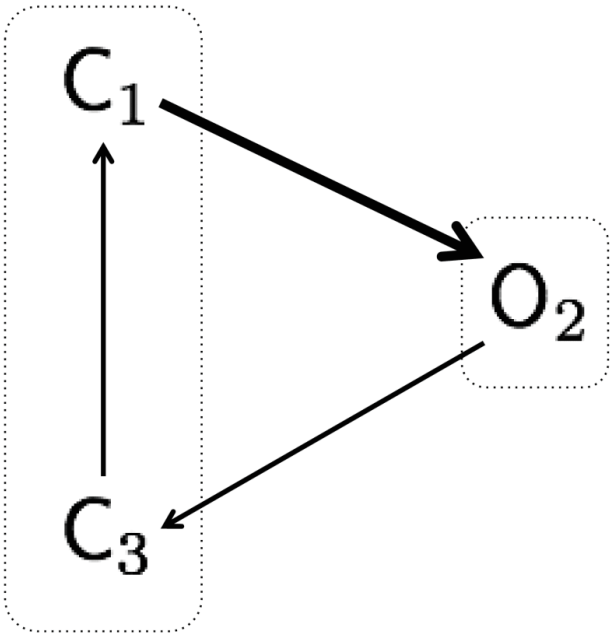


Figure 2.2: State transition graph of Channelrhodopsin (ChR2) from [5]. Dashed boxes group together states, and bold arrow is a transition whose rate is affected by input.

Therefore, in the context of this thesis, a communication channel in molecular biology can represent a pathway of signal transduction. For the example of ChR2 [5] in figure 2.2, the input alphabet X is of the set $\{0, 1\}$, representing the different states of the input signal of which is the intensity of light. This influence the rate of transition from state C_1 to O_2, denoted as the bolded arrow in figure 2.2. The output alphabet Y is of the set $\{C, O\}$, representing if the state of the receptor belong to 'closed' ($\{C_1, C_3\}$) or 'open' ($\{O_2\}$).

2.4.1 Entropy Rate and Mutual Information Rate

Furthermore, given a sequence of n random variables $X_1^n = \{X_1, X_2, ..., X_n\}$, the entropy rate of the sequence is defined as:

$$\mathcal{H}(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, ..., X_n), \tag{2.39}$$

or simply put, the amount of information needed to describe each outcome X_i over the n outcomes or steps, as n tends to infinity. If X is identically and independently distributed,

it is not hard to see $\mathcal{H}(X) = H(X)$, echoing the base definition of entropy.

Likewise, the entropy rate of a Markov chain with transition matrix P and stationary distribution vector μ, can be given by:

$$\mathcal{H}(X) = -\sum_i \mu_i (\sum_j P_{ij} \log(P_{ij})). \tag{2.40}$$

This is the average of the conditional entropy of the event of individual transitions, weighted over the stationary distribution. This is proven in section 4.2 in [6].

Moreover, as $n \to \infty$, the mutual information between consecutive states can be given by [44]:

$$I(X_n; X_{n-1}) = H(\mu) - H(X_n|X_{n-1}) \tag{2.41}$$

$$= H(\mu) - \mathcal{H}(X). \tag{2.42}$$

Likewise, the mutual information rate of two sequences of events $X_1^n = \{X_1, X_2, ..., X_n\}$ and $Y_1^n = \{Y_1, Y_2, ..., Y_n\}$ is defined as:

$$\mathcal{I}(X;Y) = \lim_{n\to\infty} \frac{1}{n} I(X_1^n; Y_1^n) \tag{2.43}$$

$$= \lim_{n\to\infty} \frac{1}{n} I(X_1, X_2, ..., X_n; Y_1, Y_2, ..., Y_n) \tag{2.44}$$

$$= \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\log\left(\frac{p(Y_1^n|X_1^n)}{p(Y_1^n)}\right)\right]. \tag{2.45}$$

This definition will serve as basis for the metric of measuring the capacities of all receptor channels in this thesis. Moreover, later in this thesis we will see rates of information per time and per sojourn on a same input instead of per sequence item.

2.5 Information Theory of Signal Transduction

The authors of [26] expanded upon this idea to create a general model of receptor activation called the "Binding IN Discrete time", or BIND channel. The paper has proposed a communication channel model with the input alphabet being discrete levels of ligand concentration with a defined lowest and highest level, $\mathcal{X} \in \{c_L, c_2, c_3, ..., c_H\}$, and the output alphabet

being the state of the receptor. Thus, if the receptor has only 1 ligand slot and can only be bounded or unbounded, then $\mathcal{Y} \in \{U, B\}$. This model forms a Markov chain with time variant that is capable of describing most cases in signal transduction.

We know that the rate of ligands binding to a receptor is a function of its concentration, and a bounded complex unbinding is not dependent on the ligand concentration. Then, there must be some values that are rates of ligand binding to the receptor a_i for each ligand concentration level c_i, and the rate of the bounded complex unbinding b. Thus, given some time interval unit, the state of the receptor will be a time varying Markov chain with transition matrix (that varies with current ligand concentration):

$$P_{Y|X=c_j} = \begin{bmatrix} 1-\alpha_j & \beta_j \\ \alpha & 1-\beta \end{bmatrix} \tag{2.46}$$

where α_j is the probability where, with current ligand concentration c_j, the unbound receptor becomes bounded after unit time interval.

The authors had shown that, to achieve maximum capacity, the input distribution must be independent and identically distributed, and the probability mass function must be all on the lowest or highest ligand concentration, i.e. $p(X = c_i) = 0$ for all $i = 2, 3, ..., n-1$. The above model will have a maximal channel capacity of:

$$C = \max_{p_H} \frac{\mathcal{H}(p_L\alpha_L + p_H\alpha_H) - p_L\mathcal{H}(\alpha_L) - p_H\mathcal{H}(\alpha_H)}{1 + (p_L\alpha_L + p_H\alpha_H)/\beta} \tag{2.47}$$

where $\mathcal{H}(p) = -p\log(p) - (1-p)\log(1-p)$ is the binary distribution entropy where probability of one of the outcome is p.

Related ideas are further developed by the same authors [5], where different receptors of varying complexity are programmed as models that helps explain their mechanism. The paper laid out the model of the three receptors (Channelrhodopsin-2, ChR2; Acetylcholine, ACh; and Calmodulin, CaM) as follows: Given that a receptor can have k states. The transition rate matrices are $k \times k$ matrices where some terms are dependent on signal strength (ligand concentration or light intensity), which then can change over time, denoted $Q(x(t))$. Likewise, the probability distribution of the states the receptors will be in over time is

denoted as a k element vector $p(t)$, where each element represent a state. This model of receptor kinetics forms a differential equation system called a Master equation, as mentioned in equation 2.15. The paper used a first order numerical method to approximate the solution of this differential equation:

$$p(t + \Delta t) = p(t) + \Delta t p(t)Q + o(\Delta t) \tag{2.48}$$

$$= p(t) \cdot (I + \Delta t Q) + o(\Delta t). \tag{2.49}$$

Taking $I + \Delta t Q$ as the state change operator matrix. This is thus analogous to the discretization of continuous time Markov chain by deriving the transition operator of a discrete step, but only retaining the first order term of the matrix exponential.

$$e^{\Delta t \cdot Q} = I + \Delta t \cdot Q + \frac{1}{2!}\Delta t^2 \cdot Q^2 + \frac{1}{3!}\Delta t^3 \cdot Q^3 + ... \tag{2.50}$$

$$= I + \Delta t \cdot Q + o(\Delta t) \tag{2.51}$$

We refer to [5], which formulated that the overall channel as an abstract mathematical object really exists as 3 different probabilistic events, X, S, and Y. X serve as the input to the channel which is the concentration of the receptor's ligand, in tiers such as $\{c_L, c_2, c_3, ..., c_H\}$ as treated as a discrete variable as in the sources. Y is the 'intermediate' output of the channel which is the 'true' state or fully observed state of the protein, in which multiple fully observed states can have the same chemical function and these states within a group are indistinguishable from each other in lab experiments. Finally, Y represents the 'actual' output of the receptor, which is the actual function observable by other parts of the cell in the organism or the experimenter measuring in a lab. Thus, $Y = f(S)$ is also a deterministic many to one mapping from S.

For example, Acetylcholine has 5 fully observed states, $\mathcal{S} = \{O_1, O_2, C_3, C_4, C_5\}$, but the first two states and the last three states each serve the same functions of 'open' and 'closed'. Thus it have 2 observable states $\mathcal{Y} = \{O, C\}$, which forms a many to one mapping from the 5 fully observed states.

Using analytic derivations and Monte Carlo simulations, the authors are able to plot functions of rates of mutual information $\mathcal{I}(X;S)$ and $\mathcal{I}(X;Y)$ over different probability of

low input intensity $p(X = c_L)$, which is de facto the only independent varying parameter in models with only distribution in highest and lowest concentration level is considered, since $p(X = c_L) + p(X = c_H) = 1$.

Chapter 3

Simulation Methodology and Validation

3.1 Introduction to Methodology

To explore receptors as information channels, we look to [5] as our guideline. An *Intensity Driven Channel* is a receptor protein in a biological systems where it can assume multiple different conformations, also called states. The transitions between the states forms a *State Graph* with which the receptor system's state perform a random walk on with rates of transitioning influenced by an input parameter. The input parameter is usually the concentration of a molecule the receptor is intended to detect, but can also be other types of physical input such as intensity of light [5].

Mathematically, the channel is modelled by a *Non-stationary Master's Equation*, which becomes a *Non-stationary Hidden Markov Chain* when calculated with discrete time steps. Let X be the value of the input parameter to the channel, $\mathcal{S}$ be the set of fully observed states, the states set that fully describes all conformation information of the receptor model we examine, and the distribution of the fully observed state laid out as a vector $\vec{s}$. Then $R(X)$ is the rate matrix dictating the rate of transition between state given the current input value X. The system is thus defined by the Master's Equation as in (2.15).

Then, the system becomes a hidden model since not all information about the receptor's state can be observed. The set of fully observed states $\mathcal{S}$ is considered the 'complete' set of states, capturing all possible microstates of the receptor. However, in practical applications, the microstates, or full information of the receptor's state, are not observable, and instead only their effect is observable.

In many systems, the practical output of the channel Y are elements of a observable, or 'macrostates' states set $\mathcal{Y}$ related to but different from $\mathcal{S}$. At any moment in time, Y is a mapping $h(S)$ through some many to one function $h : \mathcal{S} \to \mathcal{Y}$

Thus, the system is considered an informational theoretic channel with input event X and output event Y, and our goal is to examine the Shannon capacity of such channels by calculating the mutual information $I(X;Y)$ which is defined as over all possible distributions of the events X and Y.

3.1.1 Example: Channelrhodopsin (ChR2)

For example, according to [5], the receptor called Channelrhodopsin (ChR2) have 3 states, $\mathcal{S} = \{C_1, O_2, C_3\}$, with the transition graph shown in figure 3.1 and dashed boxes show the grouping of fully observed states to observable states.

In the state graph, the bold arrow represents the transition that is affected by the scalar input parameter x, which can be the intensity of light of a wavelength range, or the concentration of a ligand molecule in the proximal environment to the receptor. The rate matrix $R(x)$ is then a function of x and linear in x, and can be written as $R(x) = U + xV$, where U is the 'base' rate matrix describing the rate of transition between the microstates when intensity is at a minimum, or $x = 0$, and V is the matrix of additional rates per unit of x. Realistically, a receptor's x can have a range of intensity values between a minimum and maximum intensity changing over time. However, as shown in [26] only the minimum and the maximum intensity contribute to the calculation of the maximum capacity of the channel. Thus, all program implementation from here will only consider x as a discrete event with 2 cases, where the set of input events is $\mathcal{X} = \{L, H\}$, with L standing for input being

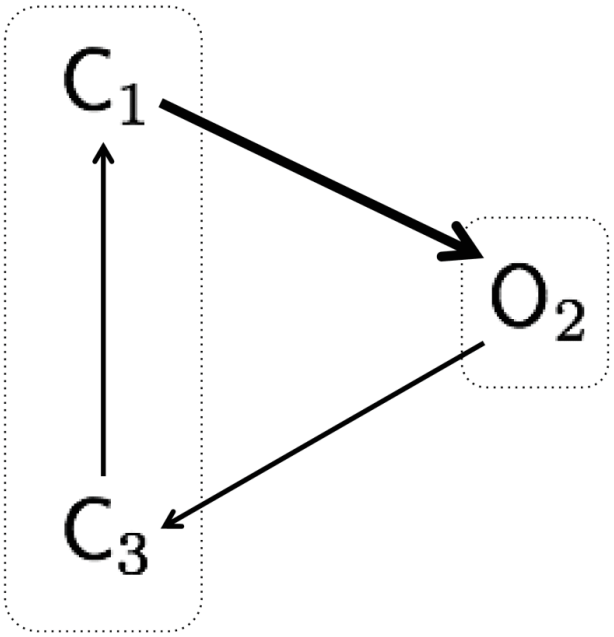


Figure 3.1: State transition graph of Channelrhodopsin (ChR2). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition.

in low (minimum) intensity and H for high (maximum) intensity. Thus, from now on the rate matrix with input x will be denoted with subscript as R_x since x is discrete, as opposed to denoting R as a function of a real scalar x.

Thus, following the example of ChR2 above, it will have rate matrix as follows:

$$R_x = \begin{bmatrix} -q_{1,2}x & 0 & q_{3,1} \\ q_{1,2}x & -q_{2,3} & 0 \\ 0 & q_{2,3} & -q_{3,1} \end{bmatrix}. \tag{3.1}$$

In the initial stage of this thesis, we follow the system in [5] and implement our channels practically as Markov chain with discrete time steps of length Δt. Then, between each simulation step, the Markov matrix A_x representing the probabilities of each state transitioning to another, is the matrix exponential of the rate matrix.

$$A_x = \exp(\Delta t R_x) \tag{3.2}$$

The rate of mutual information per seconds is formally defined as follows:

$$\mathcal{I}(X;Y) = \lim_{m\to\infty} \frac{1}{m\Delta t}\mathbb{E}\left[\log\left(\frac{p(Y_1^m|X_1^m)}{p(Y_1^m)}\right)\right]. \tag{3.3}$$

Here m is the number of time steps in a sequence and X_1^m and Y_1^m represent the random event of the set of possible X and Y event sequences of length m progressed with step length Δt, with the subscript 1 and superscript m representing the indices at the start and end of the sequence. However, it is clearly unrealistic to compute infinitely many steps of simulations and of all cases in the tree of possible events.

3.1.2 Belief Propagation

To calculate the mutual information between the input and output of the system, we employ the Monte Carlo method with the *Belief Propagation* (BP) algorithm [45]. We start with running a Monte Carlo simulation where each each instant of the ensemble is a specific individual receptor that assume a specific element of its fully observed states set in a specific point in time. Let W_U be the Markov matrix of the distribution of the joint event $\mathcal{X} \times \mathcal{S}$, and π_0 be the eigenvector of W_U with eigenvalue 1, which is the steady state distribution of the joint states. Each simulation stream starts with a draw of the choice distribution from π_0 as the initial condition in the first tick, and progress in time and change states following according to the A_X matrices, which depends on the value of input X at that time step. A simulation instance will end when a predetermined maximum step count m is reached, and output two sequences of event observations x_1^m and y_1^m.

We then calculate the probabilities for an instance's result happening $P(y_1^m|x_1^m)$ and $P(y_1^m)$ using the BP algorithm. The BP algorithm to calculate $\log(\Pr(y_1^n|x_1^n))$ can be de-

scribed as follows:

$$\alpha(s_1) = p(s_1|x_1, s_0) \qquad = p(s_1|x_1) \tag{3.4}$$

$$\alpha(s_2) = \sum_{s_1} p(s_2|x_2, s_1)p(y_1|s_1)\alpha(s_1) \qquad = p(s_2, y_1|x_1^2) \tag{3.5}$$

$$\alpha(s_3) = \sum_{s_2} p(s_3|x_3, s_2)p(y_2|s_2)\alpha(s_2) \qquad = p(s_3, y_1^2|x_1^3) \tag{3.6}$$

$$\vdots \tag{3.7}$$

$$\alpha(s_i) = \sum_{s_{i-1}} p(s_i|x_i, s_{i-1})p(y_{i-1}|s_{i-1})\alpha(s_{i-1}) \qquad = p(s_i, y_1^{i-1}|x_1^i) \tag{3.8}$$

$$\alpha(s_{k+1}) = \sum_{s_k} p(s_{k+1}|x_{k+1}, s_k)p(y_k|s_k)\alpha(s_k) = p(s_{k+1}, y_1^k|x_1^{k+1}), \tag{3.9}$$

Where $\alpha(s_k)$ is a message vector representing the distribution of the states at step k.

Practically, we first define a mask matrix which is expanded from a mask vector, $Q_{y_i} = \text{diag}(q_{y_i})$ to filter for observable output state at the ith time step. Its elements are indicator functions of if the hidden state correspond to current partially observed output, thus $(q_{y_i})_k = 1$ if $h(S_k) = y_i$, else $(q_{y_i})_k = 0$, for each $S_k \in \mathcal{S}$. Given channel input and output sequences y_1^n, x_1^n from a simulated output, $Pr(y_1^n|x_1^n)$ can be calculated as an accumulated product as follows. At the ith step, the Markov stochastic matrix $A_{x_i} = \exp(\Delta t R_{x_i})$ along with mask Q_{y_i} are applied to the message vector $\hat{\alpha}(s_{i-1})$ representing the probability of each fully observed state. It is then renormalized, with the norm being the accumulated conditional probability whose log is to be summed.

$$\beta_i = Q_{y_i} A_{x_i} \hat{\alpha}(s_{i-1}) \tag{3.10}$$

$$\nu_i = \|\beta_i\|_1 \tag{3.11}$$

$$= \begin{cases} \Pr(s_i|s_1^{i-1}) & \text{for } p(Y) \text{ stream} \\ \Pr(s_i|s_1^{i-1}, x_1^i) & \text{for } p(Y|X) \text{ stream} \end{cases} \tag{3.12}$$

$$\hat{\alpha}(s_i) = \frac{\beta_i}{\|\beta_i\|_1}, \tag{3.13}$$

The combined accumulated probability after k steps will then be:

$$\log(\Pr(s_{i+k}|s_1^{i-1})) = \sum_{j=i}^{i+k} \log(\nu_j). \tag{3.14}$$

Here $\hat{\alpha}(s_i)$ is the message vector re-normalized to have a norm of 1 in each iteration. This is how the algorithm is computationally implemented otherwise the floats will underflow to just 0 as they accumulate the product of small values.

In essence, the BP algorithm iterates through the sequence of observations generated by a simulation instance y_1^n, and at each step calculates the probability of the sub-sequence up to the kth step happening $\Pr(y_1^k)$, with the uncertainty due to the many to one mapping stored as a belief vector $a(s_k)$. The probabilities will be accumulated by the sum of logarithms to ultimately obtain $\log(\Pr(y_1^n|x_1^n))$ and $\log(\Pr(y_1^n))$ of a simulation result instance.

Once a large amount of simulation instances are ran, we can calculate the difference of logs:

$$\log\left(\frac{\Pr(y_1^n|x_1^n)}{\Pr(y_1^n)}\right) = \log(\Pr(y_1^n|x_1^n)) - \log(\Pr(y_1^n)), \tag{3.15}$$

Then, expression (3.3) can ultimately be approximated by taking the mean of all cases of simulation stream and dividing by total time elapsed each stream $n\Delta t$. This will give a rate of information per time (bits/seconds), and this will be the maximum capacity of which information can be communicated using this receptor per unit of time.

3.1.3 Code Implementation

The simulation and inference are coded according to the same code from [5], and utilized the `Numba` package in Python to program GPU kernels that can be executed on CUDA GPUs for fast parallel inference.

The validity of the new code, especially the validity of specific program functions are confirmed by comparing the outputs given the same input arrays between new code (both a GPU kernel and CPU numpy version are created) and code from [5]. The output values observed are numerically the same given the same small sample inputs.

Likewise, the helper functions such as `bigXYmatrix` that generate the join transition distribution matrix, and `getSteadyStateDist` and `getNonStationarySteadyStateDist` which extract the eigen vector steady state distribution given a Markov matrix, are also verified to be faithful recreations.

However, the Markov matrices differ slightly. The new code contains the methods for generating the receptor state transition Markov matrix, given a discretization time step Δt and value for low/high inputs, for Acetylcholine (ACh), using exact same parameters and units (distinction between seconds and milliseconds), but uses the new operator $A = exp(\Delta tR)$ instead of the old one $A = I + \Delta tR$, for generating time step matrix (Markov matrix, column sums to 1) A from rate matrix (column sums to 0) R. It also contained a modernized version with seconds as unit and exponential conversion operator, for ACh and the TCR under cKP and KS-KP models. As mentioned above, when $\Delta t \to 0$, the Taylor series that generates the exponential of a matrix $A = exp(\Delta tR)$ becomes equivalent to $A = I + \Delta tR$ since all terms at after the first order of Δt vanishes.

With a GPU powered system programmed, we receive a huge increase in computational power and efficiency allowing us to explore simulations larger in scale that are previously unrealistic. We explored more receptor channel along with TCR, with the result of validation detailed in appendix A.

3.2 The effect of Δt on Receptor Channels

According to [5, 26], for any receptor channel system, the rate of mutual information will converge to a canonical maximum value as $\Delta t \to 0$. Thus, within each possible configuration under each model, we estimate the rate of mutual information using the Monte Carlo method, repeating multiple runs with progressively smaller Δt. Once signs of convergence is observed we can conclude our test which outputs an estimation that will be close to the 'true' value of the maximum capacity of the channel.

We test our simulation program with receptor models mentioned in literature, most

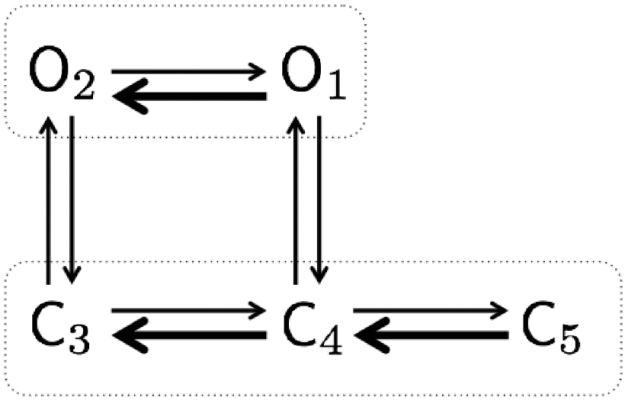


Figure 3.2: State transition graph of Acetylcholine receptor (ACh). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition.

prominently the Acetylcholine receptor (ACh), as well as other models from literature that are easy to tinker with and interpret, to examine if this effect exist in other models as well.

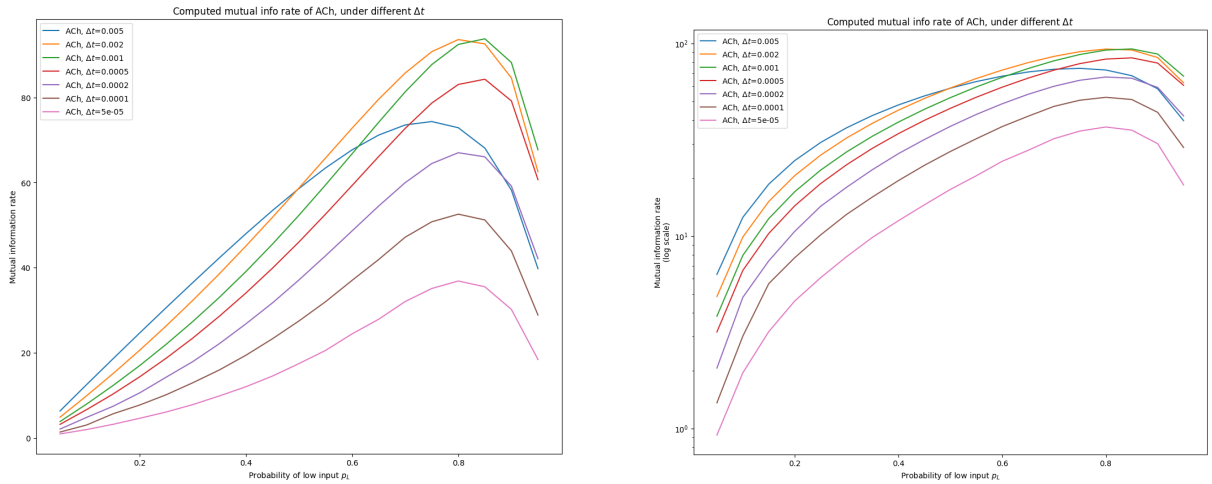
3.2.1 Acetylcholine receptor (ACh)

As shown in figure 3.2, ACh is a receptor channel with 5 states, and 3 sensitive transitions embedded between the two 'open' states or the three 'closed' states. Running our simulation on this system yields the results in figure 3.3.

Figure 3.3 demonstrates that the mutual information rate of ACh reaches a maximum as Δt is shrinking, then tends toward 0 as $\Delta t \to 0$. This is thought to be impossible as [26] stated that the maximum capacity achieving input distribution will be i.i.d. with time step $\Delta t \to 0$ and proved so for the case of models with fully observed states.

3.2.2 Modified ACh

We now come to raising a new hypothesis: Let a transition between two states in a hidden Markov receptor channel that are affected by the input value be a *sensitive transition*. And let a transition that is between two microstates belonging to the same macrostate called a *hidden transition*. Define a hidden sensitive channel as a receptor channels where all sensitive



(a) Mutual information rate of ACh (linear scale)

(b) Mutual information rate of ACh (log scale)

Figure 3.3: Mutual information rate of ACh varied by probability of low input p_L and time step length Δt in i.i.d. input.

transitions are hidden transitions. Then, if a receptor channel is a hidden sensitive channel with i.i.d. input, its rate of mutual information converges to 0 when $\Delta t \to 0$.

Since ACh have only hidden sensitive transitions, its rate of mutual information becomes 0 as $\Delta t \to 0$. Now we modify ACh with alternative state grouping of $\{\{O_2, C_3\}, \{O_1, C_4\}, \{C_5\}\}$ (in code, `mapping = [[0, 3], [1, 2], [4,]]`) instead of the standard $\{\{O_1, O_2\}, \{C_3, C_4, C_5\}\}$ (in code, `mapping = [[0, 1], [2, 3, 4]]`), such that there now does not exist any sensitive edge within a group. Its state graph can be seen in figure 3.4.

As seen in figure 3.5, the sequence of curves of mutual information rate shows signs of converging to a non-zero value when $\Delta t \to 0$, for a system that is not hidden sensitive. This supports our hypothesis but does not rigorously confirms it. Thus, in the next chapter, we give a formal proof to the *Theory of Hidden Sensitive Channels*.

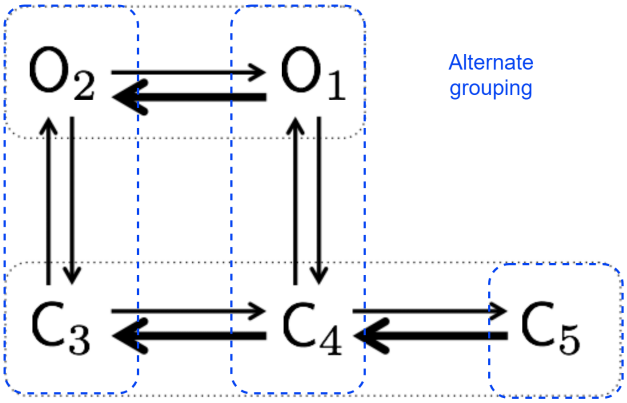
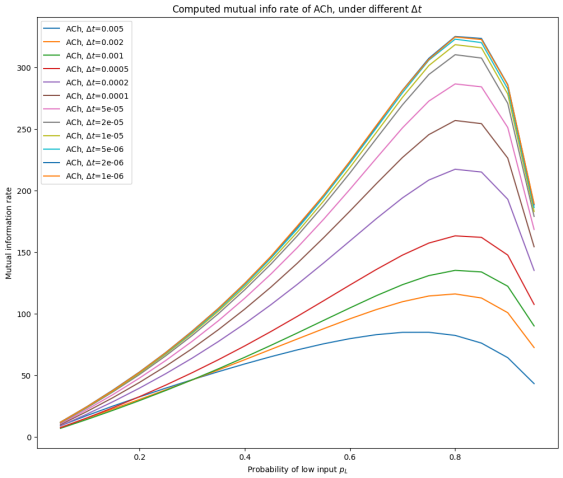
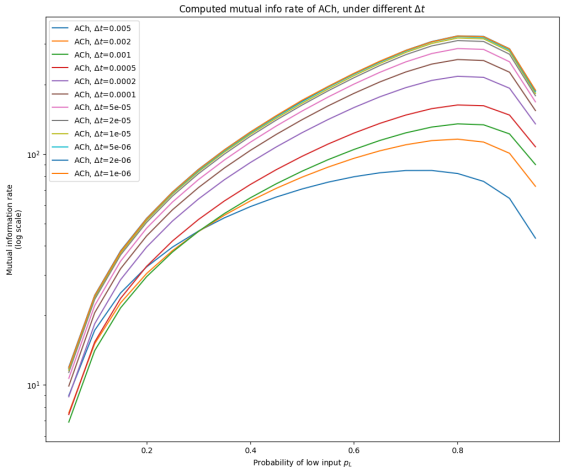


Figure 3.4: State transition graph of ACh but with different grouping (blue dashed boxes). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition.



(a) Mutual information rate of ACh



(b) Mutual information rate of ACh

Figure 3.5: Mutual information rate of ACh varied by probability of low input p_L and time step length Δt in i.i.d. input. Simulated with the alternate grouping where there are no sensitive edge within a group.

Chapter 4

Information Transmission in Signal Transduction Channels with Hidden Sensitive Transitions

Given the observations in section 3.2, we hypothesize that a hidden Markov channel with i.i.d. input over discrete time steps will have its mutual information rate fall to 0 as time step length $\Delta t \to 0$, if the channel's sensitive transitions are all hidden transitions within some group belong to the same macrostate. In particular, both models of TCR as well as ACh are hidden sensitive channels, while ChR2, CaM, and the modified ACh we devised are not.

This deviates from the previous assumption from [26] that the maximum capacity achieving input distribution of all Markov receptor channels should be i.i.d.. The source proved the case for receptor channels with fully observed states and assumed same for channels with states that are hidden by the observable mapping. However, we will soon show that it is not the case.

In this chapter we will formalize the definition of hidden sensitive channels and the definition of the conjecture stated above. Then, we provide a proof of the this conjecture, which we call hidden sensitive channel theorem.

4.1 Model and Definitions

Let $\mathcal{S}$ represent the set of receptor states. Let x_t represent the stimulus at time t, and let $W(x_t)$ be a $|\mathcal{S}| \times |\mathcal{S}|$ matrix of rates: $W(x_t) = [w_{ji}(x_t)]$, where $w_{ji}(x_t)$ is the transition rate from state i to state j. Let Q be an $|\mathcal{S}|$-dimensional vector of occupancy probabilities of each state, then Q satisfies the master equation as mentioned in equation (2.15), which is also known as *Kolmogorov's forward equation.*

Transition rates are proportional to the magnitude of the stimulus: the rate matrix $W(x_t)$ can be written

$$W(x_t) = U + x_t V, \tag{4.1}$$

where $U = [u_{ji}]$ is a matrix of "background" rates independent of x_t, and $V = [v_{ji}]$ is a matrix of rates proportional to x_t.

As noted in chapter 3, states can be grouped by an observable property. Let Y_t be an observable property associated with the state at time t, $s_t \in \mathcal{S}$; for simplicity, we will assume $Y_t \in \{0, 1\}$. There exists a subset of states $\mathcal{A} \subset \mathcal{S}$ and a function $\phi : \mathcal{S} \to \{0, 1\}$ satisfying

$$Y_t = \phi(s_t) = \begin{cases} 1, & s_t \in \mathcal{A}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

We give the following definitions:

Definition 1. *A* sensitive *transition is defined as a pair* (i, j) *with* $v_{ji} > 0$.

Definition 2. *A* hidden *transition is defined as a pair* (i, j) *where* $\phi(i) = \phi(j)$*, i.e., not observable. Otherwise, the transition is* observable.

In the remainder of the chapter we are particularly interested in *hidden*, *sensitive* transitions.

Following the approach of previous work [5, 26], our analytical approach is to develop a discrete-time model of this system and consider the continuous-time limit. Define $\Delta t > 0$ be a time step; then the above system becomes a discrete-time Markov chain with transition

probability matrix

$$P_i = I + \Delta t W(x_i) + o(\Delta t) \tag{4.3}$$

$$= I + \Delta t U + \Delta t x_i V + o(\Delta t), \tag{4.4}$$

where: (4.4) follows from (4.3) and (4.1); x_i refers to the stimulus at index i (i.e. on the interval from $i\Delta t$ to $(i+1)\Delta t$), and $o(\Delta t)$ is Bachmann-Landau little-o notation. Similarly, we use y_i and s_i to refer to the observation and state at index i. With the discrete-time model in mind, we define the information-theoretic channel with an input sequence $X_1^n = [X_1, X_2, \ldots, X_n]$ (i.e. the stimulus), and output sequence $Y_1^n = [Y_1, Y_2, \ldots, Y_n]$ (i.e., the observable feature of the receptor's state).

Concretely, the input to the channel is the observation of a random event changing over time, taken from a discrete set of possibilities $\mathcal{X} = \{\mathrm{X}_1, ...\mathrm{X}_{|\mathcal{X}|}\}$. Then, the state of the non-stationary Markov chain S is dictated by a family of transition matrices. The transition probabilities between each state are a set of $|\mathcal{X}|$ stochastic matrices $A_{\mathrm{X}_1}, ..., A_{\mathrm{X}_{|\mathcal{X}|}}$. Note the non-italic X denoting event possibility set elements.

We define $X_1^n = [X_1, X_2, ..., X_n]$ the input event sequence vector, or inputs to the receptor system over time, which corresponds to $x_1^n = [x_1, x_2, ..., x_n]$, stimulus intensity values in each tick which are numerical scalars, and $Y_1^n = [Y_1, Y_2, ...Y_n]$, the output sequence vector, the macro state/observable state over time. Here the mutual information rate of the two finite sequences of events is defined by the following expectation value expression:

$$I(X_1^n, Y_1^n) = \mathbb{E}\left[\log\left(\frac{p(Y_1^n|X_1^n)}{p(Y_1^n)}\right)\right]. \tag{4.5}$$

In practical cases, $p(Y_1^n|X_1^n)$ and $p(Y_1^n)$ are calculated from the results of each runs of a Monte Carlo simulation using the belief propagation algorithm. The sample mean of many runs becomes the estimation of the mutual information rate $\mathcal{I}(X_1^n, Y_1^n)$, as discussed in section 3.1.2.

For some time T, and let $n_T \triangleq \lfloor T/\Delta t \rfloor$; we will then let $n_T \Delta t$ be the total duration of the communication session. In the continuous limit as $\Delta t \to 0$, the information rate of the

receptor channel per time is defined as

$$\begin{aligned}
&\mathcal{I}(X;Y) \\
&= \lim_{T\to\infty}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{1}{n_T}I(X_1^{n_T};Y_1^{n_T}) &(4.6)\\
&= \lim_{T\to\infty}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{1}{n_T}\mathbb{E}\left[\log\left(\frac{p(Y_1^{n_T}|X_1^{n_T})}{p(Y_1^{n_T})}\right)\right]. &(4.7)
\end{aligned}$$

The capacity is then the maximum of $\mathcal{I}(X;Y)$ over all possible input distributions $p(x_1^{n_T})$. Note that this definition of mutual information rate differ from the original definition in (2.45). Since we now wish to measure rate of mutual information per time instead of per item in the sequence of discrete time observations, we further divide the mutual information value by the length of time step Δt. Our methodology in calculating (4.7) is to use *Monte Carlo* simulation, generating numerous trajectories $x_1^{n_T}$ and $y_1^{n_T}$, and calculating their probabilities $p(y_1^{n_T}|x_1^{n_T})$ and $p(y_1^{n_T})$.

As $x_1^{n_T}$ and $y_1^{n_T}$ jointly form a hidden Markov model, we employ belief propagation [45] to efficiently calculate the probabilities. Define α_k as a N-dimensional vector of state probabilities at time index k. Given an initial value α_0, the ith element of α_k evolves as

$$\alpha_{k+1}^{(i)} = \sum_j \Pr(s_{k+1}=i \mid s_k=j)\alpha_k^{(j)}\chi_{\mathcal{A}_{Y_k}}(i), \tag{4.8}$$

where $\chi_{\mathcal{A}_{Y_k}}(i)$ is the indicator function on $\mathcal{A}_{Y_k}$, where $\chi_{\mathcal{A}_{Y_k}}(i)=1$ if $i\in\mathcal{A}_{Y_k}$, and 0 otherwise, as mentioned in (4.2). If x_1^k is known, we write the conditional probability vector as $\hat{\alpha}_k$, and (4.8) becomes

$$\hat{\alpha}_{k+1}^{(i)} = \sum_j \Pr(s_{k+1}=i \mid x_{k+1}, s_k=j)\hat{\alpha}_k^{(j)}\chi_{\mathcal{A}_{Y_k}}(i). \tag{4.9}$$

Thus, α_n and $\hat{\alpha}_n$ can be calculated by repeatedly applying (4.8) and (4.9), respectively. At the end of the BP algorithm after n steps, we output the probabilities of the event sequence by $\alpha_n\vec{1}=\Pr(Y_1^n)$, and $\hat{\alpha}_n\vec{1}=\Pr(Y_1^n \mid X_1^n)$, where $\vec{1}$ is a length-N all-one vector [45].

If every state in $\mathcal{S}$ is distinct and observable (known as the *fully observed case*), then [5, 26] showed that the capacity-achieving input distribution is i.i.d.. For the equivalent continuous-time system, it has been explicitly shown that the capacity-achieving distribution

is implemented by a telegraph process in the limit as switching rates approach infinity [46], effectively a continuous-time analog of the i.i.d. distribution. However, in systems in which every sensitive transition is a hidden transition, we show in the next section that the mutual information rate for i.i.d. inputs as $\Delta t \to 0$ is, surprisingly, zero.

4.2 Main result

Our main theoretical result is the following:

Theorem 1. *Suppose $X_1^{n_T}$ is an i.i.d. Bernoulli random process, and $Y_1^{n_T}$ is generated as described above. Furthermore suppose all sensitive transitions in $W(x_i)$ are hidden transitions. For such a system:*

1. *As $\Delta t \to 0$, $\hat{\alpha}_k \to \alpha_k$ in probability; and*
2. *$\mathcal{I}(X;Y) = 0$ (as defined in (4.7)).*

Proof. First, define a *sojourn* as follows: the indices $[a, a+1, a+2, \ldots, b]$ form a sojourn if $Y_a = Y_{a+1} = \ldots = Y_b$. Thus, the entire sequence $Y_1, \ldots, Y_{n_T}$ is made up of a series of sojourns.

Inductive Hypothesis For any $a, b \geq 0$, observe the following cases:

- If $\hat{\alpha}_a = \alpha_a$, then $\hat{\alpha}_b = \alpha_b$ if the time steps $[a, ..., b]$ form a sojourn where Y stays the same, or;
- When a sojourn ends at step a as $Y_a \neq Y_{a+1}$, if $\hat{\alpha}_a = \alpha_a$, then $\hat{\alpha}_{a+1} = \alpha_{a+1}$.

Inductive Step: Sojourn Consider a sojourn of Y_k for $k \in \{a, a+1, \ldots, b\}$, where at the beginning of the sojourn, $\alpha_a = \hat{\alpha}_a$. Our general approach in this proof is the following. In part a) of the proof, we show that, if $\alpha_a = \hat{\alpha}_b$, then $\alpha_b \to \hat{\alpha}_b$ in probability. In part b), we construct an inductive argument to show that this property remains true for a sequence of sojourns, proving part 1 of the theorem. Finally, in part c), we show that part 2 of the theorem follows from part 1.

Convergence in probability of $\alpha_b \to \hat{\alpha}_b$ Notice that (4.9) can be performed as a matrix multiplication:

$$\hat{\alpha}_{k+1} = H_{Y_k} P_k \hat{\alpha}_k, \tag{4.10}$$

where P_k is the time-varying transition probability matrix in (4.4), and

$$H_{Y_k} \triangleq \text{diag}\Big(\Big[\chi_{\mathcal{A}_{Y_k}}(1), \chi_{\mathcal{A}_{Y_k}}(2), \ldots, \chi_{\mathcal{A}_{Y_k}}(N)\Big]\Big). \tag{4.11}$$

Then we have:

$$\begin{aligned}
\hat{\alpha}_{k+1} &= H_{Y_k} P_k \hat{\alpha}_k && (4.12)\\
&= H_{Y_k}(I + \Delta t U + \Delta t x_k V + o(\Delta t))\hat{\alpha}_k && (4.13)\\
&= H_{Y_k}\hat{\alpha}_k + \Delta t H_{Y_k} U \hat{\alpha}_k + \Delta t x_k H_{Y_k} V \hat{\alpha}_k + o(\Delta t). && (4.14)
\end{aligned}$$

By assumption, index k is in a sojourn of constant Y_k, so $\hat{\alpha}_k H_{Y_k} = \hat{\alpha}_k$. Thus,

$$\begin{aligned}
\Delta\hat{\alpha}_k &\triangleq \hat{\alpha}_{k+1} - \hat{\alpha}_k && (4.15)\\
&= \Delta t H_{Y_k} U \hat{\alpha}_k + \Delta t x_k H_{Y_k} V \hat{\alpha}_k + o(\Delta t). && (4.16)
\end{aligned}$$

Let $\tilde{x}_k = x_k - \mu$, where $\mu \triangleq E[x]$. Then,

$$\begin{aligned}
x_k &= \tilde{x}_k + \mu && (4.17)\\
\Delta\hat{\alpha}_k &= \Delta t H_{Y_k} U \hat{\alpha}_k + \Delta t x_k H_{Y_k} V \hat{\alpha}_k + o(\Delta t) && (4.18)\\
&= \Delta t H_{Y_k} U \hat{\alpha}_k + \Delta t(\tilde{x}_k + \mu) H_{Y_k} V \hat{\alpha}_k + o(\Delta t) && (4.19)\\
&= \Delta t H_{Y_k} U \hat{\alpha}_k + \Delta t \tilde{x}_k H_{Y_k} V \hat{\alpha}_k \\
&\quad + \Delta t \mu H_{Y_k} V \hat{\alpha}_k + o(\Delta t) && (4.20)\\
&= \Delta t H_{Y_k}(U + \mu V)\hat{\alpha}_k + \Delta t \tilde{x}_k H_{Y_k} V \hat{\alpha}_k + o(\Delta t). && (4.21)
\end{aligned}$$

Following a similar argument for α_k, we can see that,

$$\Delta\alpha_k = \Delta t H_{Y_k}(U + \mu V)\alpha_k + o(\Delta t). \tag{4.22}$$

Combining (4.22) and (4.21), for a process initialized with $\hat{\alpha}_k = \alpha_k$, we have,

$$\Delta\hat{\alpha}_k = \Delta\alpha_k + \Delta t\tilde{x}_k H_{Y_k} V\alpha_k + o(\Delta t). \tag{4.23}$$

Now consider the evolution of α_k over the entire sojourn $k \in \{a, a+1, \ldots, b\}$:

$$\begin{aligned}
&\hat{\alpha}_b \\
&= \hat{\alpha}_a + \sum_{k=a+1}^{b} \Delta\hat{\alpha}_k && (4.24)\\
&= \hat{\alpha}_a + \sum_{k=a+1}^{b} \Delta\alpha_k + \sum_{k=a+1}^{b} \Delta t\tilde{x}_k H_{Y_k} V\alpha_k + \sum_{k=a+1}^{b} o(\Delta t) && (4.25)\\
&= \hat{\alpha}_a + (\alpha_b - \alpha_a) + \sum_{k=a+1}^{b} \Delta t\tilde{x}_k H_{Y_k} V\alpha_k + \sum_{k=a+1}^{b} o(\Delta t) && (4.26)
\end{aligned}$$

where (4.25) follows from (4.23).

First consider the additive term in $o(\Delta t)$ in (4.26). Since the sojourn length is upper bounded by $n_T = \lfloor T/\Delta t \rfloor \leq T/\Delta t$, this term is upper bounded by $T\frac{o(\Delta t)}{\Delta t}$. Under Bachmann-Landau notation, by definition $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

Now consider the additive term in $\tilde{x}_k$ in (4.26). For each k, let $\beta_k \triangleq \alpha_k V H_{Y_k}$; then β_k is an N-dimensional vector. For the ith element $\beta_k^{(i)}$, there exists $\beta_{\max}$ such that $\beta_k^{(i)} < \beta_{\max}$ since each of α_k, V, and H_{Y_k} is bounded. The random variable $\beta_k^{(i)} x_k$ has $E[\beta_k^{(i)} x_k] = 0$ (because $E[\tilde{x}_k] = 0$), and variance

$$\begin{aligned}
\mathrm{Var}(\beta_k^{(i)}\tilde{x}_k) &= \mathrm{Var}(\Delta t\tilde{x}_k H_{Y_k} V\alpha_k) && (4.27)\\
&= (\Delta t H_{Y_k} V\alpha_k)^2 \mathrm{Var}(\tilde{x}_k) && (4.28)\\
&\leq \Big(\Delta t\beta_{\max}\Big)^2 \mathrm{Var}(\tilde{x}_k). && (4.29)
\end{aligned}$$

Furthermore, we know that for any random events A, B, if A and B are independent,

$\mathrm{Var}(A+B) = \mathrm{Var}(A) + \mathrm{Var}(B)$. Since $\tilde{x}_k$ is an i.i.d. random process, by assumption, then

$$\mathrm{Var}\left(\sum_{k=a+1}^{b} \beta_k^{(i)} \tilde{x}_k \Delta t\right) = \sum_{k=a+1}^{b} (\Delta t \beta_k^{(i)})^2 \mathrm{Var}\left(\tilde{x}_k\right) \tag{4.30}$$

$$\leq (b-a+1)\Big(\Delta t \beta_{\max}\Big)^2 \mathrm{Var}(\tilde{x}_k) \tag{4.31}$$

$$\leq \frac{T}{\Delta t}\Big(\Delta t \beta_{\max}\Big)^2 \mathrm{Var}(\tilde{x}_k) \tag{4.32}$$

$$= \Delta t T \beta_{\max}^2 \mathrm{Var}(\tilde{x}_k). \tag{4.33}$$

where (4.32) follows since the sojourn length is upper bounded by $n_T = \lfloor T/\Delta t \rfloor \leq T/\Delta t$, and similarly, for $(b-a+1)\cdot o(\Delta t)$

From (4.33), since the mean and variance of the additive term both approach zero, we conclude that as $\Delta t \to 0$, (4.26) converges in probability to

$$\hat{\alpha}_b \to \hat{\alpha}_a + (\alpha_b - \alpha_a), \tag{4.34}$$

By assumption, the initial condition for the sojourn was $\hat{\alpha}_a = \alpha_a$. Thus,

$$\hat{\alpha}_b \to \alpha_b. \tag{4.35}$$

Sequences of sojourns We will prove by induction that, for intensity channels where all sensitive transitions are hidden transitions, $\hat{\alpha}_n \to \alpha_n$ as $\Delta t \to 0$ for any amount of simulation steps n.

First, the base case: we consider the convergence of α_k and $\hat{\alpha}_k$ within the first sojourn, where $a = 0$ and $b > 0$. By assumption, initially $\hat{\alpha}_0 = \alpha_0 = \pi_0$, where π_0 is the steady state distribution of the Markov chain. Under the argument in part a) of the proof, setting $a = 0$, we have that $\hat{\alpha}_b \to \alpha_b$ in probability, i.e., for any $\epsilon, \delta > 0$ such that

$$\lim_{\Delta t \to 0} \Pr\Big(|\hat{\alpha}_b - \alpha_b| > \delta\Big) < \epsilon \tag{4.36}$$

Second, the inductive case: We first show the effect of the change in observable. Consider an index $b' > 1$ where $a = b' + 1$ and $y_{b'} \neq y_a$; then b' is at the end of a sojourn, and a

is at the beginning of the next one. Assume that $\hat{\alpha}_{b'} \to \alpha_{b'}$ in probability (the *inductive assumption*). When a sojourn ends due to $y_{b'} \neq y_a$, we have

$$\hat{\alpha}_a = H_{Y_a} P_a \hat{\alpha}_{b'} \tag{4.37}$$

$$= H_{Y_a}(I + \Delta t U + \Delta t x_a V)\hat{\alpha}_{b'} + o(\Delta t) \tag{4.38}$$

$$= H_{Y_a}\hat{\alpha}_{b'} + \Delta t H_{Y_a} U \hat{\alpha}_{b'} + \Delta t x_a H_{Y_a} V \hat{\alpha}_{b'} + o(\Delta t) \tag{4.39}$$

and we know that

$$\hat{\alpha}_{b'} = H_{Y_{b'}}\hat{\alpha}_{b'}. \tag{4.40}$$

Since $Y_{b'} \neq Y_a$, they cover disjoint subsets of $\mathcal{S}$, thus we know $H_{Y_a} H_{Y_{b'}} = \mathbf{0}$, the all zero matrix. Thus,

$$\hat{\alpha}_a = H_{Y_a}\hat{\alpha}_{b'} + \Delta t H_{Y_a} U \hat{\alpha}_{b'} \tag{4.41}$$

$$+ \Delta t x_a H_{Y_a} V \hat{\alpha}_{b'} + o(\Delta t) \tag{4.42}$$

$$= H_{Y_a} H_{Y_{b'}}\hat{\alpha}_{b'} + \Delta t H_{Y_a} U H_{Y_{b'}} \hat{\alpha}_{b'} \tag{4.43}$$

$$+ \Delta t x_a H_{Y_a} V H_{Y_{b'}} \hat{\alpha}_{b'} + o(\Delta t) \tag{4.44}$$

$$= \Delta t H_{Y_a} U H_{Y_{b'}} \hat{\alpha}_{b'} + \Delta t x_a H_{Y_a} V H_{Y_{b'}} \hat{\alpha}_{b'} + o(\Delta t) \tag{4.45}$$

$$= \Delta t H_{Y_a}(U + x_a V) H_{Y_{b'}} \hat{\alpha}_{b'} + o(\Delta t). \tag{4.46}$$

Likewise, following the same derivation as (4.41-4.46) but substituting $\hat{\alpha}_k$ and x_k with α_k and μ, we have:

$$\alpha_a = \Delta t H_{Y_a}(U + \mu V) H_{Y_{b'}} \alpha_{b'} + o(\Delta t). \tag{4.47}$$

But since our system only has hidden sensitive transitions, any element v_{ji} of V, if i and j belong to different sets between $\mathcal{A}$ and $\bar{\mathcal{A}}$, $v_{ji} = 0$. Thus $H_{Y_a} V H_{Y_{b'}} = \mathbf{0}$.

$$\hat{\alpha}_a = \Delta t H_{Y_a}(U + x_a V) H_{Y_{b'}} \hat{\alpha}_{b'} + o(\Delta t) \tag{4.48}$$

$$= \Delta t H_{Y_a} U H_{Y_{b'}} \hat{\alpha}_{b'} + o(\Delta t) \tag{4.49}$$

$$\alpha_a = \Delta t H_{Y_a}(U + \mu V) H_{Y_{b'}} \alpha_{b'} + o(\Delta t) \tag{4.50}$$

$$= \Delta t H_{Y_a} U H_{Y_{b'}} \alpha_{b'} + o(\Delta t). \tag{4.51}$$

Therefore, if $\hat{\alpha}_{b'} \to \alpha_{b'}$ in probability,

$$\hat{\alpha}_a \to \alpha_a. \tag{4.52}$$

Finally, to complete the induction, if $\hat{\alpha}_a \to \alpha_a$, and time index a is the beginning of a sojourn that ends at time index b, by the same argument as in the base case, it follows that $\hat{\alpha}_b \to \alpha_b$.

All rate matrices $W(x_i)$ are finite, meaning their elements are bounded by some maximum rate r. At each discrete step of length Δt, there is a maximum $r\Delta t$ probability to transition to a different output state each tick. Then, the average length of a sojourn δ in amount of time steps is bounded by $O(1/r\Delta t)$. As T increases, the number of sojourn scales by $T/\delta = o(Tr\Delta t)$.

Since the number of sojourns scales as $o(T\Delta t)$, by induction it is true that $\hat{\alpha}_{n_T} \to \alpha_{n_T}$. This proves the first part of the theorem.

Mutual information Thus, for any finite T, for all possible cases of X and Y observations, the belief vectors have $\hat{\alpha}_{n_T} \to \alpha_{n_T}$ as $\Delta t \to 0$, thus we have:

$$\begin{aligned} I\Big(X_1^{n_T}; Y_1^{n_T}\Big) &= \mathbb{E}\Big[\log(\Pr(Y_1^{n_T})) - \log(\Pr(Y_1^{n_T}|X_1^{n_T}))\Big] && (4.53)\\ &= \mathbb{E}\Big[\log(\alpha_{n_T}\vec{1}) - \log(\hat{\alpha}_{n_T}\vec{1})\Big] && (4.54)\\ &= 0 && (4.55) \end{aligned}$$

and thus, even as T approaches infinity, since $n_T \triangleq \lfloor T/\Delta t \rfloor$, for some constant c,

$$\begin{aligned} \mathcal{I}(X;Y) &= \lim_{T\to\infty}\lim_{\Delta t\to 0} \frac{1}{\Delta t}\frac{1}{n_T} I\Big(X_1^{n_T}; Y_1^{n_T}\Big) && (4.56)\\ &\le \lim_{T\to\infty}\lim_{\Delta t\to 0} c\frac{1}{\Delta t}\frac{\Delta t}{T} I\Big(X_1^{n_T}; Y_1^{n_T}\Big) && (4.57)\\ &= 0. && (4.58) \end{aligned}$$

which proves the second part of the theorem.

□

Chapter 5

Mutual Information Rate Maximization Distinguishes Between Two Models of TCR Activation

5.1 Implementing the TCR Models

We use an interpretation for both TCR models where we consider that it can bind to 2 possible species of ligands. The first is a ligand from self cells that should be non-activating, and the second is from intruder cells that should be activating. The ligands' binding rates are denoted as k_n^+ and k_a^+, and unbinding rates are denoted as k_n^- and k_a^-, where the subscripts denote non-activating and activating, respectively. In this model, the T-Cell is considered to be constantly contacting other cells. During 'low' input state, the T-Cell is in close contact with cells presenting non-activating ligands, while during 'high' input, the T-cell is in close contact with cells presenting activating ligands. Thus, the state of the TCR can go down two paths on the state graph, with one representing binding to non-activating ligands, while the other represents binding to activating ligands, with the states further down on each path seeing the TCR progressively being phosphorylated. This model is described in figure 5.1 and is adapted and modified from [1].

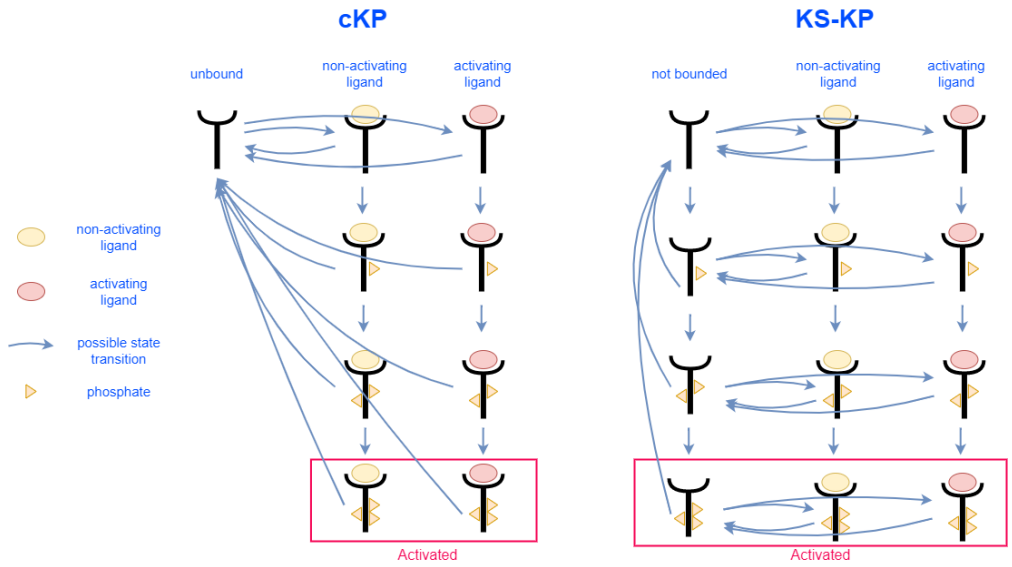


Figure 5.1: State transition graph of TCR cKP and KS-KP. It have phosphorylation threshold $n = 3$, each vertex is a possible state under a model and edges are possible transitions. Yellow is non-activating ligand, Red is activating ligand, blue arrows represents possible state transition, while yellow triangle is phosphate. Under each model, states within the red box form a macrostate group where the receptor is activated, all other states outside the red box are together the non-activated macrostate.

Further, citing [1], we setup the simulation such that the transition rate from unbound state to both non-activating ligand bound or activating ligand bound are dependent on the product of another constant, M, the overall concentration of ligands.

Thus, following figure 5.1, for the cKP model with phosphorylation count $n = 3$, there are 9 fully observed states/microstates:

$$\mathcal{S} = \{U, B_{n,0}, B_{n,1}, B_{n,2}, B_{n,3}, B_{a,0}, B_{a,1}, B_{a,2}, B_{a,3}\}. \tag{5.1}$$

Likewise, KS-KP will have 12 microstates:

$$\mathcal{S} = \{U_0, U_1, U_2, U_3, B_{n,0}, B_{n,1}, B_{n,2}, B_{n,3}, B_{a,0}, B_{a,1}, B_{a,2}, B_{a,3}\}. \tag{5.2}$$

Then, the rate matrix for cKP with phosphorylation count $n = 3$ can be expressed as:

$$R_X = \begin{bmatrix} \hat{M} & k_n^- & k_n^- & k_n^- & k_n^- & k_a^- & k_a^- & k_a^- & k_a^- \\ \hat{M}_n & -k_n^- - k_p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_p & -k_n^- - k_p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_p & -k_n^- - k_p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_p & -k_n^- & 0 & 0 & 0 & 0 \\ \hat{M}_a & 0 & 0 & 0 & 0 & -k_a^- - k_p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_p & -k_a^- - k_p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_p & -k_a^- - k_p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_p & -k_a^- \end{bmatrix}. \tag{5.3}$$

While KS-KP with phosphorylation count $n = 3$ will have a rate matrix of:

$$R_X = \begin{bmatrix} \hat{M}-k_p & k_{dp} & k_{dp} & k_{dp} & k_n^- & 0 & 0 & 0 & k_a^- & 0 & 0 & 0 \\ k_p & \hat{M}-k_p-k_{dp} & 0 & 0 & 0 & k_n^- & 0 & 0 & 0 & k_a^- & 0 & 0 \\ 0 & k_p & \hat{M}-k_{dp} & 0 & 0 & 0 & k_n^- & 0 & 0 & 0 & k_a^- & 0 \\ 0 & 0 & k_p & \hat{M}-k_{dp} & 0 & 0 & 0 & k_n^- & 0 & 0 & 0 & k_a^- \\ \hat{M}_n & 0 & 0 & 0 & -k_n^- - k_p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{M}_n & 0 & 0 & k_p & -k_n^- - k_p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{M}_n & 0 & 0 & k_p & -k_n^- - k_p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{M}_n & 0 & 0 & k_p & -k_n^- & 0 & 0 & 0 & 0 \\ \hat{M}_a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -k_a^- - k_p & 0 & 0 & 0 \\ 0 & \hat{M}_a & 0 & 0 & 0 & 0 & 0 & 0 & k_p & -k_a^- - k_p & 0 & 0 \\ 0 & 0 & \hat{M}_a & 0 & 0 & 0 & 0 & 0 & 0 & k_p & -k_a^- - k_p & 0 \\ 0 & 0 & 0 & \hat{M}_a & 0 & 0 & 0 & 0 & 0 & 0 & k_p & -k_a^- \end{bmatrix}. \tag{5.4}$$

In both matrices, $\hat{M}$ is a shorthand notation defined as follows, which take values depending on the input state X:

$$\hat{M}_n = \begin{cases} Mk_n^+ & X = L \\ 0 & X = H \end{cases} \tag{5.5}$$

$$\hat{M}_a = \begin{cases} 0 & X = L \\ Mk_a^+ & X = H \end{cases} \tag{5.6}$$

$$\hat{M} = \begin{cases} -Mk_n^+ & X = L \\ -Mk_a^+ & X = H \end{cases} \tag{5.7}$$

$$= -\hat{M}_n - \hat{M}_a. \tag{5.8}$$

5.2 Markov Distribution for Input and Upgrade to Continuous Time Simulation

Chapter 4 demonstrated that, for a hidden sensitive channel, using i.i.d. input over discrete time steps with step length $\Delta t \to 0$ will lead to a mutual information rate of 0. Since both TCR cKP and TCR KS-KP are hidden sensitive channels, we must change our paradigm and examine cases where the input event X is distributed following a Markov chain.

5.2.1 Defining a Markov Matrix by Specifications

Since the input event consist only of 2 possibilities, $\mathcal{X} = \{L, H\}$, we can use a 2 state Markov chain, a system where many properties are explicitly solvable and derivable.

Firstly, note that an i.i.d. Bernoulli distribution for X is technically a degenerate special case of X being a Markov chain. If we let $Pr(X = L) = p_L$, the Markov matrix is this following singular matrix:

$$M_X = \begin{bmatrix} p_L & p_L \\ 1 - p_L & 1 - p_L \end{bmatrix}. \tag{5.9}$$

As such, no matter what the current state x_i is, the probability of the next state x_{i+1} being in L is p_L, and is thus independent of x_i. Intuitively, it will have a steady state distribution (eigenvector with eigenvalue 1) as $[p_L, 1 - p_L]^T$.

Then, if we wish to have a Markov system that promotes staying at the same state, the matrix must have larger values on the diagonal elements, as they are the probabilities of x_{i+1} being the same state as x_i. Let a_L be the probability of staying at 'low' state each discrete step, a_H be the probability of staying at 'high' state each discrete step, then:

$$M_X = \begin{bmatrix} a_L & 1 - a_H \\ 1 - a_L & a_H \end{bmatrix}. \tag{5.10}$$

This matrix will have a eigenvector, or steady state distribution as:

$$\mu_{M_X} = \begin{bmatrix} \frac{1-a_H}{2-a_L-a_H} \\ \frac{1-a_L}{2-a_L-a_H} \end{bmatrix}. \tag{5.11}$$

To better study the case of input as persistent signals staying on the same state, also called *sojourns*, we want an expression for the average length of 'stay' in a state. From the definition of the Markov chain matrix's elements, clearly:

$$Pr(x_{i+1} = H | x_i = H) = a_H \tag{5.12}$$

$$Pr(x_{i+2} = H, x_{i+1} = H | x_i = H) = a_H^2 \tag{5.13}$$

$$Pr(x_{i+k} = H, ..., x_{i+1} = H | x_i = H) = a_H^k \tag{5.14}$$

$$Pr(x_{i+k+1} = L, x_{i+k} = H, ..., x_{i+1} = H | x_i = L) = (1 - a_H) a_H^{k-1} (1 - a_L), \tag{5.15}$$

where the length of a sojourn k follows a geometric distribution:

$$p(k) = (1 - a_H) a_H^{k-1} \qquad k \in \{1, 2, 3, ...\} \tag{5.16}$$

$$\sum_{j=1}^{\infty} p(j) = \sum_{j=1}^{\infty} (1 - a_H) a_H^{j-1} \tag{5.17}$$

$$= \frac{1 - a_H}{1 - a_H} = 1. \tag{5.18}$$

Moreover, the geometric distribution defined as

$$\Pr(X = k) = (1 - p)^{k-1} p \tag{5.19}$$

has a mean of $1/p$. Meaning that, if we wish to design a Markov input distribution system with mean sojourn length for 'high' input T_H, we can solve the following relation to get a_H, which is an element in the Markov matrix:

$$T_H = \frac{1}{p} \tag{5.20}$$

$$= \frac{1}{1 - a_H} \tag{5.21}$$

$$a_H = 1 - \frac{1}{T_H}. \tag{5.22}$$

From expression (5.10), the Markov matrix M_X is defined by 2 parameters, probability of staying at 'low' after a tick a_L, and probability of staying at 'high' after a tick a_H. However, for our purposes, we need a procedure to create a Markov matrix defined by a different set of 2 parameters, namely the average length of a sojourn at 'high' state T_H, and the probability of the input being in 'high' state p_H. One can also consider p_H as the proportion of time overall when

input state is in 'high'. Thus, p_H is just the element corresponding to $x = H$ in the steady state distribution μ_{M_X}, which gives us:

$$p_H = \frac{1 - a_L}{2 - a_L - a_H} \tag{5.23}$$

$$a_L = \frac{(2 - a_H)p_H - 1}{p_H - 1}. \tag{5.24}$$

Therefore, any 2 state Markov chain matrix can be created from spectral parameters that are T_H and p_H, following the rules described by expressions (5.22) and (5.24).

5.2.2 Simulation Results with Markov Input Distribution

With the system of input with Markov distribution setup, we can start examining its efficacy and attempt to find out the mean sojourn length on high T_H and probability of low input p_L that will give the maximum mutual information rate in different hidden sensitive receptors.

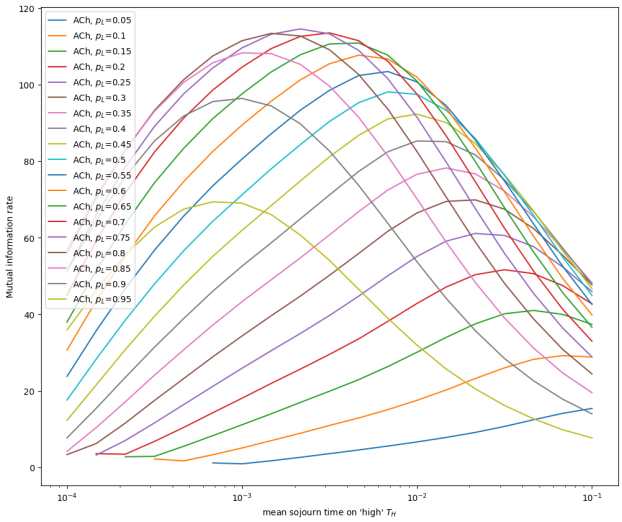
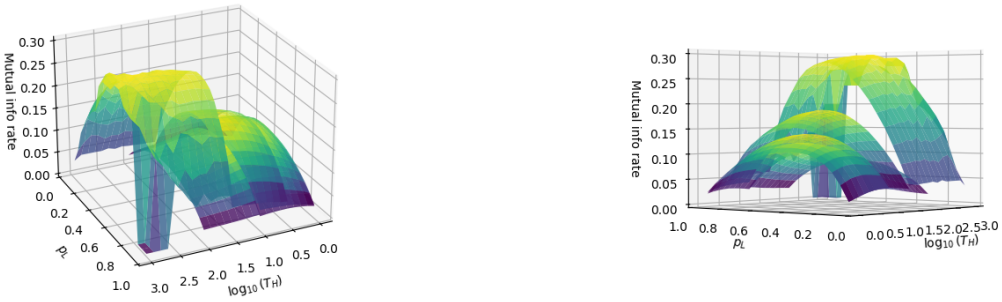


Figure 5.2: Rate of mutual information of ACh under Markov distributed input, discrete time step system, simulated with $\Delta t = 5 \cdot 10^{-5}$.

In figure (5.2) we can see a well formed 2 parameter curve that shows approximately what T_H and p_L values that maximizes the mutual information rate of the ACh receptor.



(a) Mutual information rate of TCR-KSKP

(b) Mutual information rate of TCR-KSKP

Figure 5.3: Each surface are of slightly different setup, such as different Δt or total simulation time. $\Delta t \in \{0.0002, 0.0005, 0.001, 0.005\}$

Simulating TCR With Markov Input

Experiments inquiring the mutual information landscape of TCR with Markov input distribution can now be carried out. We start with TCR KS-KP due to its more numerically moderate distributions. Moreover, its larger state space which requires larger matrices to implement provide us an upper bound estimation for what simulation size or length will be realistic given our computational resources. However, it is soon found out that despite demonstrating a maxima in the landscape given some Δt value, the maxima shifts upward and towards larger T_H parameter value as Δt is adjusted to smaller value, with no clear signs of convergence.

To demonstrate the fundamental property of convergence as $\Delta t \to 0$, we need to run simulations with progressively smaller Δt. Note that the Δt values eventually becomes smaller than ones used in figure 5.3, and the fact that larger T_H yields larger mutual information rate is reasonable as we are likely on the side of the landscape smaller than the optimal T_H. This eventually becomes an issue as larger T_H and smaller Δt eventually cause significant numerical artifacts due to the insufficient typical set syndrome. This happens when the Monte Carlo ensemble cannot adequately explore a sufficient subset of the space of all possible paths and the simulation result statistics not able to be averaged out.

Thus, a new simulation method will be needed. In the following subsection, we will devise a simulation system that works in continuous time by transforming all summations over discrete

ticks into a single matrix exponential for each sojourn. Thus, we will have a system that runs every simulation with Δt effectively being 0 by default, which will free us from needing to run multiple simulations with smaller Δt for convergence. Moreover, this also relieves us of second guessing if the lack of convergence is due to insufficient typical set or other numerical errors.

5.2.3 Creation of Continuous Time Monte Carlo Simulations

To keep computational time reasonable, we explore continuous time based simulation methods. Here each step of a sequence for inputs is a sojourn over a state, if we can run simulations based on constant number of input state switches, and the run time complexity are dependent only or mainly on number of sojourns, then we can estimate with proper accuracy the mutual information rate as $\Delta t \to 0$, no matter what T_H is, under the same amount of compute time.

Similar to the geometric distribution above, the length of a sojourn follows the exponential distribution:

$$f(t, \lambda) = \lambda e^{-\lambda t} \tag{5.25}$$

$$\frac{1}{\lambda} = T_{\mathrm{X}}. \tag{5.26}$$

Where $\mathrm{X} \in \mathcal{X}$, which in this context is either of the two state input set with 'low' and 'high', $\mathcal{X} = \{L, H\}$. If the input system consist only of 2 states, then for each current state sojourn, there is only one other state to switch to. Thus, generating a specific case of input sequence will give a list of sojourn times and the list of corresponding input states will simply be an alternating sequence of L, H.

To attain a specific case or sequence of output events, we utilize the Gillespie method. Here the code should output a sequence of elements in the fully observed state set $\mathcal{S}$, along with the time length of each output sojourns. Then, another function can map it to a sequence of elements in the observable state set $\mathcal{Y}$. Starting from an initial state, which can be any specific state $s_1 \in \mathcal{S}$, drawn from a choice distribution weighted by initial state vector. For any step i, given s_i is the mth element in $\mathcal{S}$, and that the current input sojourn is x_j, we use the mth column of R_{x_j}, excluding the element on the diagonal for the transition rates to every other state. Note that a diagonal element in the rate matrix stands for the rate of current state to itself, or rather the rate of leaving

the current state, equal to the negative sum of all other elements in the column. Then, the length of the current sojourn also follow the exponential distribution

$$f(t, \lambda) = \lambda e^{-\lambda t} \tag{5.27}$$

$$\frac{1}{\lambda} = \sum_{i=1, i \neq m}^{|\mathcal{S}|} (R_{x_j})_{i,m} = -(R_{x_j})_{m,m}. \tag{5.28}$$

This is the time until the current state changes to a different state. To decide which of the other $|\mathcal{S}| - 1$ state it will change to, simply draw from a choice distribution weighted by the set of other elements, meaning by the vector $[(R_{x_j})_{1,m}, (R_{x_j})_{2,m}, ..., (R_{x_j})_{m-1,m}, 0, (R_{x_j})_{m+1,m}, ...(R_{x_j})_{|\mathcal{S}|,m}]^T$.

If at any step, the time until next event happen added to current cumulative time goes over the end of the current input sojourn, meaning it goes into the next input sojourn, this means that no event has happened until the end of the current input sojourn. Simply start the Gillespie process again at the beginning of the new input sojourn with the rate matrix corresponding to the new input state.

Relating back to the discrete time step system with Markov chain distributed input, M_X is the stochastic Markov chain matrix for deciding the sequence of channel inputs. In the continuous time system, it has a corresponding rate matrix W_X, which can be expressed as follows:

$$W_X = \begin{bmatrix} -\frac{1}{T_L} & \frac{1}{T_H} \\ \frac{1}{T_L} & -\frac{1}{T_H} \end{bmatrix}. \tag{5.29}$$

Like other rate matrices, its column sums up to 0. T_L and T_H are the average sojourn time at 'low' and 'high' input, respectively. Intuitively, the two elements not on the main diagonal represents 'rate of input state transitioning from low/high to high/low per time unit', which equals to one over sojourn time. Then, following definitions, W_X and M_X are related by

$$M_X = \exp(\Delta t W_X), \tag{5.30}$$

Which is the rule to convert a rate matrix to discrete time step system with step length Δt. Both matrices has an eigenvector, with corresponding eigenvalue 0 for W_X and 1 for M_X, being the steady state distribution, $[\frac{T_L}{T_L+T_H}, \frac{T_H}{T_L+T_H}]^T$.

As $\frac{\Delta t}{T_H} \to 0$, or in other words $\Delta t \to 0$ when T_H stays the same, we have $M_X = I + \Delta t W_X$ with $T_H = \Delta t k_H$, where k_H is the length of sojourns on H input state measured in number of Δt length steps. With this M_X reduces to the same as the one defined by expressions (5.10), (5.22), and (5.24) based on the derivation with geometric distribution.

5.2.4 Continuous Time Belief Propagation Algorithm for Mutual Information

From the simulation we get two sequences of amorphous sojourn time lengths, an input and an output, each with corresponding states of each sojourns. Now, notice how the belief propagation for mutual information calculation in the discrete time step system works. Let the term 'joint sojourn' denote a time period within which neither the input state nor output state changes, where an output state change can be hidden if we look at the observable state set $\mathcal{Y}$. Within every joint sojourn, x and y stays the same, and the BP algorithm repeatedly apply the same per-step transition matrix A_x and the same mask H_y in every discrete time step. As mentioned above, the steps are:

For $P(Y|X)$:

$$\beta_k = H_Y A_x \alpha_{k-1} \tag{5.31}$$

$$\nu_k = \|\beta_k\|_1 \tag{5.32}$$

$$\alpha_k = \frac{\beta_k}{\nu_k} \tag{5.33}$$

$$\log(P(Y|X)) = \sum_{k=1}^{n} \log(\nu_k), \tag{5.34}$$

Here, we know that each α will only have non zero elements in the elements corresponding to the current y, and at every step $A_x \alpha_{k-1}$ is filtered out by the mask H_y. Let T be the time taken in this joint sojourn thus $T = n\Delta t$ and we wish to take the limit $\Delta t \to 0$.

Therefore we will basically have:

$$\beta_1 = H_Y A_x \alpha_0 \tag{5.35}$$

$$\nu_1 = \|\beta_1\|_1 = \|H_Y A_x \alpha_0\|_1 \tag{5.36}$$

$$\alpha_1 = \frac{\beta_1}{\nu_1} \tag{5.37}$$

$$\beta_2 = H_Y A_x \alpha_1 \tag{5.38}$$

$$\nu_2 = \|\beta_2\|_1 = \frac{\left\|(H_Y A_x)^2 \alpha_0\right\|_1}{\nu_1} \tag{5.39}$$

$$\vdots \tag{5.40}$$

$$\nu_3 = \|\beta_3\|_1 = \frac{\left\|(H_Y A_x)^3 \alpha_0\right\|_1}{\nu_1 \nu_2} \tag{5.41}$$

$$\nu_1 \nu_2 \nu_3 = \left\|(H_Y A_x)^3 \alpha_0\right\|_1 \tag{5.42}$$

$$\prod_{k=1}^{n} \nu_k = \|(H_Y A_x)^n \alpha_0\|_1 \tag{5.43}$$

$$\sum_{k=1}^{n} \log(\nu_k) = \log(\|(H_Y A_x)^n \alpha_0\|_1) \tag{5.44}$$

$$\text{as } \Delta t \to 0 \tag{5.45}$$

$$\sum_{k=1}^{n} \log(\nu_k) = \log(\|H_Y (I + \Delta t R_x)^n \alpha_0\|_1) \tag{5.46}$$

$$= \log(\|(I + \Delta t H_Y R_x)^n \alpha_0\|_1) \tag{5.47}$$

$$\text{let } \delta t = n \Delta t \tag{5.48}$$

$$= \log(\|\exp(H_Y R_x \cdot \delta t) \alpha_0\|_1). \tag{5.49}$$

For $P(Y)$, we need only regard the sojourns in Y.

$$\hat{\beta}_k = \hat{H}_Y M_U \hat{\alpha}_{k-1} \tag{5.50}$$

$$\hat{\nu}_k = \left\|\hat{\beta}_k\right\|_1 \tag{5.51}$$

$$\hat{\alpha}_k = \frac{\hat{\beta}_k}{\hat{\nu}_1} \tag{5.52}$$

$$\log(P(Y)) = \sum_{k=1}^{n} \log(\hat{\nu}_k). \tag{5.53}$$

We can follow the same thought process as above. The BP algorithm for $P(Y)$ considers each possible joint input-output states, which are elements of the cross product set $\mathcal{X} \times \mathcal{S}$. Thus, $\hat{H}_y$ is

the expanded mask matrix mapping states in $\mathcal{X} \times \mathcal{S}$ to $\mathcal{Y}$, which is just the $\mathcal{S}$ to $\mathcal{Y}$ mask matrix H_Y repeated twice on the diagonal for each X $\in \mathcal{Y}$ since the input state $\mathcal{X}$ does not influence the mapping of fully observed output state to observable output state. Likewise, M_U is a matrix of chunks stemming from an outer product between the Markov chain transition probability matrices A_L and A_H, and the input state Markov chain matrix M_X, thus forming the stochastic matrix that is the transition probability matrix of the Markov chain over the join state set $\mathcal{X} \times \mathcal{S}$. Note that:

$$M_U = \left[\begin{array}{c|c} A_L & A_H \\ \hline A_L & A_H \end{array}\right] \odot M_X \tag{5.54}$$

$$\text{as } \Delta t \to 0 \tag{5.55}$$

$$= \left[\begin{array}{c|c} I + \Delta t R_L & I + \Delta t R_H \\ \hline I + \Delta t R_L & I + \Delta t R_H \end{array}\right] \odot (I + \Delta t W_X) \tag{5.56}$$

$$= \left[\begin{array}{c|c} I + \Delta t R_L & I + \Delta t R_H \\ \hline I + \Delta t R_L & I + \Delta t R_H \end{array}\right] \odot \begin{bmatrix} 1 - \frac{\Delta t}{T_L} & \frac{\Delta t}{T_H} \\ \frac{\Delta t}{T_L} & 1 - \frac{\Delta t}{T_H} \end{bmatrix} \tag{5.57}$$

$$= \left[\begin{array}{c|c} (I + \Delta t R_L)(1 - \frac{\Delta t}{T_L}) & (I + \Delta t R_H)\frac{\Delta t}{T_H} \\ \hline (I + \Delta t R_L)\frac{\Delta t}{T_L} & (I + \Delta t R_H)(1 - \frac{\Delta t}{T_H}) \end{array}\right] \tag{5.58}$$

$$= \left[\begin{array}{c|c} I + \Delta t(R_L - \frac{1}{T_L} \cdot I) + O(\Delta t^2) & \frac{\Delta t}{T_H} I + O(\Delta t^2) \\ \hline \frac{\Delta t}{T_L} I + O(\Delta t^2) & I + \Delta t(R_H - \frac{1}{T_H} \cdot I) + O(\Delta t^2) \end{array}\right] \tag{5.59}$$

$$= \left[\begin{array}{c|c} I + \Delta t(R_L - \frac{1}{T_L} \cdot I) & \frac{\Delta t}{T_H} I \\ \hline \frac{\Delta t}{T_L} I & I + \Delta t(R_H - \frac{1}{T_H} \cdot I) \end{array}\right] \tag{5.60}$$

$$= I + \Delta t \left[\begin{array}{c|c} R_L - \frac{1}{T_L} \cdot I & \frac{1}{T_H} I \\ \hline \frac{1}{T_L} I & R_H - \frac{1}{T_H} \cdot I \end{array}\right] \tag{5.61}$$

$$W_U = \left[\begin{array}{c|c} R_L - \frac{1}{T_L} \cdot I & \frac{1}{T_H} I \\ \hline \frac{1}{T_L} I & R_H - \frac{1}{T_H} \cdot I \end{array}\right] \tag{5.62}$$

$$M_U = I + \Delta t W_U. \tag{5.63}$$

This reduce the joint state Markov chain matrix M_U back to the form of a rate matrix W_U which can be used in the continuous time belief propagation system. We can see that each of W_U's

column sums up to 0, since $R_X - \frac{1}{T_X} \cdot I + \frac{1}{T_X} \cdot I = R_X$ and R_X have columns that sums up to 0.

From here, we can just compute $P(Y_{t\in[t_j,t_j+\delta t]}|X_{t\in[0,t_j+\delta t]}, Y_{t\in[0,t_j]})$, the partial probability contributed to $P(Y|X)$ by a joint sojourn that spans over the time period $[t_j, t_j + \delta t]$, within which the states x and y are the same, with an initial message vector α_{t_j}, as follows:

$$\log(P(Y_{t\in[t_j,t_j+\delta t]}|X_{t\in[0,t_j+\delta t]}, Y_{t\in[0,t_j]})) = \sum_{k=1}^{n} \log(\nu_k) \tag{5.64}$$

$$= \log(\left\|\exp(H_Y R_x \cdot \delta t)\alpha_{t_j}\right\|_1) \tag{5.65}$$

Likewise, $P(Y_{t\in[t_j,t_j+\delta t]}|Y_{t\in[0,t_j]})$, the partial probability contributed to $P(Y)$ by a sojourn, within which y is the same, with an initial message vector $\hat{\alpha}_{t_j}$, is calculated as follows:

$$\log(P(Y_{t\in[t_j,t_j+\delta t]}|Y_{t\in[0,t_j]})) = \sum_{k=1}^{n} \log(\hat{\nu}_k) \tag{5.66}$$

$$= \log(\left\|\exp(\hat{H}_Y W_U \cdot \delta t)\hat{\alpha}_{t_j}\right\|_1) \tag{5.67}$$

When a joint sojourn in the $P(Y|X)$ algorithm ends due to x switching to a different state. Simply start applying the rate matrix corresponding to the new x, along with the mask for y, $H_y R_x$, and repeat the calculation in (5.47) and keep adding to the result.

When a joint sojourn in the $P(Y|X)$ algorithm ends due to y switching to a different state, or when a sojourn in the $P(Y)$ algorithm ends, we can note that, in the case of the discrete time algorithm (for $P(Y|X)$), at the tick right after a y value change, and we see that $\alpha_{k-1} = H_{y_{k-1}}\alpha_{k-1}$, we have:

$$\beta_k = H_{y_k} A_x \alpha_{k-1} \tag{5.68}$$

$$\nu_k = \|\beta_k\|_1 \tag{5.69}$$

$$= \|H_{y_k} A_x \alpha_{k-1}\|_1 \tag{5.70}$$

$$= \left\|H_{y_k}(I + \Delta t R_x) H_{y_{k-1}}\alpha_{k-1}\right\|_1 \tag{5.71}$$

$$= \left\|\Delta t H_{y_k} R_x H_{y_{k-1}}\alpha_{k-1}\right\|_1 \tag{5.72}$$

$$= \Delta t \left\|H_{y_k} R_x H_{y_{k-1}}\alpha_{k-1}\right\|_1 \tag{5.73}$$

Since the previous mask $H_{y_{k-1}}$ leaves α_{k-1} with non zero values only in elements corresponding to y_{k-1}, while the new H_{y_k} mask only takes vector elements corresponding to y_k, only the matrix

elements in the off diagonal block $H_{y_k} R_x H_{y_{k-1}}$ will contribute to the new message vector. Likewise for $P(Y)$, we have a similar expression, though the $\hat{\nu}_k$ norm combines $|\mathcal{X}|$ counts of off diagonal chunks of multiple R_x, for all the X $\in \mathcal{X}$. Once again note that usually $|\mathcal{X}| = 2$.

In the BP algorithm for the discrete time simulation system, both the $P(Y|X)$ and $P(Y)$ cases of the expression (5.73) goes to 0 as $\Delta t \to 0$. Then, the logarithm of them will approach $-\infty$, which are then added to the accumulating log probability that will become the result. However, note that, in the same simulation instance, the number of y changes across the $P(Y|X)$ and $P(Y)$ algorithms are the same, and exist in the same locations as they have the same sequence of y. Thus, if the step size Δt are the same, which should always be the case, then these constant $\log(\Delta t)$ terms will all cancel each other out in the subtraction $\log(P(Y|X)) - \log(P(Y))$ at the end that will give us the mutual information of each simulation instance.

Thus, in the continuous time system, when a change in output macrostate y is encountered, the BP algorithm need only calculate

$$\nu_k = \left\| H_{Y_k} R_x H_{Y_{k-1}} \alpha_{k-1} \right\|_1 \tag{5.74}$$

$$\hat{\nu}_k = \left\| \hat{H}_{Y_k} W_U \hat{H}_{Y_{k-1}} \alpha_{k-1} \right\|_1 \tag{5.75}$$

and add them to the accumulating sum of the respective part of the algorithm, as the only thing that matters is the difference between them.

At the end, the difference between the accumulated sum of logs of the two streams will be the probability differential in continuous time of the simulation instance. And just like above sections, taking the mean of all instances in a Monte Carlo ensemble will give us an approximation of the system's mutual information.

5.3 Mutual Information Rate Shows KS-KP to be More Realistic

With the above progress, we can now explore the mutual information rate landscape of TCR cKP and KS-KP over different constants of system setups.

5.3.1 Mutual Information Landscape over Activating Ligand's Chemical Kinetic Constants

After exploration and consulting literature, it is deemed that p_L and T_H should be specific constants instead of parameters to vary. Thus, we look on to exploring the mutual information landscape regarding the chemical kinetic constants of the activating ligand.

As mentioned in section 5.1, since the T-cell is swimming in a sea of non-activating ligands aiming to identify the one foreign cell with activating ligands, we simplify this to a system of 2 species of ligand, each with their own path to go down in the state graph. The non-activating ligand will have binding and unbinding rates k_n^+ and k_n^-, which we keep as constants. Meanwhile, we explore the variation in the kinetics rate constants of the activating ligand, k_a^+ and k_a^-, which forms the axes of our mutual information landscape. During 'low' input, the T-cell is in contact with the 'background noise' of the non-activating ligands, which have a very low but non zero probability to triggering activation, and during 'high' input, the T-cell will come into close contact with a foreign cell, meaning it have chance to bind to its activating ligands, and go down the path in the state graph where it can be activated, which should be more likely than if it had bound to a non-activating ligand, due to the different unbinding rate of ligand types k_n^- and k_a^-. Since in practice the ratio of non-activating ligand to activating ligand is overwhelmingly large, we should take p_L as $\lim_{p\to 1^-} p$, a number less than but infinitely close to 1. However, in practice, $p_L = 0.99$ will suffice, as it will give enough time for the TCR to return to the steady state distribution of the case of constant 'low' input, manifested practically as only contacting with non-activating ligands after detaching from the single foreign cell for an overwhelmingly long period of time. And we put mean sojourn time on 'high' input T_H as 30 seconds, after communicating with our biophysics collaborators. Technically the act of close contact between T-cell and the other cell does not have a set mean length of time, but 30s is a reasonable estimation for our simplified model.

Moreover, the rate of mutual information will no longer be measured per time, but instead, per sojourns on 'high' input state. This allow us to measure the average impact per instance of the event that is a foreign cell with its activating ligands coming into close contact of the T-cell. Thus,

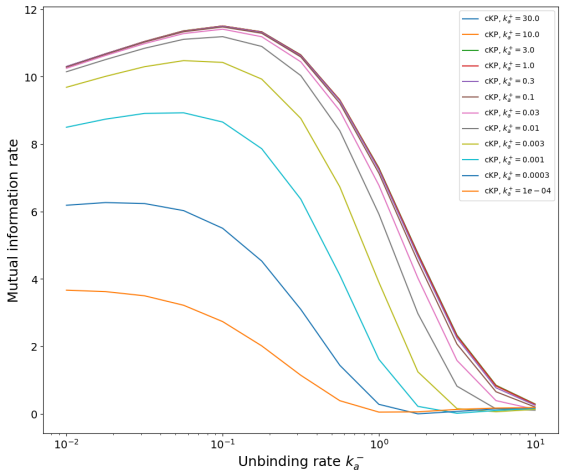


Figure 5.4: Mutual information per activating ligand of cKP as function of k_a^- on log scale. Different curves represent different k_a^+ values on log scale from 10^{-4} to 10^2.

the rate of mutual information is now defined as:

$$\mathcal{I}(X,Y) = \lim_{m\to\infty} \frac{2}{c}\mathbb{E}\left[\log\left(\frac{p(y_1^m|x_1^m)}{p(y_1^{m'})}\right)\right] \tag{5.76}$$

$$= \lim_{m\to\infty} \frac{T_H}{p_H \sum_{i=0}^{m} T_i}\mathbb{E}\left[\log\left(\frac{p(y_1^m|x_1^m)}{p(y_1^{m'})}\right)\right], \tag{5.77}$$

Here x_1^m and y_1^m are the sequence of joint input and output event sojourns, and m is the number of joint sojourns, and m' is the total number of sojourn that regards only the state of y. Meanwhile c is the number of X sojourns, when the input X stays the same, counted without regards to output Y observations. Thus $c/2$ represents the number of instances of continuous H input happening over all m joint sojourns as there are only two possible input states. Letting T_i be the length of the ith sojourn, T_H be the average length of a H input sojourn, and p_H be the average portion of time of input being in H. Then T_H/p_H will be the average length of a L and H cycle, and $\sum_{i=0}^{m} T_i$ will be the total time elapsed in the simulation instance of m joint sojourns. In the long run as $m \to \infty$, $c/2$ is equivalent to $\sum_{i=0}^{m} T_i/(T_H/p_H)$.

Thus, table 5.1 shows the constants we will use for the following simulations.

Description	symbol	value	unit
Binding rate of non-activating ligand	k_n^+	$4 \cdot 10^{-4}$	$\mu m^2 s^{-1}$
Unbinding rate of non-activating ligand	k_n^-	2.0	s^{-1}
Default concentration of ligands	M	330	μm^{-2}
Phosphorylation rate	k_p	1.0	s^{-1}
Dephosphorylation rate	k_{dp}	5.0	s^{-1}
Mean sojourn time of activating ligand in contact	T_H	30	s
Portion of time of activating ligand in contact	p_H	0.01	

Table 5.1: Specific constants used in our model, based on [1–4].

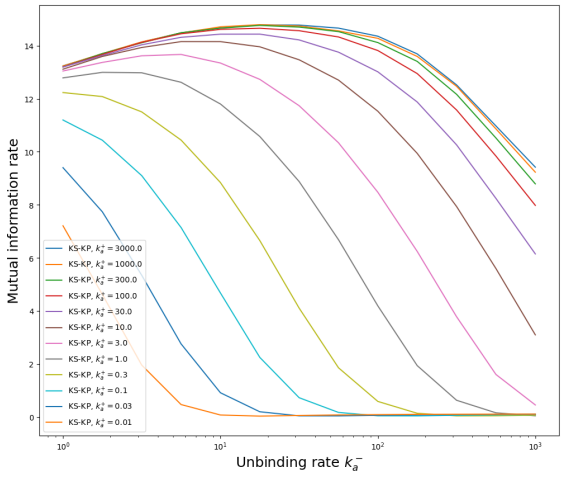


Figure 5.5: Mutual information per activating ligand of KS-KP as function of k_a^- on log scale. Different curves represent different k_a^+ values on log scale from 10^{-2} to 10^4.

5.3.2 The Mutual Information Maximizing Value of k^- and its Internal Maxima

Exploring this system leads us to figures 5.4 and 5.5. For both the cKP and KS-KP system, when k_a^+ is constant, an internal maxima over the k_a^- axis is observed. When k_a^- stay constant, the maximum mutual information rate is achieved when $k_a^+ \to \infty$, where it converges to a constant value. This is understandable as when k_a^+ becomes infinite, binding to ligands become instantaneous upon the T-cell contacting another cell, thus nullifying the unbound state. However, since the phosphorylation rate k_p is still finite, the progress toward activation will not be instantaneous, thus, the rate of mutual information per event of close contact converges to a finite value.

However, when k_a^+ is constant, an internal maxima of k_a^- is observed for each value of k_a^+, with it shifting towards larger k_a^- value and settling as the value of k_a^+ also gets closer to approach its convergent value of maximum rate of mutual information.

5.3.3 Experimentally described TCR activating ligands follows the trend of kinetic parameter pairs that result in optimal mutual information for the KS-KP model

Therefore, we have now discovered a way to distinguish between cKP and KS-KP. By optimizing over k_a^- for each k_a^+, we obtain the internal maxima of k_a^- for each models. If it can be demonstrated that one of the two models achieve maximum rate of mutual information for parameters closer to real life values, we can show that said model is likely to be more realistic, as evolutionary logic dictates that a receptor channel system will have evolved towards achieving maximum mutual information for its 'intended' working environment.

In this case, we compare the varied binding and unbinding constants k_a^- and k_a^+ in the k_a^- internal optima over k_a^+ to the experimentally derived kinetic constants of activating ligands in relevant literature.

As seen in figure 5.6, for small k_a^+ values, such as those close to the default non-activating ligands, the two models have similar maxima of k_a^-. However, two models diverge as k_a^+ becomes larger, with cKP's internal maxima tapering off, while KS-KP's internal maxima continues to shift until a later plateau. We also plot the experimentally found k_a^- and k_a^+ values for TCR

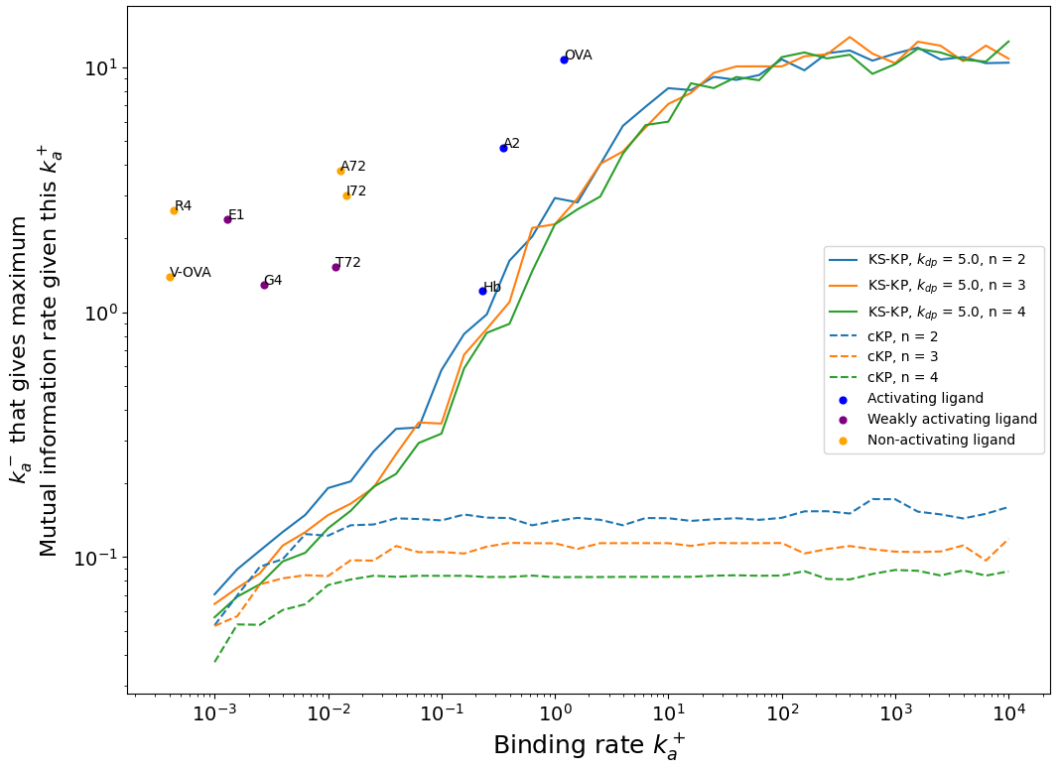


Figure 5.6: Internal maxima in the k_a^- axis for each k_a^+ in mutual information per activating ligand, over cKP and KS-KP with different phosphorylation threshold n. Both axes are on log scale.

ligands from [2,47], which includes activating ligands (blue) and non-activating ligands (yellow and purple). Here we find that the activating ligands follow the same trend as the k_a^- and k_a^+ values at the locations of internal mutual information maxima for the KS-KP model.

The main sources here [2, 47] collected their data experimentally by micro-manipulating T-cells in an adhesion frequency assay to contact with red blood cells or beads, and measured the contact times to infer binding and unbinding rates. This gives binding rate data with concentration measured in units per area which is the surface area of the secondary body, as opposed to per volume for when ligand molecules are freely floating around in a test tube. This makes it highly applicable to our results as this gives the most realistic numerical values that are closest to the dynamics of T-cells in real life.

Moreover, for the KS-KP model, the phosphorylation count threshold n is shown to not have a great impact on the region of maxima, while cKP shows more separation between the sub-models varying in n. This can be due to the existence of unbound phosphorylated states in KS-KP masking the phosphorlation step parameter, making them appear similar. This means that cKP has sensitivity to the value of n while KS-KP lacks it, and more work into other simulation and evaluation methods are likely needed to discover the precise value of n in the ground truth model.

As mentioned above, we assume that the TCR has evolved specifically for the purpose of recognizing these non-self, activating ligands, and thus maximizes its channel capacity regarding its 'intended' targets, as described in the literature [18]. Thus, the kinetics constants of the activating ligands should then coincide with those that give a maximal mutual information rate, which is its channel capacity. Therefore, our results ultimately suggest that KS-KP being the more favourable model to describe the mechanism of TCR activation.

Chapter 6

Conclusion

6.1 Contributions of This Thesis

In conclusion, we have advanced previous research by applying the hidden Markov model as an information channel for biological receptors model in [5] to the problem of information theory metric of T-cell receptor laid out in [1] and concluded that KS-KP is the more likely model to describe TCR mechanism.

In the process of doing so, we also developed a more computationally efficient system of Monte Carlo Markov chain simulation and probability calculation using belief propagation for estimating rate of mutual information based on [5, 45] which is massively parallelizable on Nvidia CUDA GPUs. This allow us to simulate larger receptor channel systems, which estimates the mutual information value to greater accuracy, and using less computational time, which led us to new discoveries. Likewise, this system can be applied to other research projects tackling similar problems of estimating the information theoretic channel capacity of a biological receptor.

Then, using our highly efficient simulation system, we discovered new a paradigm and completed the theorem of the capacity of biological signal transduction in [26] by changing previous assumptions and proving the hidden sensitive channel theorem, which dictates that Markov chain channels with only hidden sensitive transitions will have 0 mutual information, thus capacity, on an identically and independently distributed (i.i.d.) input when the time step length approaches 0. As both models of TCR used by this thesis are hidden sensitive channels, we figured a comparison must

be simulated in a regime with Markov input, in other words where the input state are persistence over periods of times.

After that, we further upgraded our system to simulate and estimate mutual information in continuous time, with the Markov chain ensemble and belief propagation algorithms effectively running in time step length of 0. This provide even more efficient inference and evades the issue of insufficient typical set exploration and unsure convergence by time step length.

Finally, using the tools and knowledge we built and discovered, we are able to complete the goal of distinguishing between the cKP and KS-KP models of TCR using the information channel capcaities of the model's mathematical structure and deduced that KS-KP is the more likely model of the mechanism of TCR activation according to evolutionary logic.

6.2 Future Direction

For the future direction of this thesis' line of work, other researchers can utilize the Monte Carlo simulation and belief propagation systems on GPU we devised for many other biological receptor systems, which is designed to be modular, flexible, and adaptable.

Likewise, the result of this thesis leaves a few incomplete questions for others to explore.

Firstly, our argument that KS-KP being that more likely model of TCR mechanism is not an absolute conclusion stemming from mathematical proof, and there exist other candidate models that can be explored in future research [10, 14]. More exploration into the information theoretic metrics of the TCR models can support or refute our arguments.

Secondly, we showed that the phosphorylation threshold n cannot easily distinguish a model under the act of comparing rate of mutual information, especially regarding KS-KP models. Thus, more exploration into this topic, coupled with experimental progress, can allow us to uncover more about this parameter.

Thirdly, even though we have proven the hidden sensitive channel theorem, in system such as Calmodulin receptor (CaM) from [5] where there are simultaneously hidden and non-hidden sensitive transitions, one can look more into the mathematics of quantifying the impact of the two types of transition to the rate of mutual information over different i.i.d. time step or persistent sojourn lengths in the input state.

Finally, inquiry into integrating other parts of the T-cell activation pathway into our receptor model can be explored in the future, which gives a more complete picture of the mechanism of TCR activation.

Bibliography

[1] A. S. Moffett, K. A. Ganzinger, and A. W. Eckford, "Comparing kinetic proofreading and kinetic segregation for t cell receptor activation," *Physical Review Research*, vol. 7, no. 2, p. 023003, 2025.

[2] J. Huang, V. I. Zarnitsyna, B. Liu, L. J. Edwards, N. Jiang, B. D. Evavold, and C. Zhu, "The kinetics of two-dimensional tcr and pmhc interactions determine t-cell responsiveness," *Nature*, vol. 464, no. 7290, pp. 932–936, 2010.

[3] J. Pettmann, A. Huhn, E. Abu Shah, M. A. Kutuzov, D. B. Wilson, M. L. Dustin, S. J. Davis, P. A. van der Merwe, and O. Dushek, "The discriminatory power of the t cell receptor," *Elife*, vol. 10, p. e67092, 2021.

[4] R. A. Fernandes, K. A. Ganzinger, J. C. Tzou, P. Jönsson, S. F. Lee, M. Palayret, A. M. Santos, A. R. Carr, A. Ponjavic, V. T. Chang, *et al.*, "A cell topography-based mechanism for ligand discrimination by the t cell receptor," *Proceedings of the National Academy of Sciences*, vol. 116, no. 28, pp. 14002–14010, 2019.

[5] A. W. Eckford and P. J. Thomas, "The channel capacity of channelrhodopsin and other intensity-driven signal transduction receptors," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 4, no. 1, pp. 27–38, 2019.

[6] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley, 2012.

[7] R. S. Ganti, W.-L. Lo, D. B. McAffee, J. T. Groves, A. Weiss, and A. K. Chakraborty, "How the t cell signaling network processes information to discriminate between self and agonist ligands," *Proceedings of the National Academy of Sciences*, vol. 117, no. 42, pp. 26020–26030, 2020.

[8] C. C. m. professional, "T cells: Types and function," Mar 2025.

[9] U. Alon, *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2019.

[10] R. A. Mariuzza, P. Agnihotri, and J. Orban, "The structural basis of t-cell receptor (tcr) activation: An enduring enigma," *Journal of Biological Chemistry*, vol. 295, no. 4, pp. 914–925, 2020.

[11] G. Tkačik, C. G. Callan Jr, and W. Bialek, "Information flow and optimization in transcriptional regulation," *Proceedings of the National Academy of Sciences*, vol. 105, no. 34, pp. 12265–12270, 2008.

[12] G. Tkačik and W. Bialek, "Information processing in living systems," *Annual Review of Condensed Matter Physics*, vol. 7, no. 1, pp. 89–117, 2016.

[13] C. T. Bergstrom and M. Rosvall, "The transmission sense of information," *Biology & Philosophy*, vol. 26, no. 2, pp. 159–176, 2011.

[14] P. A. Van Der Merwe and O. Dushek, "Mechanisms for t cell receptor triggering," *Nature Reviews Immunology*, vol. 11, no. 1, pp. 47–55, 2011.

[15] C. A. Janeway Jr, P. Travers, M. Walport, and M. J. Shlomchik, *Immunobiology: The Immune System in Health and Disease*. Garland science, 2001.

[16] A. K. Sewell, "Why must t cells be cross-reactive?," *Nature Reviews Immunology*, vol. 12, no. 9, pp. 669–677, 2012.

[17] M. M. Davis, J. J. Boniface, Z. Reich, D. Lyons, J. Hampl, B. Arden, and Y.-h. Chien, "Ligand recognition by $\alpha\beta$ t cell receptors," *Annual review of immunology*, vol. 16, no. 1, pp. 523–544, 1998.

[18] A. W. Eckford, "The philosophy and science of biological information theory: A brief overview and recent results," in *International Zurich Seminar on Information and Communication (IZS 2026)*, pp. 131–136, ETH Zurich, 2026.

[19] O. Rivoire and S. Leibler, "The value of information for populations in varying environments," *Journal of Statistical Physics*, vol. 142, pp. 1124–1166, 2011.

[20] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, "Information transduction capacity of noisy biochemical signaling networks," *science*, vol. 334, no. 6054, pp. 354–358, 2011.

[21] A. Rhee, R. Cheong, and A. Levchenko, "The application of information theory to biochemical signaling systems," *Physical biology*, vol. 9, no. 4, p. 045011, 2012.

[22] S. Uda and S. Kuroda, "Analysis of cellular signal transduction from an information theoretic approach," in *Seminars in cell & developmental biology*, vol. 51, pp. 24–31, Elsevier, 2016.

[23] S. Uda, T. H. Saito, T. Kudo, T. Kokaji, T. Tsuchiya, H. Kubota, Y. Komori, Y.-i. Ozaki, and S. Kuroda, "Robustness and compensation of information transmission of signaling pathways," *Science*, vol. 341, no. 6145, pp. 558–561, 2013.

[24] T. Tsuruyama, "Analysis of cell signal transduction based on kullback–leibler divergence: Channel capacity and conservation of its production rate during cascade," *Entropy*, vol. 20, no. 6, p. 438, 2018.

[25] S. Ito and T. Sagawa, "Maxwell's demon in biochemical signal transduction with feedback loop," *Nature communications*, vol. 6, no. 1, p. 7498, 2015.

[26] P. J. Thomas and A. W. Eckford, "Capacity of a simple intercellular signal transduction channel," *IEEE Transactions on information Theory*, vol. 62, no. 12, pp. 7358–7382, 2016.

[27] J. J. Stewart, C. Y. Lee, S. Ibrahim, P. Watts, M. Shlomchik, M. Weigert, and S. Litwin, "A shannon entropy analysis of immunoglobulin and t cell receptor," *Molecular immunology*, vol. 34, no. 15, pp. 1067–1082, 1997.

[28] J. Henderson, Y. Nagano, M. Milighetti, and A. Tiffeau-Mayer, "Limits on inferring t cell specificity from partial information," *Proceedings of the National Academy of Sciences*, vol. 121, no. 42, p. e2408696121, 2024.

[29] L. Bene, M. Bagdány, and L. Damjanovich, "T-cell receptor is a threshold detector: sub-and supra-threshold stochastic resonance in tcr-mhc clusters on the cell surface," *Entropy*, vol. 24, no. 3, p. 389, 2022.

[30] P. François and G. Altan-Bonnet, "The case for absolute ligand discrimination: modeling information processing and decision by immune t cells," *Journal of Statistical Physics*, vol. 162, pp. 1130–1152, 2016.

[31] V. Tserunyan and S. Finley, "Information-theoretic analysis of a model of car-4-1bb-mediated nfκb activation," *Bulletin of Mathematical Biology*, vol. 86, no. 1, p. 5, 2024.

[32] S. F. Lee, D. Klenerman, K. Ganzinger, V. Chang, J. McColl, and M. Palayret, "Super-resolution imaging of t cell triggering supports the kinetic segregation model in the adaptive immune response," *Biophysical Journal*, vol. 104, no. 2, p. 428a, 2013.

[33] K. Choudhuri, D. Wiseman, M. H. Brown, K. Gould, and P. A. van der Merwe, "T-cell receptor triggering is critically dependent on the dimensions of its peptide-mhc ligand," *Nature*, vol. 436, no. 7050, pp. 578–582, 2005.

[34] V. N. Gudivada, D. Rao, and V. V. Raghavan, "Big data driven natural language processing research and applications," in *Handbook of statistics*, vol. 33, pp. 203–238, Elsevier, 2015.

[35] I. M. Kramer, B. D. Gomperts, and P. E. Tatham, *Signal transduction*. Elsevier/Academic Press London, 2016.

[36] A. C. Guyton and J. E. Hall, *Guyton and Hall textbook of medical physiology*. Elsevier, 2011.

[37] B. D. Gomperts, P. E. Tatham, *et al.*, *Signal transduction*. Academic Press, 2009.

[38] J. Ngoenkam, W. W. Schamel, and S. Pongcharoen, "Selected signalling proteins recruited to the t-cell receptor–cd3 complex," *Immunology*, vol. 153, no. 1, pp. 42–50, 2018.

[39] F. Horn and R. Jackson, "General mass action kinetics," *Archive for rational mechanics and analysis*, vol. 47, pp. 81–116, 1972.

[40] N. J. Tro, T. D. Fridgen, L. Shaw, and R. S. Boikess, *Chemistry: A molecular approach*, vol. 5. Pearson Boston, MA, 2017.

[41] T. Székely Jr and K. Burrage, "Stochastic simulation in systems biology," *Computational and structural biotechnology journal*, vol. 12, no. 20-21, pp. 14–25, 2014.

[42] D. T. Gillespie, "Stochastic simulation of chemical kinetics," *Annu. Rev. Phys. Chem.*, vol. 58, no. 1, pp. 35–55, 2007.

[43] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *The Journal of chemical physics*, vol. 115, no. 4, pp. 1716–1733, 2001.

[44] altugkarakurt, "Probability theory - what is mutual information of a markov chain? - mathematics stack exchange."

[45] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on information theory*, vol. 47, no. 2, pp. 498–519, 2001.

[46] D. T. Chen, A. W. Eckford, and P. J. Thomas, "Capacity of a class of fully observed multistate poisson-type signal transduction channels," in *2025 IEEE International Symposium on Information Theory (ISIT)*, pp. 1–6, IEEE, 2025.

[47] J. Hong, S. P. Persaud, S. Horvath, P. M. Allen, B. D. Evavold, and C. Zhu, "Force-regulated in situ tcr–peptide-bound mhc class ii kinetics determine functions of cd4+ t cells," *The Journal of Immunology*, vol. 195, no. 8, pp. 3557–3564, 2015.

[48] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965.

Appendix A

Validation of Program Code

A.1 Distribution of Log Probability Differences of models

We start by building simulations of previously researched models to demonstrate the validaty of the programmed GPU simulation system. To do so, we set the model parameters to an arbitrary set of constants and examine the distribution of log probability differences of each simulation instances of a Monte Carlo ensemble. These are the values whose mean will be calculated and outputted as the estimation of mutual information rate of this model setup.

A.1.1 Acetylcholine receptor (ACh)

Firstly, we examine the mutual information rate of the Acetylcholine receptor (ACh), whose state graph is described in figure A.1.

From runs with ACh, we found that the output mutual information rate from a set of independent runs of the simulations follows almost perfectly the Gaussian normal distribution. Using the the Shapiro-Wilk test for normality [48], the output array of probability log differences achieved statistic being 0.993 and 0.9997 in the immediate examples, this suggests that the output array's values follows closely to a normal distribution.

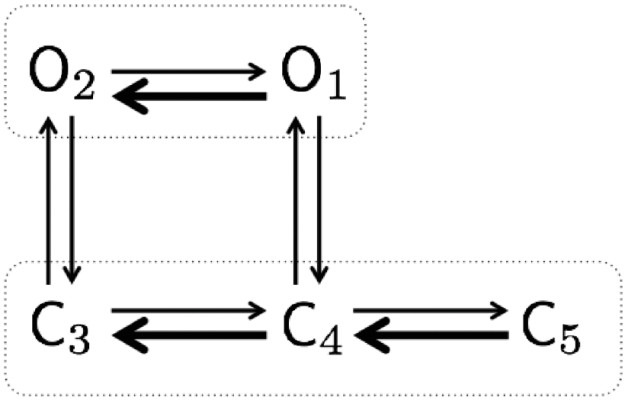
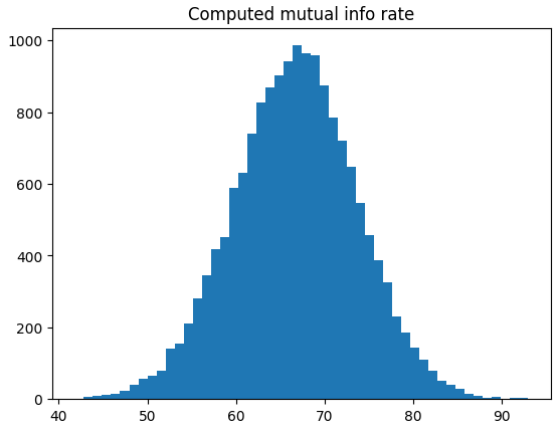
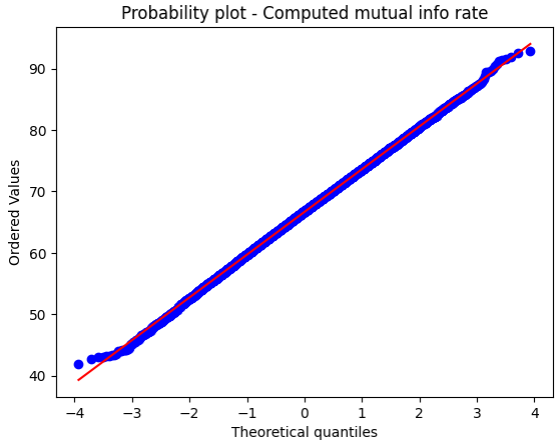


Figure A.1: State transition graph of Acetylcholine receptor (ACh). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition.



(a) Histogram of the simulation output of ACh

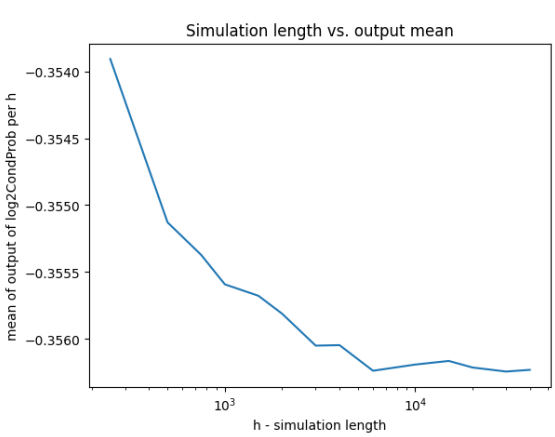


(b) Probability plot of the simulation output of ACh

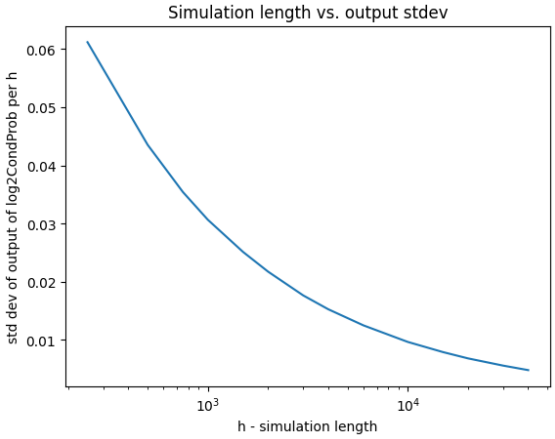
This make sense citing the central limit theorem, as the metrics output from the simulations are each a sum of the values of many independent and identical draws from some distribution. Even though the input/output state sequence vectors has their elements each dependent on previous one, the overall system is ergodic i.e. the input/output structure is non-changing. Thus, the mutual information rate measured/calculated from each one input/output sample in a time step manifests as the aforementioned independent and identical distribution draws.

Parameter	symbol	value
Simulation runs amount	`w`	16384
Length of each simulation	`h`	(varies)
Time step	`dt`	0.0005
Probability of 'low' on input	`p_L`	0.65

`h`	mean	stdev
4000	66.7274	7.7561
12000	66.3616	4.5017
24000	66.2606	3.1402
36000	66.1975	2.6002
48000	66.1697	2.2342
60000	66.1870	2.0201



(c) simulation length vs. mean



(d) simulation length vs. standard deviation

Figure A.2: Effect of simulation length on the distribution of Conditional Probability value

simulation length vs. output spread An inverse relation between simulation length and standard deviation is observed. This makes sense citing the core relation regarding sample size in central limit theorem. Moreover, the mean mutual info rate seems to trend downward with signs of

converging as simulation length increases, with the longer simulation smoothen out this error. This error is ultimately resolved by fixing a combination of an off-by-one error regarding the simulation length and incorrect initial condition, thus the initial condition of each simulation stream is not statistically null.

A.1.2 TCR KS-KP

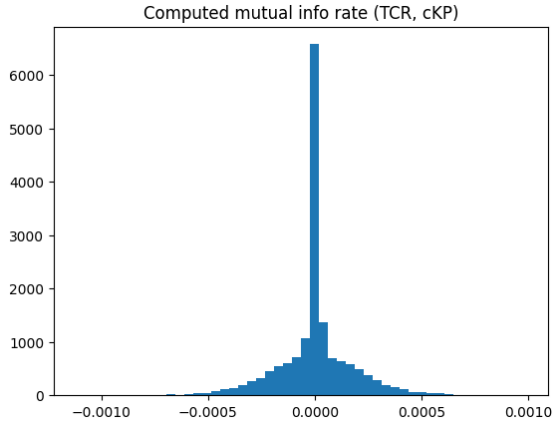
For TCR, initial explorations and validation runs are done using the KS-KP model which have more states and transitions and is thus more computationally intensive. This allow us to establish an upper bound of computational time needed for either model which lets us stop attempts with setups that would have taken unrealistic amount of time. A simulation with the following parameters is ran to obtain the following results.

Parameters:

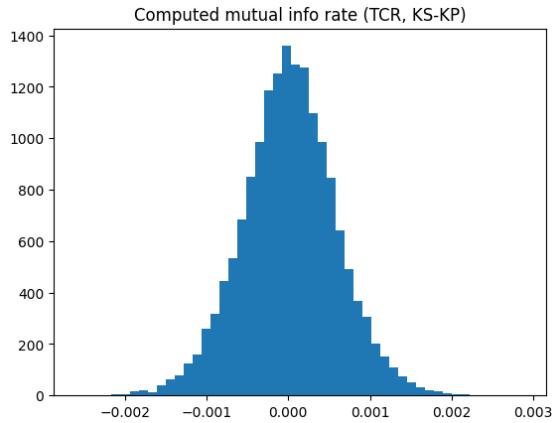
Parameter	symbol	value
Simulation runs amount	`w`	16384
Length of each simulation	`h`	200000
Time step	`dt`	0.0005
Total concentration of ligands	`M`	330
Proportion of activating (non-self) ligands when the 'input' is 'high'	`G`	0.08
Phosphorylation rate	`k_p`	1.0
Dephosphorylation rate	`k_dp`	5.0
Probability of 'low' on input	`p_L`	0.65

Result:

model	mean	std deviation	S-W statistic
cKP	1.0455326203786616e-06	0.00017281743816607086	0.9109601382746773
KS-KP	1.1866494982469654e-05	0.0005800097102858504	0.9976934309293832

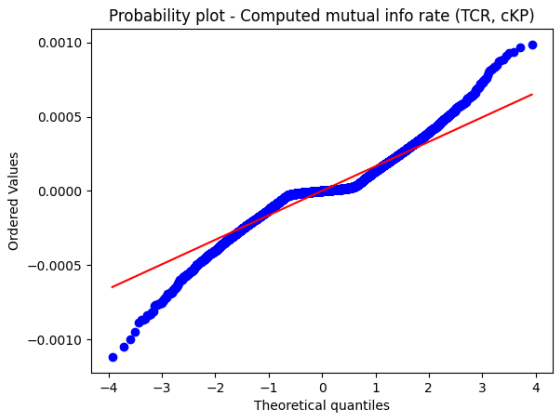


(a) Histogram of the simulation output of TCR (cKP)

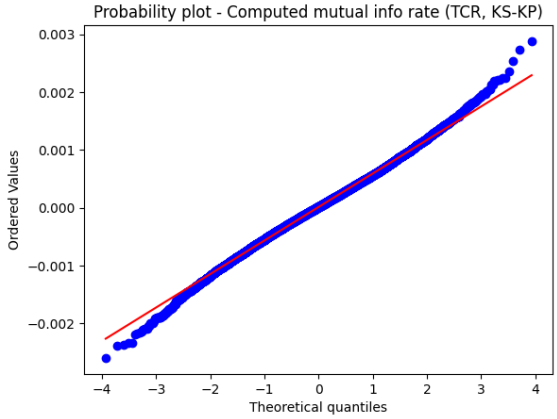


(b) Histogram of the simulation output of TCR (KS-KP)

Figure A.3: Histogram of the simulation output of two models of TCR, 200000 time steps



(a) Probability plot of the simulation output of TCR (cKP)



(b) Probability plot of the simulation output of TCR (KS-KP)

Figure A.4: Probability plot of the simulation output of two models of TCR, 200000 time steps

The reason of of the huge spike at 0.0 may be due to the more extreme distribution of overall state probability in the cKP model, where some instances in the ensemble never experience an observable output state of 1 due to the very low probability of reaching to an activated state. Thus, the overall rate of mutual information of that simulation instance stays exactly at 0, as

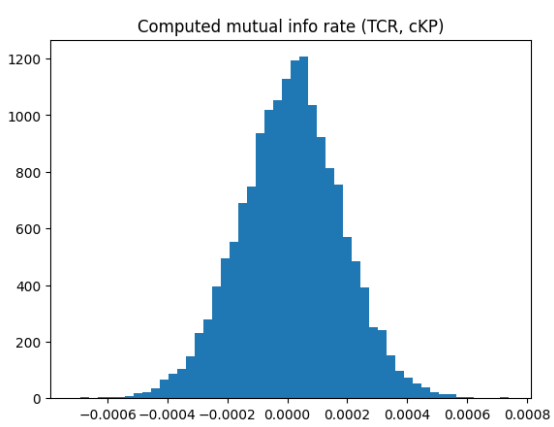
no change at all in the channel output means no information is transmitted. Therefore, one can hypothesize that using larger time steps can lead to more well behaved distribution.

Then, we run the simulation with a larger time step Δt with the following parameters:

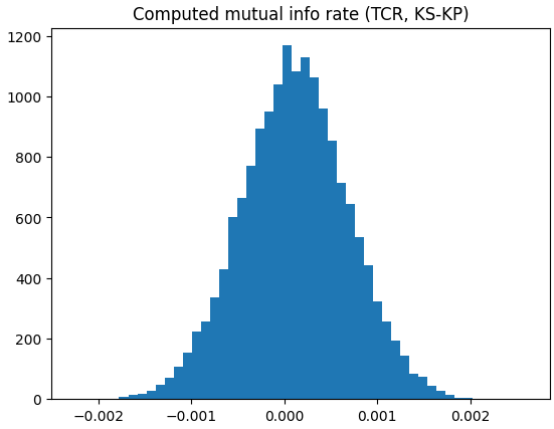
Parameter	symbol	value
Simulation runs amount	`w`	16384
Length of each simulation	`h`	200000
Time step	`dt`	0.005
Total concentration of ligands	`M`	330
Proportion of activating (non-self) ligands when the 'input' is 'high'	`G`	0.08
Phosphorylation rate	`k_p`	1.0
Dephosphorylation rate	`k_dp`	5.0
Probability of 'low' on input	`p_L`	0.65

Result:

model	mean	std deviation	S-W statistic
cKP	1.111899284795012e-05	0.00017121070698023462	0.9991091401236779
KS-KP	0.00011206704984570402	0.0005748048066800736	0.9998324977371903



(a) Histogram of the simulation output of TCR (cKP)



(b) Histogram of the simulation output of TCR (KS-KP)

Figure A.5: Histogram of the simulation output of two models of TCR, 200000 time steps

Here our hypothesis turn out to be true, with a distribution that visually looks closer to being a bell curve and Shapiro-Wilk statistic much closer to 1.0. However, one thing worth noting is that the rate of mutual information increased by an order of magnitude with the standard deviation of the distribution being very similar.

A.2 Mutual Information Rate as $\Delta t \to 0$ of receptors

A.2.1 Calmodulin receptor (CaM), i.i.d. input

Calmodulin receptor (CaM) is another receptor channel described in [5]. Its state graph can be seen in figure A.6. It have many transitions that are sensitive to input parameter, with some situated between microstates belonging to the same macrostate, and some going across different macrostates.

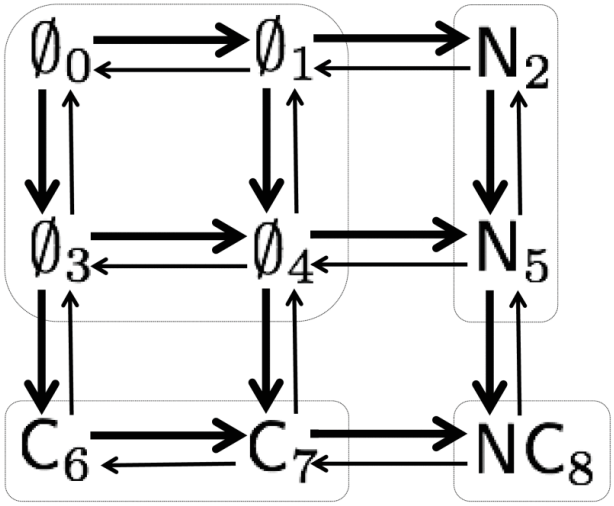
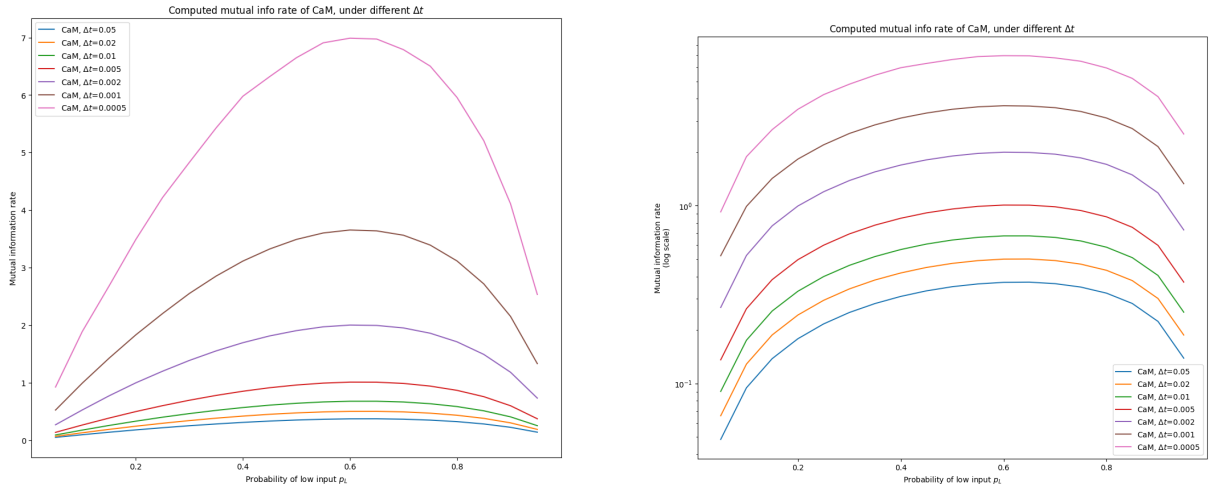


Figure A.6: State transition graph of Calmodulin receptor (CaM). Dashed boxes group together microstates to macrostates, and bold arrow is sensitive transition.



(a) Mutual information rate of CaM (linear scale)

(b) Mutual information rate of CaM (log scale)

Figure A.7: Mutual information rate of CaM varied by probability of low input p_L and time step length Δt in i.i.d. input.

As seen in figure A.7, the mutual information rate of CaM does not seem to collapse to 0 when $\Delta t \to 0$, but this also could have been that our experiments has not yet crossed the Δt value that will lead to the maximum mutual information rate, after which the rate will decrease to 0 as Δt approaches 0.

This result contributes to the discovery of the theorem of hidden sensitive channels as CaM is a receptor channel with both hidden and non-hidden transitions, while its mutual information rate does not collapse to 0.

A.2.2 TCR KS-KP

Here we start with the less numerically extreme but more computationally intensive model of KS-KP, and examine the effect Δt have as it becomes smaller to see what values it converges to and what parameter constants achieve maximum capacity.

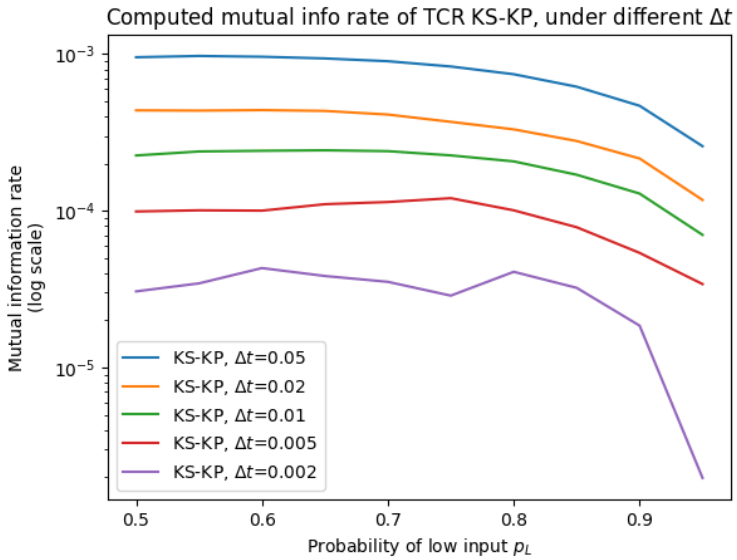


Figure A.8: Mutual info rate of TCR KS-KP, given different Δt

However, this preliminary result shows that the choice of Δt value significantly affect the result of the TCR Monte Carlo simulation, but instead of converging to a finite but positive value, it collapses to 0 as $\Delta t \to 0$. This is interesting as [26] prescribes that, in the discrete time channel model of a biological receptor, maximum capacity is achieved by identically and independently distributed inputs with time step Δt approaching 0.

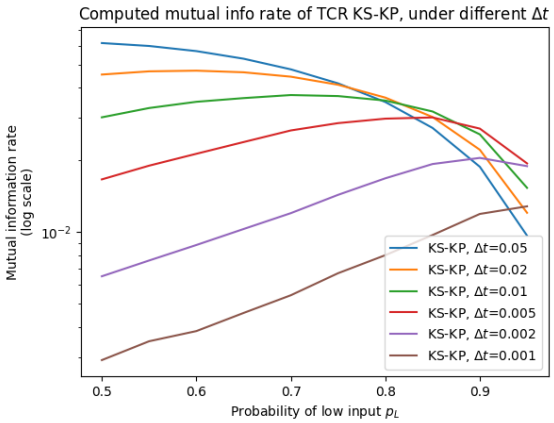
To resolve this contradiction, either the simulation program has bugs, or the previously established theories are incorrect. After extensive testing, code inspection, and cross referencing output between the previous and current implementation on the same input, it is shown that the two versions are exactly equivalent in each code function and no logical bugs has been detected.

We further tried using 64-bit float variables, as well as running the simulation with much higher ligand concentration, or lower phosphorylation step count to 1, to increase the overall probabilities of triggering a activation, thus to eliminate potential rounding errors or numerical artifacts due to the small Δt.

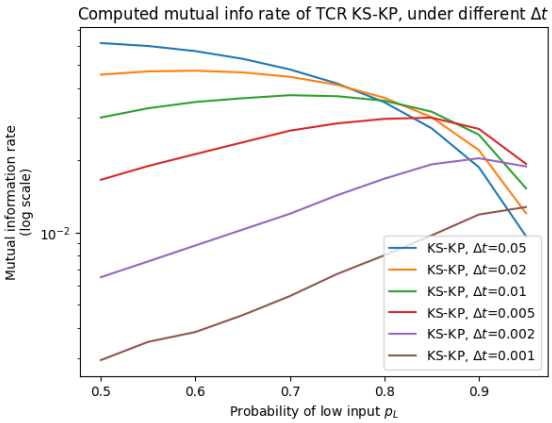
Parameters:

Parameter	symbol	value
Simulation runs amount	`w`	131072
Length of each simulation	`h`	1000
Time step	`dt`	varies
Total concentration of ligands	`M`	5000
Proportion of activating (non-self) ligands when the 'input' is 'high'	`G`	1.0
Phosphorylation rate	`k_p`	1.0
Dephosphorylation rate	`k_dp`	1.5
Probability of 'low' on input	`p_L`	varies
Number of phosphorylation to activate	`n`	3

In this trial, we further promote the overall probability of being able to trigger a binding state by greatly reducing the dephosphorylation rate and increasing `G` to 1.0, meaning, when input is 'high', the probability of an ingress ligand being non self is 100%. Moreover, trials using `float32` instead of `float64` can be tested to see if the phenomena of mutual information rate curve going down instead of converging with a clear maximum is indeed due to floating point rounding errors.



(a) Mutual info rate of TCR KS-KP, given different Δt, using `float64`



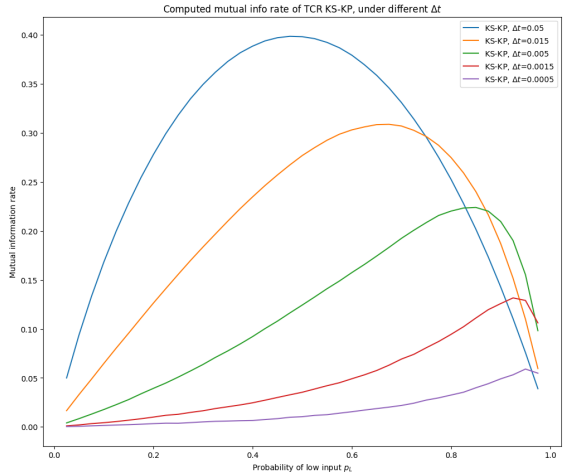
(b) Mutual info rate of KS-KP, given different Δt, using `float32`

Figure A.9: These ones are ran with higher ligand concentration: $M = 5000$.

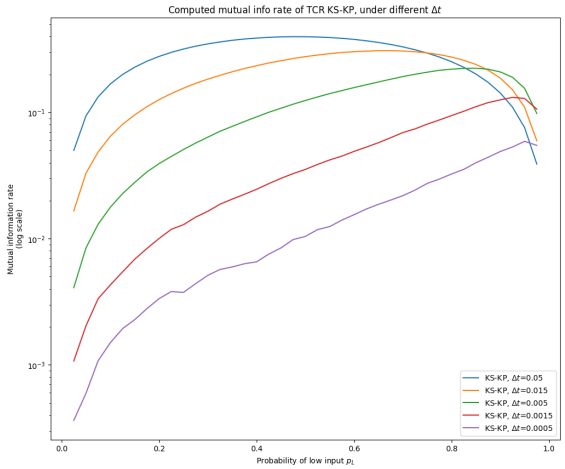
However, we see that there are essentially no difference between the two graphs with their sequence of curves. Note that the results should theoretically be identical due to using the same seeds for random states generation. This show that the lower curve might not be a floating point rounding error. Moreover, the time taken to run the simulations has been halved when using `float32`.

Parameters:

Parameter	symbol	value
Simulation runs amount	`w`	131072
Length of each simulation	`h`	1000
Time step	`dt`	varies
Total concentration of ligands	`M`	5000
Proportion of activating (non-self) ligands when the 'input' is 'high'	`G`	1.0
Phosphorylation rate	`k_p`	6
Dephosphorylation rate	`k_dp`	12
Probability of 'low' on input	`p_L`	varies
Number of phosphorylation to activate	`n`	1



(a) Mutual information rate of TCR KS-KP with 1 phosphorylation (linear scale)



(b) Mutual information rate of TCR KS-KP with 1 phosphorylation (log scale)

Figure A.10: These ones also have less realistic parameters: $M = 5000$, $k_p = 6$, $k_{dp} = 12$.

Here we see that, even with highly unrealistic parameters regarding ligand concentration, phosphorylation rate, and phosphorylation step count to promote the overall probability of activated state being triggered, the fact of which the mutual information rate becoming 0 as $\Delta t \to 0$ remains, showing that this phenomenon is not due to a rounding error.

This result is the initial moment where the theorem of hidden sensitive channels is discovered. After an extensive search in the code base and model definitions for errors it is concluded that having its mutual information collapse to 0 is just how TCR KS-KP behaves. Soon the connection that ACh, TCR cKP, and TCR KS-KP are all receptor models that has only sensitive transitions that are hidden, was made and ultimately led to the discovery and formalization of the hidden sensitive theorem.

Appendix B

Calculating Matrix Exponential Efficiently

In practice, in order to compute the mutual information rate through the continuous time system, we need to compute $\exp(H_y R \cdot \delta t)$ for different sojourn lengths δt and the same few $H_y R$, where this R represents any of R_L, R_H, W_U, the rate matrices, along with H_y, the mask for mapping to macrostates.

For any square matrix B, one can compute $\exp(B \cdot t)$ for arbitrary t and the same B repeatedly in an efficient way using the eigen diagonalization of B.

$$B = V \Lambda V^{-1} \tag{B.1}$$

$$\exp(B \cdot t) = V \exp(\Lambda \cdot t) V^{-1}. \tag{B.2}$$

Where Λ is the diagonal matrix of eigenvalues of B, and V is the matrix of eigenvectors of B. From this, we can just cache V, $\lambda = \text{diag}(\Lambda)$ (the eigenvalues just as a vector/1D tensor), and V^{-1} for each possible $H_y R$ in the code memory and we can easily compute any $\exp(H_y R \cdot t)$ in every sojourn.

This worked very well for receptors like ACh. However, it completely collapses when working with TCR cKP and TCR KS-KP.

Due to the abundance of one-way transition in the state graph of both TCR KS-KP and TCR cKP, the rate/Markov matrix will have many eigenvectors that are almost parallel, causing V^{-1} to have elements with extremely large values. Thus, when they are acting on a vector, this leads

to numerical instability where rounding errors are magnified and the message vector end up with unreasonable values. Therefore, a different method of computing $\exp(B \cdot t)$ efficiently is needed.

To devise the new method, note the following property of exponential functions, given any $A, B \in \mathbb{R}^{n \times n}$ and $t_1, t_2 \in \mathbb{R}$:

$$\exp(A + B) = \exp(A)\exp(B) \tag{B.3}$$

$$\text{if} \tag{B.4}$$

$$AB = BA, \tag{B.5}$$

And certainly two matrices created from the matrix multiplied to different scalars commute:

$$(t_1 B)(t_2 B) = (t_1 t_2)B^2 \tag{B.6}$$

$$= (t_2 t_1)B^2 \tag{B.7}$$

$$= (t_2 B)(t_1 B), \tag{B.8}$$

Therefore, we have:

$$\exp(B \cdot (t_1 + t_2)) = \exp(B \cdot t_1 + B \cdot t_2) \tag{B.9}$$

$$= \exp(B \cdot t_1)\exp(B \cdot t_2). \tag{B.10}$$

Knowing this, we can build a cache of exponentials $\exp(B \cdot 2^m)$ for magnitudes

$$m \in \{m_{min}, ..., -2, -1, 0, 1, 2, ...m_{max}\}, \tag{B.11}$$

to be combined to calculate $\exp(B \cdot t)$ for any t within a reasonable range.

$$t = \sum_{m=m_{min}}^{m_{max}} r_m 2^m, \qquad r_m \in \{0, 1\} \tag{B.12}$$

$$\exp(B \cdot t) = \prod_{m=m_{min}}^{m_{max}} C_m \tag{B.13}$$

$$C_m = \begin{cases} \exp(B \cdot 2^m) & r_m = 1 \\ I & r_m = 0. \end{cases} \tag{B.14}$$

Moreover, if we use a different base $b > 2$ for the cache, and precompute $\exp(B \cdot rb^m)$ for every radices $r \in \{1, 2, ..., b-1\}$ in every magnitude $m \in \{m_{min}, ..., -2, -1, 0, 1, 2, ...m_{max}\}$ to be combined, we can effectively 'trade memory cost for computational speed'.

Using this method, we can now fully implement the continuous time simulation system even as the state graph of the system is abundant in one way transitions.