From Biology To Algorithms

Nicholas David Lindsay Owens

Submitted for the degree of Doctor of Philosophy

University of York

Department of Electronics

June 2010

Abstract

This thesis describes a path from a model of a biological system to a biologically-inspired algorithm. The thesis commences with a discussion of the principled design of biologically-inspired algorithms. It is argued that modelling a biological system can be tremendously helpful in eventual algorithm construction. A proposal is made that it is possible to reduce modelling biases by modelling the biological system without any regard to algorithm development, that is, with only concern of understanding the biological mechanisms. As a consequence the thesis investigates a detailed model of T cell signalling process. The model is subjected to stochastic analysis which results in a hypothesis for T cell activation. This hypothesis is abstracted to form a simplified model which retains key mechanisms. The abstracted model is shown to have connections to Kernel Density Estimation, through developing these connections the *Receptor Density Algorithm* is developed. By design, the algorithm has application in tracking probability distributions. Finally, the thesis demonstrates the algorithm on a related but different problem of detecting anomalies in spectrometer data.

Contents

List of Tables

List of Figures

Acknowledgements

My first thanks must go to my supervisors Jon Timmis and Andy Tyrrell for their guidance and allowing me to investigate the varied topics of this thesis. Then, to Andy Greensted for his support and making me feel welcome in the department. I would also like to thank Susan Stepney for her guidance before my PhD and for giving me Penrose tiling distractions during my PhD.

I am indebted to my family for their support and remarkable efforts through the years.

To to all my friends in York, particularly Ralph for driving me to Mair's Delicatessen. A special thanks must go to Paul and Helena for their tremendous generosity in putting a roof over my head in the final months.

Finally, thank you to Sioned and Dunja for proofreading and cycling. Especially to Dunja for all her support.

Declaration

This work is the author's own with the exception of the receptor density algorithm robot implementation. Some of this work has already been presented in [Owens et al., 2008], [Owens et al., 2009], [Owens et al., 2010].

CHAPTER 1

Biologically-Inspired Computation

A *Biologically-Inspired* computational system is an information processing system whose structure and function have been designed with inspiration from a biological system. Systems designed in this manner intend to replicate properties of a biological system, and aim to deliver new perspectives and solutions to computational information processing problems. *Bio-inspired algorithms* are a class of biologically-inspired systems restricted to algorithmic concerns.

The development of a bio-inspired algorithm is the topic of this thesis. Particularly, inspiration will be taken from the *immune system* and so the resultant algorithm is known as an *Artificial Immune System*. This chapter focuses on the principled design of bio-inspired algorithms and the connection between a biological system and a resultant bio-inspired algorithm.

The chapter is organised as follows: Section 1.1 offers some history, along with motivating arguments and challenges for the bio-inspired development of systems. Section 1.2 introduces current bio-inspired algorithm design methodologies which outline a need for modelling biology when building algorithms. The section also gives a description of the current state of Artificial Immune Systems and discusses the sources of its inspiration. Section 1.3 presents factors important when modelling biological systems, and particularly when modelling for algorithm development. Section 1.4 discusses the issue of moving from biology to algorithms. Finally, Section 1.5 provides an outline of this thesis.

1.1 History, Motivation and Challenges

Walter Bradford Cannon in his 1932 book entitled *The Wisdom of the Body* [Cannon, 1932] defined the term *homeostasis* to describe the ability of biological systems to maintain their internal state despite a largely varying external environment[1]. Toward the end of the

[1]Homeostasis is the concatenation of *homeo-* meaning similar and *-stasis* meaning standing, both from Greek. The origins of the word's concept were well established before its conception. The prominent example being the work of Claude Bernard and his definition of the constancy of internal environment, *milieu intérieur* [Bernard, 1927].

introductory chapter that defines homeostasis, is the following less well known paragraph, [Cannon, 1932]:

> It seems not impossible that the means employed by the more highly evolved animals for preserving uniform and stable their internal economy (i.e., for preserving homeostasis) may present some general principles for the establishment, regulation and control of steady states, that would be suggestive for other kinds of organization – even social and industrial – which suffer from distressing perturbations. Perhaps a comparative study would show that every complex organization must have more or less effective self-righting adjustments in order to prevent a check on its functions or a rapid disintegration of its parts when it is subjected to stress. And it may be that an examination of the self-righting methods employed in the more complex living beings may offer hints for improving and perfecting the methods which still operate inefficiently and unsatisfactorily.

Remarkably, this quotation from 1932 exactly describes the bio-inspired endeavour, and moreover presents themes in structural correspondence and fault-tolerance. These are topics at the forefront of current biologically-inspired endeavours [Levi and Kernbach, 2010]. Continuing historically, the work of Turing and Von Neumann (names traditionally associated with conventional computational approaches) was concerned with biological inspiration and modelling [Turing, 1952, 1992; Neumann, 1966]. The origins of the majority of today's bio-inspired fields were not, however, developed until much later[2]. Now research fields include and are not limited to, Evolutionary Computing [Mitchell, 1998]; Artificial Neural Networks [Bishop, 1995]; Artificial Immune Systems [de Castro and Timmis, 2002; Timmis et al., 2008a]; Swarm Intelligence Systems [Bonabeau et al., 1999] and Artificial Life [Bedau, 2003].

1.1.1 Motivation

There are two very general properties of biological systems that are responsible for the interest of computer scientists in biology[3]. The first is the *robustness* of biological systems and is related to homeostasis. Robustness is the property that allows a system to maintain its function against internal and external perturbations, and is the "ubiquitously observed property of biological systems" [Kitano, 2004]. All biological systems must be "robust" enough to survive sufficiently long to reproduce. Further discussions on the origins of robustness and its relation to evolution are given in Section 1.1.2.

As an aside, the definition of robustness is very similar to that of homeostasis. Many authors have felt the general interpretation of homeostasis and its suffix *-stasis*, to be too static. In order to convey the notion a dynamic stability, such as a return to a trajectory when perturbed, new definitions, predominantly suffixes to *homeo-* have been defined. A notable example is *homeorhesis* with *-rhesis* meaning flow [Waddington, 1957; Saunders,

[2]This delay is in part due to the wait for appropriate computational power to be developed.

[3]There are many reasons why computer scientists are interested in biology, the concepts of robustness and self-organisation are two that currently stand out.

1993]. However, these concepts are often invoked in sufficiently general situations that such semantics do not matter and robustness may be the least loaded of the terms. The second important property of biological systems is *self-organisation*; [Camazine et al., 2001] defines a self-organised system as one that exhibits patterns/structure at a global scale that are the result of many interactions at a lower scale. The defining feature is the lack of central control. The global order is not encoded explicitly at the any level, but emerges implicitly from local interactions between components. Such self-organised patterns are known as *emergent* phenomena, and cannot be understood by examination of individual components alone. The lower scale components are often heterogeneous in structure. In biological systems self-organised phenomena are observed as one moves up each level of physical and temporal scale (for example, molecules to cells, cells to organs, organs to organisms, ...) [Cohen, 2000]. Generally conventionally engineered systems do not display this manner of robustness and self-organisation, and a desire to have these properties has sparked quite considerable engineering effort [Levi and Kernbach, 2010].

The immune system is an excellent example of a robust, self-organised system [Cohen, 2000], and worth special mention as the topic of this thesis. The system has quite remarkable abilities: recognition/discrimination [Germain and Stefanová, 1999]; maintenance [Cohen, 2000]; inference from danger/context [Matzinger, 2002]; and memory [Murphy et al., 2008]. The specifics of these are largely responsible for the body of work in artificial immune systems to date [Timmis et al., 2008a]. Immunologists have even taken computational interpretations of the immunology [Cohen, 2007]. The immune system is regarded to take the state of the body as input and *computes* an immune response as output. This view of a biological system as an *information processing* system is of clear relevance to a bio-inspired endeavour. Though, as word of caution, computation is in the eye of the beholder; simply interpreting a biological system performing computation does not imply any eventual contribution to biologically-inspired computation.

The concept of biological inspiration is not limited to computational concerns. A first example is mathematics [Sturmfels, 2005; Cohen, 2004]. The references discuss that a continued analysis of biology will not only further biological fields, it will stimulate new mathematics. This manner of inspiration is closely related to that of computation. In a second example, and in contrast to computational engineering discussed above, classical engineering too has received biological inspiration [Ingber, 2010]. Related, and in some respects the dual to biological inspiration, is performing computation with biological systems. For example, slime mould is used to solve routing problems in [Adamatzky and Jones, 2009].

1.1.2 Challenges

There are details of biology which inhibit the abstraction of biological detail to conventionally executable information processing. This presents no problems if the objective is to discover new approaches to information processing without the need to turn them into serviceable algorithms. But, challenges are presented if the resultant algorithm is to be tractable in computational complexity in time and space.

Biological systems are evolved, and evolved systems do not progress in the sense of betterment, they increase in complexity [Cohen, 2000; Edelman and Gally, 2001]. The existence of a creature, cell or molecule provides the opportunity for the interaction with another creature, cell, or molecule. Evolution copies and mutates; a mutation may warrant a new interaction, but the old interaction still exists by the non-mutated copy. Over time, evolution will make systems more complex [Cohen, 2000; Edelman and Gally, 2001]. Closely related to this evolved complexity is *degeneracy*, this is the ability of structurally different components to perform the same function [Cohen, 2000]. Importantly, degenerate components also perform different functions; there is seldom a one-to-one mapping between function and component in biology. An example from genetics is the surprisingly small effects that gene knockout (an experiment which involves the removal of specific genes) can have on an organism [Edelman and Gally, 2001]. This is so even when the gene exhibits a measurably active behaviour under normal conditions [Cohen, 2000]. This fact is related to degenerate genes, and also (as discussed in [Cohen, 2000]) organisms self-organise their genes during development. The absence of a gene forces an organisation around a different gene. The *development* of an organism is a key process [Cohen, 2000; Waddington, 1957; Saunders, 1993]. It seems that the continual tampering of evolution (combined with environmental perturbation) has forced biological systems to become robust to this sort of tampering and perturbation. The implication for bio-inspired algorithms: the processes that make the system interesting actually act against their ability to be understood and abstracted, and an algorithm may be required to "develop".

Biological systems process information through the substrate of chemistry, and this has quite different properties to conventional computational substrates. For example, immunology renders the information contained in the conformation and chemical composition of certain molecules interpretable by molecular-recognition so-called *antigen-receptors* (Chapter 3 provides the full background). The details of the molecular-recognition involve a formation of a complicated non-covalent bond [Carneiro and Stewart, 1994]. It is this step that allows the immune system to perform comparisons between effectively any molecule, and molecular-recognition forms an important step in many information processing pathways. It would seem to be a desirable property, yet this molecular-recognition step is not easily replicable in an algorithm. This point has caused consternation in many Artificial Immune Systems (see [Stibor, 2006; McEwan and Hart, 2009] for description) and will be the cause for discussion in this thesis. The point on substrates applies in the opposite direction: biology may have to work hard to perform certain tasks a conventional computational system can perform easily.

The discussion now continues with an overview of methodologies with which to design and build biologically-inspired algorithms.

1.2 Methodologies and Artificial Immune Systems

This section describes methodologies for the design of bio-inspired algorithms, and their discussion necessarily involves historical considerations. The proceeding discussion will

be in terms of immunology and Artificial Immune Systems (AIS), although, the general concepts are applicable to any domain of biology and bio-inspired computing.

Early work in AIS had interdisciplinary origins and concerned furthering immunological understanding in terms of computational concerns [Farmer and Packard, 1986; Bersini, 1992]. However, as [Stepney et al., 2005] points out, as the field grew the quality of the biological inspiration diminished. The inspiration was often naive and many algorithm developers were accused of "reasoning by metaphor" [Stepney et al., 2005]. Algorithms were both poorly understood and unable to produce the behaviour of the biology that inspired them. An account of this can be found in the majority of recent AIS review papers [Hart and Timmis, 2008; Timmis, 2007; Timmis et al., 2008a].

To move the field back towards its original track, [Stepney et al., 2005] presented the *conceptual framework* for the development of artificial immune systems. The framework advocated careful biological modelling and abstraction to precede the development of algorithms. Further, it was proposed that the modelling work should be backed up by a solid theoretical analysis of the resulting algorithm. The steps of this framework are discussed in Section 1.2.1. In [Stepney et al., 2005] it is also argued that this is necessarily an interdisciplinary process and requires the collaboration of biologists, mathematicians and computer scientists, at the least. The outcome of an instantiation of the conceptual framework should be a well understood algorithm with a traceable route to back to biology. Moreover, the principled manner in which algorithms are developed should allow contributions back to biology; the whole process should enrich both fields. It is also suggested by [Stepney et al., 2005] that once an algorithm is developed one can apply a *meta-framework* to compare commonalities between bio-inspired algorithms. This suggestion is in part related to the observation that many population-based bio-inspired algorithms have very similar methods [Newborough and Stepney, 2005]. This idea of comparison between models is returned to in Section 1.4.

A recent thesis [Andrews, 2008] discusses the conceptual framework in detail and performs instantiations of the framework, moving from biology to algorithms. In reflections on the framework, [Andrews, 2008] notes that the the conceptual framework provides good, if brief, advice and also draws attention to a circularity: if one wishes to design a bio-inspired algorithm with an application in mind, how does one know which system to model without modelling first? The question of identifiability of systems is returned to later in this section and in Section 1.3. For the present, it is noted that when turning to biology for algorithmic ideas one should expect more work than developing an algorithm conventionally.

In work following from the conceptual framework, [Timmis et al., 2008b] define the more application oriented *immuno-engineering*. It supplies the conceptual framework with additional elements related to engineering. These largely consist of the acknowledgement of constraints of the intended application, so that given these constraints biological detail can be discarded accordingly. The implicit message is to be inspired by the biology, and not subject to its limitations. This is a theme common in much artificial immune systems work. Further, [Timmis et al., 2008b] states that the modelling for bio-inspired algorithms helps delineate which aspects are necessary to computationally re-generate a biological system's

behaviour, and which aspects are surplus. In [Timmis et al., 2008b] it is also advised that developing immune algorithms applicable to a generic computational problem may be difficult, and it may be more realistic to ground algorithm development in a specific application. In a similar vein to the conceptual framework of [Stepney et al., 2005], and possibly in a stronger manner, the work defining immuno-engineering [Timmis et al., 2008b] emphasises the importance and potential of feeding results back to contribute to a biological field.

Certainly, this chapter agrees with the concept of immuno-engineering in that one must carefully consider the application when applying algorithms. However, the chapter also expresses clarifications and caveats to some of the statements in the papers on immuno-engineering and the conceptual framework. The first of these will come in the following discussion on the connection between bio-inspired computation and biology.

1.2.1 Connections to Biology

The conceptual framework paper [Stepney et al., 2005] supplies a diagram (replicated in [Timmis et al., 2008b]) which describes the path between biology and algorithms. This diagram, with alteration and simplification, is given in Figure 1.1. The alteration draws attention to the connection between biology and bio-inspired algorithms. The left of the figure describes the biological research, the right bio-inspired algorithm development. The defining property of the division is that all right side models contain detail or simplification whose sole purpose is due to an intended bio-inspired algorithm. Thus, all models on the left side have only biological concerns, and all assumptions must be made subject to biological justification. Note that a biological model whose final purpose is an algorithm may start on the left or right.

In any left-starting bio-inspired algorithm development there must come a point where computational concerns force the model from left to right. It is possible to bypass the abstracted model and jump straight to algorithms. However, one would imagine that this would produce unsatisfactory results unless the model of biology was exceedingly abstract and general, see Section 1.3.

Returning to a left-starting model that shifts to the right to become an abstracted model. Once on the right the purpose of the abstracted model is to allow biological detail to be stripped away, heading toward an algorithm, whilst retaining the desired behaviour of the model. Assumptions can be made that are biologically invalid (though there must come a point where it ceases to be bio-inspired). The node A on the boundary of Figure 1.1 represents all implicit and explicit details in the abstracted model which are not biologically justifiable. Making A explicit brings to the forefront the connection between the biological system and the bio-inspired algorithm. If an algorithm designer wishes to contribute back to biology from this abstracted model/algorithm stage, the contributions must translate back through A. Moreover, one can make use of assumptions A understanding the properties that have been transferred from the biological model to the abstracted algorithm. The purpose of node A is to comprehend what is gained and lost when one transits through A

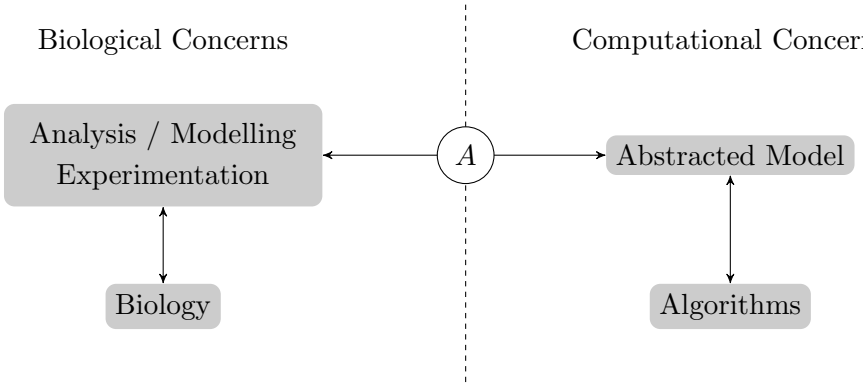

Figure 1.1: The path from biology to bio-inspired algorithms. Adapted from the diagram in [Stepney et al., 2005]. And, as noted in [Stepney et al., 2005] all arrows come with a necessary *bias*. See text for description of figure.

in either direction. It should be noted that the details made explicit by A are, for the most, implicit in the discussions of frameworks in [Stepney et al., 2005; Timmis et al., 2008b].

1.2.2 The current state of Artificial Immune Systems

It is worth giving a very brief summary to the current state of AIS, which will actually lead to suggestions of the modelling required to develop new AIS. As noted above, recent AIS review papers [Hart and Timmis, 2008; Timmis, 2007; Timmis et al., 2008a] echo the sentiment of the conceptual framework [Stepney et al., 2005]. In fact, [Timmis, 2007] describes AIS as having reached an impasse. The reasons for this become clear when one regards the source of immunological inspiration. For effectively all *pre*-conceptual framework AIS, the source of inspiration has been the major general theories of the immune system of the last half-century: *Clonal Selection Theory*, including *Clonal Expansion & Clonal Deletion* and *Immune Network Theory*. The details of some of these theories are discussed in Chapter 3; for the other theories and descriptions of associated algorithms the reader is directed to [Timmis et al., 2008a]. The concept of general theories and models is discussed in more detail in Section 1.3, the key detail of a general model is that it applies without specifics and so it translates well to algorithms. AIS inspiration sources *since* the conceptual framework have been based on *Danger Theory* and the *Tuneable-Activation-Threshold Hypothesis*, another two general theories (again, see Chapter 3 for descriptions). It should be noted that the algorithms developed from danger theory and the tuneable-activation-threshold hypothesis were developed with more detailed biological inspiration. The danger theory algorithm [Greensmith et al., 2010] was developed in conjunction with immunologists investigating danger theory. It has had some success, but has suffered from a lack of theoretical understanding, although this is changing [Stibor et al., 2009]. The algorithm based on tunable-activation-thresholds arises from the aforementioned thesis [Andrews, 2008]. The algorithm shows promise, but is yet to be accompanied by a theoretical explanatory analysis.

The AIS community has effectively exhausted the general models of immunology, so the

next AIS will have to either come from a non-general theory of immunology, or it will have to revisit the general theories. Herein lies the problem: the lure of immunology is its desirable properties such as recognition, maintenance and memory; the difficulty of immunology is that many of these properties are readily experimentally demonstrable, but the biological mechanisms responsible for these properties are not well understood. As such they seem accessible concepts but they do not have the concrete grounding that algorithms require. As an example, immune memory can be reliably invoked; vaccinations are clear evidence of this. Yet there is no agreed upon general model that completely accounts for immune memory (though many models of immune memory exist). Thus, either new AIS will have to be built from existing modelling and analysis of less general aspects of immunology, and not claim the grandeur of the immune system's macroscopic properties, or, AIS will have to play a role in developing new general models of immunology. The former endeavour seems the more realistic, and there is no reason to think it will return any less in terms of algorithmic utility.

1.3 Modelling for Bio-inspired Algorithms

The discussions of the previous section have demonstrated the need for modelling when building a biologically inspired algorithm. This section discusses the issue of modelling biology in general terms and then discusses the issues when modelling for bio-inspired algorithms.

The complexities of biology discussed in Section 1.1.2 present challenges in modelling long before abstraction to algorithms is reached. However, an understanding of the strategies that may be employed when modelling biology provides a way forward. In particular, this discussion revolves around work of [Levins, 1966] and his discussion of model building in population biology. The ideas apply equally in any field of biology. In particular, the work describes an inherent trade-off in three so-called *desiderata* of biological models:*generality*, *realism* and *precision*. The thesis of [Levins, 1966] is that one cannot maximise all three and retain a useful model. The complexity of biology is such that maximising all three results in a model is that complex as to be incomprehensible. The work of Levins has been influential, and with influence came criticism [Orzack and Sober, 1993][4], a retort [Levins, 1993], and further commentary [Odenbaugh, 2002, 2005; Bullock and Silverman, 2008]. The process has clarified the intended message of [Levins, 1966], and so it is possible to define (still loosely) the following:

- *Generality:* a general model applies broadly to biology. For example, applying to all cell types is more general than just applying to one.
- *Realism:* a realistic model incorporates detail about the system. For example, models with simplifying assumptions that discard detail are less realistic. For example,

[4]The criticism of [Orzack and Sober, 1993] largely missed the point that the trade-off was in practicality. [Orzack and Sober, 1993] argue that one always write down another model which is more general, realistic and precise and so there can be no trade-off.

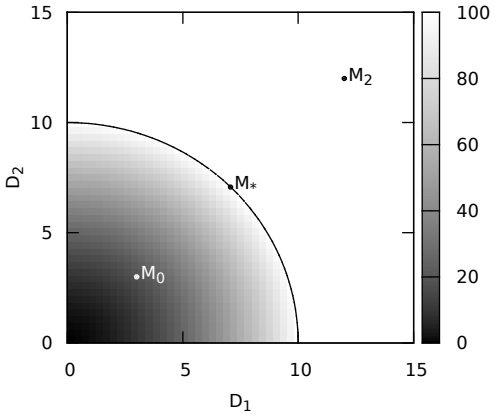

Figure 1.2: The trade-off between factors in modelling, reproduced from [Levins, 1993; Odenbaugh, 2002], only displayed for 2-dimensional D_1 and D_2. A model should try and maximise D_1 and D_2. Shading denotes the tractability of understanding of the model, *black* the simplest to intractable *white*. Thus, M_0 can be improved to $M_\star$, but M_2 is not a tractable model. All numbers, scales and shading are purely illustrative.

ignoring the volume change during the differentiation of a cell.

- *Precision:* a precise model makes specific predictions about the system. For example, the culmination of precision is a numerical model that can be computationally calculated.

In [Levins, 1993] it is stated that these are not the only three considerations, but perhaps the most important. The trade-off is described such that one can maximise two factors at the expense of a third. Three examples are given in [Levins, 1966], which here are converted to the immunological domain. A model that is realistic, precise and not general, is the exact numerical calculation of the short-term trajectories of a molecule in a biochemical network in a given cell type (see Chapter 3). A model that is general, precise and not realistic, is the calculation of required immune repertoire sizes from a shape-space model [Perelson and Weishbuch, 1997]. A model that is general, realistic and not precise, makes qualitative arguments that are often graphical, for example assuming functions are increasing, decreasing, concave or convex [Levins, 1966]. An example of the computational equivalent is the Unified Modelling Language (UML) used in [Read et al., 2009]. Figure 1.2 depicts the trade-off in two dimensions. As an alternative [Levins, 1966] suggests that rather than trying to satisfy all three requirements, one should develop a *cluster of models*. All cluster models are related, but have slightly different views of the system. A total picture of the system can be incrementally constructed which offers a more tractable route to understanding.

Apart from being good modelling advice, these ideas have implications when modelling for bio-inspired algorithms. If one develops a bio-inspired algorithm from an initial biological model, then the abstraction process (right Figure 1.1) corresponds to shifts in the generality, realism and precision. Note that abstraction is different to generality; abstracting a model does not make it more general, but general models may be easier to abstract. This is a manner in which bio-inspired algorithms can contribute back to biology: the abstraction of the biological model produces a new perspective on the model. Certainly, it is likely that

realism is lost, so the contribution of the model depends on the assumptions A (Figure 1.1), and if the new model is a simple abstraction it may offer nothing new over the biological model.

Regarding, the current inspiration for AIS discussed in the previous section, almost all previous inspiration comes from general models with varying levels of realism and precision. If one desires a new AIS it can be constructed from an existing model with any level of generality, reality or precision (this point is noted in [Andrews, 2008]). Note that it may require less knowledge of biology to develop a successful real and precise model, than a general one. Good real and precise models are obtainable by focusing on one particular area; to get a good general model often requires encompassing knowledge of many systems.

Given the arguments of the last section, on methodologies and the state of AIS, it seems necessary that the starting place for any new AIS must be on the left hand side of Figure 1.1. That is, the starting place must be a model for biology's sake, with no AIS concerns. Once the model is constructed, understood and ideally verified one can start the simplifications to shift to the right of Figure 1.1.

Recall the statement of [Andrews, 2008] that one cannot know which system to model without modelling. Certainly, it is necessary to chose a system by some heuristic. It is possible to "reason by metaphor" and chose a system the looks structurally appropriate for a solution to a given application. This must come with the caveat that there is no guarantee that the modelling work will result in a solution suitable for the application, especially when the model is realistic and precise. However, it may be too great a bias to have a specific application in mind. An alternative is to choose biology with respect to an *approximate application*, such that it appears to have the properties of a general class of problems. Then, the biological system can dictate the progress, rather than the application. A good rule of thumb on choosing systems is to select systems which look to be reliant on the structure of interactions, rather than the specific details of them (e.g. avoid systems which rely on molecular recognition). A further alternative is to select biological systems which look to be performing relevant information processing and then to choose an application based on the outcome of modelling. Clearly, the assigned application should be suitable compared with the biology. For example, if the model requires constant interaction between biological components and their environment then the application should also (see discussions of *embodiment* in [Stepney, 2007; Timmis et al., 2008a]).

1.4 Algorithmic Development

This section concerns the point at which the shift to the right of Figure 1.1 occurs. Modelling moves away from completely biological concerns towards abstract algorithm concerns. Clearly, these steps must have an aim, it is necessary to know where one is heading with the abstraction steps. This is the sentiment of [Timmis et al., 2008b], that it is easier to develop with an application in mind as it helps to abstract the model. However, the advice of Chapter 1 that it is better to have approximate application rather than the bias of a specific application. Again, biology should lead the process of abstraction, not the application.

All of the following: [Stepney et al., 2005; Timmis et al., 2008a,b,c; Stibor, 2008] have called for a greater use of theoretical results in AIS. The abstracted model stage on the right of Figure 1.1 is an ideal place to commence theoretical understanding of the nascent algorithm. Note that this process is simply an extension of the biological modelling, except the allowable steps have changed. Without including theoretical analysis the burden is placed on an empirical investigation of the algorithm. Simply because the concepts are well understood at the biological stage, does not imply that they will translate to the algorithm stage. When developing for a specific or approximate application, then one can ground the algorithm by comparison to the theoretical results of the area. This should result in a clear understanding of the algorithm, and its relation to existing techniques. This will evaluate the contribution of the algorithm.

It is of relevance to consider the link back to biology, and the set of assumptions A. Certainly, the development of an algorithm should play to the expected computational capabilities of the approximate application. Thus, it seems likely that A will grow in size as the algorithm is developed. If attempts to make the details of A are made explicit then an interesting opportunity arises. This relates to the meta-models of [Stepney et al., 2005] discussed in Section 1.2. That is, it is possible to make comparison of multiple algorithm design processes via the details of A. One would look for common transitions between biological models and abstracted models. If nothing else a library of such transitions could contribute to further new algorithm developments.

1.5 Outline of Thesis

This chapter has provided a brief introduction to bio-inspired algorithms. The focus has been on the presentation of the methodologies associated with artificial immune systems. This introduction has omitted a review of current AIS work, as stated earlier the reader is directed to [Hart and Timmis, 2008; Timmis, 2007; Timmis et al., 2008a]. An understanding of current AIS is not a prerequisite for understanding this thesis.

This thesis will design a new artificial immune system. The approach will adhere to the advice and discussion of this chapter. The source of immunological inspiration will be the T lymphocyte or T cell. This is chosen for to reasons related to its discrimination abilities and the aforementioned Tuneable Activation Threshold Hypothesis. In line with the discussions of this chapter, this thesis opts to investigate a realistic and precise model of T cells rather than a general one. The model will involve specific molecular interactions both inside and on the surface of the T cell. The thesis will be guided by the following hypothesis:

> *Through investigation and analysis of a realistic and precise model of T cell signalling it is possible to abstract properties and via an abstracted model design an algorithm which replicates the discrimination abilities of the T cell.*

The hypothesis will be addressed by the following:

- Analysis of an existing model of T cell signalling for the sake of the model, that is, without discussion of algorithms.

- The abstraction of results of analysis into an abstracted model of T cell signalling and identification of an *approximate application*. The details of assumptions A that connect the model and biology are made clear.
- Analysis of the abstracted model with respect to results of and methods of the *approximate application*, such that the resulting algorithm is well understood.
- Demonstration of the algorithm with validating theoretical results.

These steps are achieved in 5 chapters and an appendix:

Chapter 2 presents methods for modelling and analysis of biological systems. The focus is on tools associated with chemical kinetics.

Chapter 3 first provides an introduction to immunology and then a background to T cell signalling. Due to the constraints of this thesis an existing model of T cell signalling [Altan-Bonnet and Germain, 2005] is analysed, the model is certainly realistic and precise. The analysis uses the methods of Chapter 2 and the chapter culminates in a new hypothesis for T cell activation.

Chapter 4 constructs an abstract model of T cell signalling, the hypothesis presented in Chapter 3 dictates the detail included in this abstracted model. Careful attention is given to the set of assumptions A and the connection to the biological model. The *approximate application* of *anomaly detection* is assigned. Through analysis of the abstracted model and connections with techniques from the field of anomaly detection the *Receptor Density Algorithm* is developed. The algorithm is demonstrated on two classes of anomaly detection problem.

Chapter 5 applies the receptor density algorithm to a real-world problem: the anomaly detection in data generated by analytical chemistry devices, particularly spectrometers. Two case studies are presented: anomaly detection in mass spectrometry data and anomaly detection with a robot-mounted ion mobility spectrometer.

Chapter 6 presents conclusions of the thesis. A discussion is given of the relation between the algorithm and the biological model and implications for the biology discovered during algorithm development.

Appendix A provides supplementary material for the modelling work in Chapter 3.

1.5.1 Contribution of Thesis

The main contributions of this thesis are outlined below. A more detailed version of this list may be found in Chapter 6. The contributions are listed by chapter:

Chapter 3 The compositional analysis of the SHP1/ppERK signalling model. Particularly, the clarification that ppERK is not involved in positive feedback; the influence of the negative feedback base state; and the hypothesis of T cell activation.

Chapter 4 The development of the lattice of generalised receptors, and consequently all theoretical properties of a generalised receptor. The definition and theoretical properties of the decay density estimate. The definition and conditions for anomaly of the Receptor Density Algorithm.

Chapter 5 the application of the receptor density algorithm to the detection of anomalies in spectrometer data. The details of signature anomaly matching.

CHAPTER 2

Modelling Methods

The model analysed in Chapter 3 will be defined in terms of chemical reaction equations. This chapter gives a brief overview of the associated tools that are required in Chapter 3.

The chapter is organised as follows: Section 2.1 describes approaches to modelling chemical kinetics, the focus is on stochastic chemical kinetics and abstractions from this stochastic picture. Section 2.2 discusses useful computational tools associated with stochastic analysis of chemical reaction equations. Finally, describes a notation for chemical reaction equations which will prove useful in Chapter 3.

2.1 Chemical Kinetics

Chapter 3 analyses a biological model originally presented in [Altan-Bonnet and Germain, 2005]. The model employs chemical kinetics and is described by a system chemical reaction equations which contain of the order of 10^2 molecular species that exist in copy numbers ranging from 10^3 to 10^6 within the closed volume of a cell. The model was originally analysed in [Altan-Bonnet and Germain, 2005] by means of ordinary differential equations. The objective of this section is to overview the concepts associated with chemical kinetics and to reveal the assumptions associated with ordinary differential equation analysis of chemical kinetics.

Consider a volume $\Omega^\dagger$ that contains N molecular species: $S_0, S_1, \ldots, S_{N-1}$, which can interact via M chemical reaction channels: $R_0, R_1, \ldots, R_{M-1}$. At the most fundamental level the interactions between the molecules in Ω are described by quantum mechanics. Calculating the paths of molecules at this level is computationally intractable at the scale of the model analysed in the next chapter. Any model which considers the trajectories of individual molecules is referred to as a *microscopic* model, and all are intractable at the scale of interest here.

It is possible to reduce the computational burden whilst retaining information on the kinetics of the N chemical species. If it is valid to assume that Ω is in thermal equilibrium then the volume is said to be well-mixed and the spatial information of molecules may be

†In the model in this thesis the volume Ω will be a single cell.

discarded [1] and only the discrete quantities of each molecular species in Ω tracked. This is known as the *mesoscopic* abstraction, and due to the incomplete knowledge of the system the reactions proceed stochastically. The assumption of thermal equilibrium is justified in the model considered in the next chapter, and all subsequent discussions are restricted to those of a well-mixed volume Ω.

Denote $X_i(t) \geq 0$ as the non-negative integer quantity of species S_i at time t, then $\mathbf{X}(t) \equiv (X_0(t), \ldots, X_{N-1}(t))$ is the state of Ω at time t. Each reaction R_j is defined by two quantities, first, a state change vector $\boldsymbol{\nu}_j \equiv (\nu_{0j}, \nu_{1j}, \ldots, \nu_{(N-1)j})$, such that ν_{ij} is the change in the population of species S_i as a result of the application of reaction R_j. If $\mathbf{X}(t) = \mathbf{x}$ and reaction R_j occurs then the system will be in state $\mathbf{x} + \boldsymbol{\nu}_j$. The second defining detail of a reaction R_j is a propensity function a_j:

$$a_j(\mathbf{x})dt = \text{the probability that given } \mathbf{X}(t) = \mathbf{x} \text{ that reaction } R_j \text{ will occur somewhere in } \Omega \text{ in the next time interval } [t, t + dt). \tag{2.1}$$

The above is the key concept in the formulation of stochastic reaction kinetics, and through direct consequences of the above it is possible to derive a hierarchy of techniques that may be used to analyse chemical reaction kinetics. It is first useful to connect the concepts of R_j, a_j and $\boldsymbol{\nu}_j$ to the more familiar chemical reactions equations.

2.1.1 Chemical Reaction Equations

Every reaction R_j can be described by a chemical reaction equation of the general form:

$$\alpha_{1j}S_1 + \ldots + \alpha_{Nj}S_N \xrightarrow{r_j} \beta_{1j}S_1 + \ldots + \beta_{Nj}S_N, \tag{2.2}$$

the constants α_{ij} and β_{ij} define the stoichiometry $\nu_{ij} = \alpha_{ij} - \beta_{ij}$. The *reactants* of a reaction R_j are the set of S_i with non-zero α_{ij} and the *products* of R_j are the set of S_i with non-zero β_{ij}. The constant r_j is known as the *reaction rate constant* and combined with the number of ways of arranging the reactants, the constant defines the propensity function a_j. If the state of the system is $\mathbf{X}(t) = \mathbf{x}$ then,

$$a_j(\mathbf{x}) = r_j \binom{X_1}{\alpha_{1j}} \binom{X_2}{\alpha_{2j}} \cdots \binom{X_{N-1}}{\alpha_{(N-1)j}} \tag{2.3}$$

The justification for the existence of a reaction rate constant is given in [Gillespie, 1992] for two types of reaction: *unimolecular* and *bimolecular*. Unimolecular reactions have a single reactant and bimolecular reactions have two reactants. In general higher order reactions do not occur [Gillespie, 1992], for example, in a trimolecular reaction the probability of three reactants colliding at the same position with the appropriate orientation to react is so small that these reactions may be ignored. Moreover, any higher order reaction may be modelled

[1] If it is invalid to consider the entirety of Ω to be well mixed it is often possible to divide Ω into disjoint subvolumes Ω_i and use a *reaction-diffusion* modelling scheme [Stundzia and Lumsden, 1996; Andrews et al., 2009]. Molecules may react (in the manner outlined in this chapter) within a sub-volume or they may diffuse to a neighbouring sub-volume.

by a sequence of lower order reactions. Specific examples of unimolecular and bimolecular reactions are:

$$\text{Unimolecular: } S_i \xrightarrow{r_j} S_q, \qquad a_j(\mathbf{x}) = r_j X_i(t) \tag{2.4}$$

$$\text{Bimolecular: } S_i + S_k \xrightarrow{r_j} S_q, \quad a_j(\mathbf{x}) = r_j X_i(t) X_k(t), \qquad i \neq k. \tag{2.5}$$

$$\text{Bimolecular Symmetric: } S_i + S_i \xrightarrow{r_j} S_q, \quad a_j(\mathbf{x}) = r_j \tfrac{1}{2} X_i(t)(X_i(t) - 1). \tag{2.6}$$

2.1.2 Stochastic Kinetics

This section returns to the stochastic formulation of chemical kinetics and considers two direct consequences. The first is the *Chemical Master Equation (CME)* [Gillespie, 1992; Kampen, 2007]. The CME is a set of differential equations describing the evolution of $\mathbb{P}(\mathbf{x}, t|\mathbf{x_0}, t_0)$ which is the probability that Ω is in state $\mathbf{X}(t) = \mathbf{x}$ given the initial state $\mathbf{X}(t_0) = \mathbf{x_0}$. The CME it is defined:

$$\frac{\partial \mathbb{P}(\mathbf{x}, t|\mathbf{x}_0, t_0)}{\partial t} = \sum_{j=0}^{M-1} \left[a_j(\mathbf{x} - \boldsymbol{\nu}_j)\mathbb{P}(\mathbf{x} - \boldsymbol{\nu}_j|\mathbf{x}_0, t_0) - a_j(\mathbf{x})\mathbb{P}(\mathbf{x}, t|\mathbf{x}_0, t_0) \right] \tag{2.7}$$

Note that the CME has an equation for every state in the state space of $\mathbf{X}(t)$, and the size of this state space is exponential in N the number of molecular species. Thus, the CME is only analytically solvable in a small number of cases. It is possible to obtain numerical solutions for appropriately sized and structured state spaces; the details of this are discussed in Section 2.2.

A second consequence of the stochastic formulation of chemical kinetics is the existence of a Monte Carlo simulation algorithm known as the Gillespie Algorithm [Gillespie, 1977]. It is concerned with generating sample trajectories of molecular populations from an initial state. Consequently, the following distribution is of importance:

$$\begin{aligned} p(\tau, j|\mathbf{x}, t) = {} & \text{the probability that given } \mathbf{X}(t) = \mathbf{x} \text{ the next reaction} \\ & \text{in the system will occur at time } [t + \tau, t + \tau + d\tau) \\ & \text{and will be reaction } R_j. \end{aligned} \tag{2.8}$$

Then, $p(\tau, j|\mathbf{x}, t)$ is given by,

$$p(\tau, j|\mathbf{x}, t) = a_j(\mathbf{x}) \exp(-a_0(\mathbf{x})\tau), \qquad a_0(\mathbf{x}) = \sum_{j=0}^{M-1} a_j(\mathbf{x}) \tag{2.9}$$

There are many variants of the Gillespie Algorithm [Gibson and Bruck, 2000; Cao et al., 2004; Li and Petzold, 2006; McCollum et al., 2006; Phillips and Cardelli, 2007], all are concerned with efficient sampling of (2.9). The specifics of the *direct method* (a well known fundamental Gillespie algorithm variant) are given as they are illustrative of the concepts necessary to sample (2.9). Given that Ω is in state $\mathbf{x}$ at time t and η_1, η_2 are uniform random

numbers drawn from $[0, 1]$, then calculate:

$$\tau = [1/a_0(\mathbf{x})] \ln(1/\eta_1), \tag{2.10}$$

$$j = \text{smallest integer satisfying } \sum_{k=0}^{j} a_k(\mathbf{x}) > a_0(\mathbf{x})\eta_2. \tag{2.11}$$

Then, the simulation time is updated $t \to t + \tau$ and the simulation state is updated $\mathbf{x} \to \mathbf{x} + \boldsymbol{\nu}_j$. All Gillespie Algorithm simulations in this thesis are performed with the Stochastic Pi Machine (SPiM) [Phillips and Cardelli, 2007] which implements a variant of the direct method.

2.1.3 Approximations to Stochastic Kinetics

Through approximations to the stochastic description of chemical kinetics it is possible to derive other well-known modelling abstractions. In particular, the *macroscopic* modelling abstraction can be derived. This is the modelling abstraction which makes use of ordinary differential equations and is used by [Altan-Bonnet and Germain, 2005] in the original presentation of the model analysed in the next chapter. Understanding the derivation of the macroscopic abstraction reveals the assumptions that should hold for the results of macroscopic analysis to be valid.

The following arguments are recounted from [Gillespie, 2000]. Consider the random variable:

$$\begin{aligned} Q_j(\mathbf{X}(t), \tau) = &\text{ the number of occurrences of reaction } R_j \text{ in the} \\ &\text{time interval } [t, t + \tau] \text{ for } \tau > 0. \end{aligned} \tag{2.12}$$

Thus, the number of molecules of species S_i at time $t + \tau$ is given by:

$$X_i(t + \tau) = X_i(t) + \sum_{j=0}^{M-1} Q_j(\mathbf{X}(t), \tau)\nu_{ij}. \tag{2.13}$$

Calculating the exact details of $Q_j(\mathbf{X}(t), \tau)$ is as difficult as solving the CME (2.7). However, if two conditions are satisfied a good simple approximation to $Q_j(\mathbf{X}(t), \tau)$ can be obtained [Gillespie, 2000]. The first condition is as follows:

Condition 2.1. *Require that τ is sufficiently small such that the propensity functions do not change "appreciably" during $[t, t + \tau]$. That is, for all reactions R_j:*

$$a_j(\mathbf{X}(t')) \approx a_j(\mathbf{X}(t)), \qquad \forall t' \in [t, t + \tau]. \tag{2.14}$$

Since only unimolecular and bimolecular are considered the application of a single reaction will only change molecular populations by 1 or 2. Thus, if the sizes of all populations are large compared to 2 then Condition 2.1 can be satisfied. Assuming that Condition 2.1 can be satisfied then all reactions occurring in the interval $[t, t + \tau]$ are essentially independent

of each other. So each $Q_j(\mathbf{X}, \tau)$ is a statistically independent Poisson random variable $\mathcal{P}(a_j(\mathbf{X}(t), \tau)$ [†]. So, Condition 2.1 allows (2.13) to be approximated by:

$$X_i(t+\tau) = X_i(t) + \sum_{j=0}^{M-1} \mathcal{P}(a_j(\mathbf{X}(t), \tau)\nu_{ij}. \tag{2.15}$$

The second conditions is as follows:

Condition 2.2. *Require τ to be large enough that the expected number of occurrences of each reaction in the interval $[t, t+\tau]$ is much greater than 1. For all reactions R_j:*

$$\mathbb{E}[\mathcal{P}(a_j(\mathbf{X}(t)), \tau)] = a_j(\mathbf{X}(t))\tau \gg 1. \tag{2.16}$$

Condition 2.2 looks to be in difference to Condition 2.1, certainly it may be impossible to satisfy both conditions simultaneously and in that case the following approximations are invalid. However, if the molecular populations are sufficiently large then Condition 2.2 can hold for small τ. Condition 2.2 allows the Poisson random variable $\mathcal{P}(a_j(\mathbf{X}(t)), \tau)$ to be approximated by a normal random variable $N(a_j(\mathbf{X}(t))\tau, \sqrt{a_j(\mathbf{X}(t))\tau})$ [‡]. Note that a Poisson random variable is discrete and a normal random variable is continuous, thus this approximation converts the molecular populations to continuous real quantities. For clarity, the vector of continuous molecular populations is written $\mathbf{W}(t)$, so species S_i has continuous quantity $W_i(t)$ at time t. Then the update equation (2.15) can be approximated by,

$$W_i(t+\tau) = W_i(t) + \sum_{j=0}^{M-1} N(a_j(\mathbf{X}(t))\tau, [a_j(\mathbf{X}(t))\tau]^{\frac{1}{2}})\nu_{ij}. \tag{2.17}$$

Note that since $N(\mu, \sigma) = \mu + \sigma N(0, 1)$ the above becomes:

$$W_i(t+\tau) = W_i(t) + \sum_{j=0}^{M-1} a_j(\mathbf{W}(t))\tau\nu_{ij} + \sum_{j=0}^{M-1} [a_j(\mathbf{W}(t))\tau]^{\frac{1}{2}} N_j(0,1)\nu_{ij}, \tag{2.18}$$

where the each standard normal is denoted $N_j(0, 1)$ so that it is explicit that there are M statistically independent normal random variables. Observe that the expression (2.18) is composed of both deterministic components: $a_j(\mathbf{X}(t))\tau\nu_{ij}$ and fluctuating components: $[a_j(\mathbf{X}(t))\tau]^{\frac{1}{2}} N_j(0, 1)\nu_{ij}$. As Conditions 2.1 and 2.2 apply one can regard τ as a infinitesimal dt and, as argued in [Gillespie, 2000], the following differential formulation implied:

$$\frac{dW_i(t)}{dt} = \sum_{j=0}^{M-1} a_j(\mathbf{W}(t))\nu_{ij} + \sum_{j=0}^{M-1} [a_j(\mathbf{W}(t))]^{\frac{1}{2}} \Gamma_j(t)\nu_{ij}, \tag{2.19}$$

[†]The number of independent events n that occur at rate λ over time t has a Poisson distribution $\mathcal{P}(\lambda, t)$ with probability mass function $p_{\mathcal{P}}(n, \lambda, t) = e^{-\lambda t}(\lambda t)^n/n!$, expected value $\mathbb{E}[\mathcal{P}(\lambda, t)] = \lambda t$ and variance var $[\mathcal{P}(\lambda, t)] = \lambda t$.

[‡]A normal random variable $N(\mu, \sigma)$ has expectation μ and variance σ^2 and probability density function $p_N(x, \mu, \sigma) = (\sigma\sqrt{2\pi})^{-1} \exp(-(x-\mu)^2/(2\sigma^2))$.

where $\Gamma_j(t)$ are temporally uncorrelated, statistically independent Gaussian white noises ($\Gamma_j \equiv \lim_{dt\to 0} N(0, 1/dt)$). The expression (2.19) is known as the *Chemical Langevin Equation (CLE)* and is a commonly used modelling abstraction. Note that the ratio of the size of the fluctuating component of (2.19) to the size of deterministic component is $[a_j(\mathbf{W}(t))]^{-1/2}$. The size of the relative fluctuations is proportional to the inverse of the square root of the reactant population sizes. If one considers the *thermodynamic limit* in which both Ω and the number of molecules in Ω go to ∞ such that the concentrations of each molecular species remains constant, then the size of relative fluctuations becomes *vanishingly small*. The fluctuations can be discarded and one derives the *Reaction Rate Equation (RRE)* the system of ordinary differential equations that constitutes the commonly used *macroscopic* abstraction. This is usually written in terms of a concentration variable $\mathbf{Z}(t) = \mathbf{W}(t)/\Omega$ and requires altered propensity functions $\hat{a}_j(\mathbf{W}(t))$:

$$\hat{a}_j(\mathbf{W}(t)) = \begin{cases} W_i(t)^2/\Omega & \text{for symmetric bimolecular reactions} \\ a_j(\mathbf{W}(t))/\Omega & \text{otherwise.} \end{cases} \tag{2.20}$$

Then the RRE is given:

$$\frac{dZ_i(t)}{dt} = \sum_{j=0}^{M-1} \hat{a}_j(\mathbf{Z}(t))\nu_{ij}. \tag{2.21}$$

The advantage of this deterministic formulation is the ease with which it may be solved and the powerful analysis tools associated with ordinary differential equations. The disadvantage is that at least Conditions 2.1 and 2.2 must hold for the RRE to be valid. Many biological systems, including the one analysed in Chapter 3, have molecules present in small quantities such that the CLE and the RRE are not valid.

It is interesting to note how results are affected if the RRE is applied in a situation where Conditions 2.1 and 2.2 do not hold. Models closely related to the one considered in Chapter 3 are compared at a mesoscopic and a macroscopic abstractions in [Artyomov et al., 2007; Lipniacki et al., 2008]. Both [Artyomov et al., 2007; Lipniacki et al., 2008] note that bistabilities can exist in the stochastic model which do not exist in the deterministic model. Sufficiency conditions for this discrepancy are given in [Artyomov et al., 2007] in terms of the structure of the chemical reaction network and the size of molecular populations. These sufficiency conditions are satisfied by some of the reactions in the model in Chapter 3.

The model in Chapter 3 does not always fulfil Conditions 2.1 and 2.2, and the sufficiency conditions for a stochastic bistability described in [Artyomov et al., 2007] are satisfied for certain reactions in Chapter 3. As a consequence the analysis in Chapter 3 will be stochastic at the mesoscopic abstraction via the CME and the Gillespie Algorithm.

2.2 Analysis with the CME and Gillespie Algorithm

This section describes useful computational tools associated with the CME and Gillespie Algorithm. The focus is on the CME, and particularly that the CME is a equivalent form of the Chapman-Kolmogorov equation for Markov Processes [Kampen, 2007]. Thus, the CME

defines a continuous time Markov Chain (CTMC). Notation conventions are now introduced to aid the descriptions of CTMCs. First, let state vector $\mathbf{X}(t)$ take any state in state space $\mathcal{X}$. Second, $\mathbf{x}_i \in \mathcal{X}$ is used to denote the state that occurs at time t_i, so $\mathbf{X}(t_i) = \mathbf{x}_i$. Finally, $\mathbf{y}_i \in \mathcal{X}$ is used to index over all states, so $\mathcal{X} = \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{|\mathcal{X}-1|}\}$. Then a CTMC is defined as follows:

Definition 2.1. *A continuous time Markov chain is a continuous time stochastic process* $\{\mathbf{X}(t) : t \geq 0\}$ *with state space* $\mathcal{X}$ *that satisfies the Markov property:*

$$\mathbb{P}(\mathbf{X}(t_n) = \mathbf{x}_n | \mathbf{X}(t_{n-1}) = \mathbf{x}_{n-1}, \ldots, \mathbf{X}(t_1) = \mathbf{x}_1, \mathbf{X}(t_0) = \mathbf{x}_0) = \mathbb{P}(\mathbf{X}(t_n) = \mathbf{x}_n | \mathbf{X}(t_{n-1}) = \mathbf{x}_{n-1}) \tag{2.22}$$

Where $0 \leq t_0 \leq t_1 \leq \cdots \leq t_{n-1} \leq t_n$, *and* $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{x}_n, \in \mathcal{X}$.

The CTMC considers the jumps between states in $\mathcal{X}$ defined by the chemical reactions, and so one can consider the *transition probabilities*:

$$p_{ij}(t) = \mathbb{P}(\mathbf{X}(t) = \mathbf{y}_j | \mathbf{X}(s) = \mathbf{y}_i) \qquad \text{for } 0 \leq s \leq t \text{ and } i, j \in \mathcal{X}. \tag{2.23}$$

If $\mathbf{P}(t)$ is the matrix with entries $p_{ij}(t)$, then $\mathbf{P}(t)$ is described by the following differential equation:

$$\dot{\mathbf{P}}(t) = \mathbf{G}\mathbf{P}(t), \tag{2.24}$$

$\mathbf{G}$ is known as the *infinitesimal generator* matrix or *rate* matrix and is defined by the propensity functions of the chemical reactions:

$$g_{ij} = \begin{cases} \displaystyle\sum_{k=0}^{M-1} \delta_{\mathbf{y}_i + \boldsymbol{\nu}_k, \mathbf{y}_j} a_k(\mathbf{y}_i) & \text{for } i \neq j, \\ \displaystyle -\sum_{k=0,\ k \neq i}^{|\mathbf{X}|-1} g_{ik} & \text{for } i = j, \end{cases} \tag{2.25}$$

the delta is a multi-dimensional Kronecker delta. Each g_{ij} for $i \neq j$ describes the exponential rate of transition from state $\mathbf{y}_i$ to $\mathbf{y}_j$, and g_{ii} can be thought of as the rate of leaving state $\mathbf{y}_i$, and note that:

$$\sum_j g_{ij} = 0 \qquad \forall i. \tag{2.26}$$

Using this framework three properties of CTMCs/CMEs are calculated in this thesis: first, the time evolution of $\mathbf{X}(t)$ from a initial state (*transient analysis*); second, the long term behaviour and equilibrium states (the *stationary distribution*); third, the expected time to reach a set of states $\mathcal{C} \subseteq \mathcal{X}$ from a state $\mathbf{y_h} \in \mathcal{X}$ (the *hitting time*). The calculation of each of the three properties is now discussed in turn.

The transient distribution is given by the general solution to (2.24), if $\mathbf{p}(t)$ is the row vector with entries $p_i = \mathbb{P}(\mathbf{X}(t) = \mathbf{y}_i)$ then the solution to (2.24) from an initial distribution $\mathbf{p}(0)$ is given by the matrix exponential:

$$\mathbf{p}(t) = \mathbf{p}(0)\mathrm{e}^{t\mathbf{G}} = \mathbf{p}(0) \sum_{n=0}^{\infty} \frac{(t\mathbf{G})^n}{n!}. \tag{2.27}$$

In this thesis the matrix exponential is calculated numerically using two methods. First, for small full (non-sparse) $\mathbf{G}$ MATLAB's `expm(G)` function is used, which performs a scaling and squaring algorithm with a Padé approximation as detailed in [Moler and Loan, 2003]. Second, for large sparse $\mathbf{G}$ methods provided in EXPOKIT [Sidje, 1998] are employed, they use Krylov subspace projection techniques to calculate $\mathbf{p}(0)\mathrm{e}^{t\mathbf{G}}$ directly (i.e. $\mathrm{e}^{t\mathbf{G}}$ is not explicitly calculated) for an initial distribution $\mathbf{p}(0)$.

The long term behaviour of a CTMC is dictated by its stationary or invariant distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots \pi_{|\mathcal{X}|})$ which exists if the CTMC is irreducible and recurrent [Norris, 1998]. This is the the case for the all the CTMCs considered in this thesis. The stationary distribution may be found by solving the following:

$$\boldsymbol{\pi}\mathbf{G} = \mathbf{0}, \qquad \text{with the condition } \sum_i \pi_i = 1. \tag{2.28}$$

A numerical solution to the stationary distribution equation can often be difficult to find [Philippe et al., 1996], particularly when $\mathbf{G}$ is large and sparse, which is often the case in Chapter 3. The inverse iteration method [Philippe et al., 1996] is used, rather than solve (2.28) directly, solve and scale:

$$\mathbf{G}^T\boldsymbol{\Psi} = \boldsymbol{\Psi}_1, \qquad \text{then scale } \boldsymbol{\pi}^T = \frac{\boldsymbol{\Psi}}{\mathbf{1}_{|\mathcal{X}|}^T\boldsymbol{\Psi}}, \tag{2.29}$$

$\boldsymbol{\Psi}_1$ is the the column vector length $|\mathcal{X}|$ which has all entries zero except for 1 in the $|\mathcal{X}|$th position; and $\mathbf{1}_n$ column vector length n with all ones.

Finally, the expected time to reach a set of states $\mathcal{C} \subseteq \mathcal{X}$ can be found by calculating the vector of expected first hit times $\boldsymbol{h}^{\mathcal{C}} = (h_i^{\mathcal{C}} : i \in \mathcal{X})$, with $h_i^{\mathcal{C}}$ the expected time to reach a state $\mathbf{c} \in \mathcal{C}$ starting from state $\mathbf{y}_i$. The vector of hit times can be found by solving the following [Norris, 1998]:

$$\begin{aligned} h_i^{\mathcal{C}} &= 0 && \text{for } \mathbf{y}_i \in \mathcal{C} \\ -\sum_j g_{ij} h_k^{\mathcal{C}} &= 1 && \text{for } \mathbf{y}_i \notin \mathcal{C}. \end{aligned} \tag{2.30}$$

2.2.1 A Volume Scaling

The following technique will feature in Chapter 3. The model in Chapter 3 will involve receptors on the surface of a cell. A receptor has an internal and external component, and the receptor diffuses at a rate slower than molecules which can move freely inside the cell. Additionally, the model in the next chapter comes supplied with molecular concentrations and deterministic macroscopic rates, it is necessary to convert these to discrete molecule numbers and stochastic mesoscopic rates. The rate conversion applies to bimolecular reactions and amounts to a change of scale and units (from $\mathrm{mol}^{-1}s^{-1}$ to s^{-1}).

The molecule numbers of the entire model in Chapter 3 are only just within range of the Gillespie algorithm. The following applies a scaling to reduce the state space down to the concern of a small number of molecules near a single cell-surface receptor. A fraction

$\epsilon \in (0, 1]$ is applied to the volume Ω of a cell. A macroscopic rate d_j is converted to a mesoscopic rate r_j for reaction R_j with respect to scaling ϵ by the following:

$$r_j = \frac{d_j}{N_A \Omega \epsilon}, \tag{2.31}$$

$N_A = 6.022 \times 10^{23}$ is the Avogadro Number, and allows simulation of exact molecule numbers. If a molecule has an initial concentration $c \in \mathbb{R}$ then the initial discrete number of molecules $N \in \mathbb{Z}$ is taken to be $N = \lfloor cN_A V\epsilon \rfloor$. Three different choices of ϵ are used in Chapter 3, and justification for these choices is given in Section 3.11.1.

It is noted that the results obtained with an ϵ scaling are not guaranteed to be valid. However, they may provide intuition, or the correct qualitative picture. This must be confirmed at a $\epsilon = 1$ level. Given the arguments in Chapter 1, small ϵ would fail to reproduce any emergent properties present at the $\epsilon = 1$ level. Often in emergent systems the basic local rules are known but the global behaviour remains to be understood. In the Chapter 3, the basic local rules of a this cell-surface receptor are not known. The ϵ scaling offers a method of addressing this issue.

2.3 Reaction Equation Notation

In Chapter 3 there are many chemical reactions which apply to different species but share a common structure. To describe reaction equations more succinctly set notation is used to group together common reactions. A reaction applied to a set represents application of the reaction to every member of the set, so that,

$$\{\mathrm{A},\, \mathrm{B}\} + \mathrm{C} \xrightarrow{r} \{\mathrm{A},\, \mathrm{B}\}\mathrm{C}, \tag{2.32}$$

represents the reactions:

$$\mathrm{A} + \mathrm{C} \xrightarrow{r} \mathrm{AC} \quad \text{and} \quad \mathrm{B} + \mathrm{C} \xrightarrow{r} \mathrm{BC}\ . \tag{2.33}$$

Concatenation of terms represents a complex containing the elements of the concatenation. Generally any set manipulation can be used to reduce the reaction description:

$$\mathcal{S} = (\{\mathrm{A}\} \times \{\emptyset,\, \mathrm{B},\, \mathrm{C}\} \times \{\mathrm{D}\}) \cup \{\mathrm{GH},\, \mathrm{GJ}.\} \tag{2.34}$$

Represents:

$$\mathcal{S} = \{\mathrm{AD},\, \mathrm{ABD},\, \mathrm{ACD},\, \mathrm{GH},\, \mathrm{GJ}\}. \tag{2.35}$$

Then reactions may be described using $\mathcal{S}$, for example the binding of F may occur for all $i \in \mathcal{S}$:

$$\mathcal{S}{+}\mathrm{F} \xrightarrow{r} \mathcal{S}\mathrm{F}. \tag{2.36}$$

Components may be supplied with indices:

$$\mathrm{A}_n + \mathrm{B} \xrightarrow{r} \mathrm{A}_n\mathrm{B}, \quad n = 1, 2, 3 \tag{2.37}$$

which represents:

$$\begin{aligned} \mathrm{A}_1 + \mathrm{B} &\xrightarrow{r} \mathrm{A}_1\mathrm{B}, \\ \mathrm{A}_2 + \mathrm{B} &\xrightarrow{r} \mathrm{A}_2\mathrm{B}, \\ \mathrm{A}_3 + \mathrm{B} &\xrightarrow{r} \mathrm{A}_3\mathrm{B}. \end{aligned} \tag{2.38}$$

Algebraic manipulations of indices apply to all values that the index is defined, for example:

$$\begin{aligned} \mathrm{A}_n,\ n = 1,2,3 &\implies \mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \\ \mathrm{A}_{n+1},\ n = 1,2,3 &\implies \mathrm{A}_2, \mathrm{A}_3, \mathrm{A}_4. \end{aligned}$$

This notation can increase the possibility for ambiguity of determining reactants and products. Any ambiguities can be resolved by conservation, i.e. the same numbers of molecules on both sides of the equation. In all reactions if a component that may be associated to a complex and that component is not explicitly given in the reaction, then it is assumed that the reaction may proceed with and without the component associated. For example if B and D may be bound to A, then the reaction:

$$\mathrm{A} + \mathrm{C} \xrightarrow{r} \mathrm{AC}, \tag{2.39}$$

includes all the following reactions:

$$\mathrm{A} + \mathrm{C} \xrightarrow{r} \mathrm{AC}, \qquad \text{neither B nor D bound,} \tag{2.40}$$

$$\mathrm{AB} + \mathrm{C} \xrightarrow{r} \mathrm{ABC}, \qquad \text{B bound,} \tag{2.41}$$

$$\mathrm{AD} + \mathrm{C} \xrightarrow{r} \mathrm{ADC}, \qquad \text{D bound,} \tag{2.42}$$

$$\mathrm{ABD} + \mathrm{C} \xrightarrow{r} \mathrm{ABDC}, \qquad \text{B and D bound.} \tag{2.43}$$

Finally, the next chapter discusses molecules undergoing *phosphorylation*, the details of this will be given in Chapter 3. For the moment, if A has p phosphorylations it is represented A^p. If no phosphorylation state is given, i.e. A, this represents that the phosphorylation state of A has no affect on the reaction.

CHAPTER 3

T Cell Signalling Models

This chapter describes the modelling work of this thesis. In particular, it describes stochastic analysis of a model originally presented in [Altan-Bonnet and Germain, 2005]. The chapter commences with a background to immunology. This is followed by a description of T cell signalling, after which, the model will then be presented and analysed.

The immunology discussed in this chapter is that of the vertebrates. Their immune system is more complicated than their non-vertebrate or plant counterparts, and has a rich complexity that sparks dispute between immunologists as to the immune system's purpose and function. Certainly, a purpose of the immune system is to protect its *host* and remove invading *pathogens* such as bacteria and viruses. The disputes often regard the general concepts of this protection, and moreover, whether this protection is the limit of immune function [Cohen, 2000]. However, all would agree that the immune system must *recognise* (certain states of the body, or molecular patterns) and then *respond* based on the recognition. The model in this chapter regards the specifics of how one type of immune recognition occurs. The complexities of the immune system prevent all but a small subset of immunology that is relevant to the model from being presented here. A fuller introduction to the mechanical components of the immune system can be found in a textbook such as [Murphy et al., 2008]. The reference is given with the caveat that immunological textbooks often contain gross simplifications, but are nonetheless a good starting point. It should be stated that although the immunological background is a simplified version of the true picture, a background is given to understand the subsequent modelling work.

3.1 The Mechanisms of the Immune System

The vertebrate immune system can broadly be split into two parts: the *innate* and the *adaptive*. The innate immune system is passed through the *germ-line*, that is through the genes of individuals. The innate is (relatively) constant during the lifetime of an individual and adapts on an evolutionary time-scale. The blue-print of the adaptive immune system is also passed on genetically, but the adaptive immune system has the ability to alter restricted regions of its genome through genetic rearrangements. The result is the adaptive immune system is able to generate proteins that were not encoded in an individual's genome at

birth. The ability to perform these genetic rearrangements is necessarily passed through the genome but the specific rearrangements performed by an individual during its lifetime are not passed on.

The innate immune system acts as a first line of defence for the host, communicates information to the adaptive immune system, and performs certain maintenance tasks. The innate response is fast (in comparison to the adaptive response), and the majority of its defence abilities rely on detecting common molecular signatures of pathogenic material. The molecular signatures detectable by the innate immune system are those that evolution has found to be reliable indicators of foreign material. The reaction of the innate immune system can be severe, many examples of *allergy* are situations where the innate immune system has inappropriately responded. The communication of information to the adaptive immune system is the result of cells in the innate immune system, namely *antigen presenting cells* (APCs), sampling molecular material from across the host and then presenting this molecular material to the adaptive immune system through interactions between APCs and cells of the adaptive immune system.

The adaptive immune system allows an organism to be protected against pathogens to which the innate immune system is blind. This protection is performed via a class of cells known as the *lymphocytes*, which are able to recognise molecular patterns via receptors on their surfaces. The lymphocytes have the ability to dictate the adaptive immune response, for example some lymphocytes can kill cells and others can activate cells of the innate and the adaptive. It is the lymphocytes to which the APCs display their collected molecular material. This interaction between lymphocytes and antigen presenting cells is a key mechanism of the adaptive immune response. A more detailed description of the behaviour of the antigen presenting cells and the lymphocytes is now given.

3.1.1 Antigen Presenting Cells

Every cell which contain a *nucleus* is an antigen presenting cell (APC). The term *antigen* was first used for any process that could invoke *anti*body *gen*eration, more generally it is a molecule that is recognised by the immune system. The antigen presenting cells take antigenic material and process it into *peptides* which are short sequences of amino acids. The peptides are then loaded in *major histocompatability complex* (MHC) molecules which are then displayed on the surface of the APC. The peptide-MHC complex will be referred to as pMHC. There are two classes of MHC: *Class I* and *Class II*.

There are two types of APC: *professional* and *non-professional*. Their differences are in the origin of their displayed peptides. Professional APCs *phagocytose* molecules and certain cells, that is, they engulf and break-down these molecules and cells. The peptides derived from the phagocytosed proteins are then presented on the surface of the APC by MHC class II. The non-professional APCs derive their peptides for presentation from protein present in their *cytosol* (that is, present inside the cell). The non-professional APCs display peptides by MHC class I. In summary, professional APCs display a sample of proteins present locally to that APC, that is, they present a sample of the environmental state. The non-professional

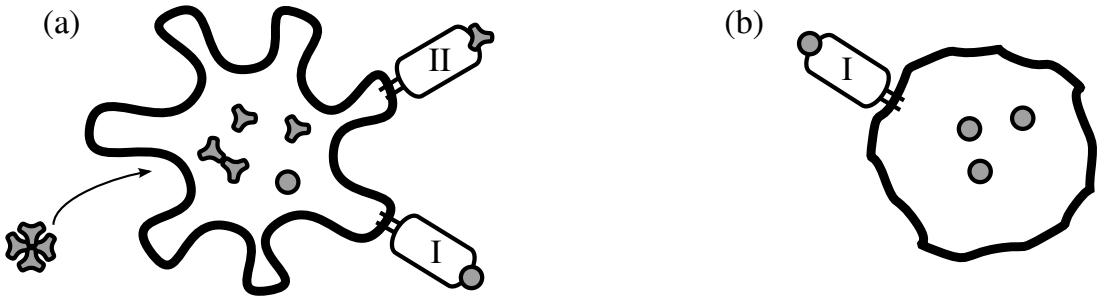

Figure 3.1: *Left: (a)* A professional APC, displaying class I and class II pMHC on its surface. Note that fragments of the extra-cellular proteins are displayed by pMHC class II. *Right: (b)* A non-professional APC, displaying class I pMHC on its surface presenting fragments of a sample of its internal proteins.

APCs, then, present a sample of their internal state. The professional APCs also present a sample of their internal state via MHC class I, in this sense they are both professional and non-professional. The displaying of pMHC is shown in Figure 3.1.1, note the quantities and proportionals in this figure are cartoon-like. In reality a single pMHC molecule is many orders of magnitude smaller than the cell. There are of the order of 10^4 pMHC molecules on the surface of an APC.

As stated earlier, cells which do not contain a nucleus, for example red blood cells, do not display MHC. The nucleus of a cell contains the cell's DNA and a number of proteins relevant to gene expression and cell replication. Thus, cells without a nucleus are unable to replicate. It is interesting to speculate that a reason why non-nucleated cells do not require this direct connection to the immune system is because they do not have the powerful expression and replication protein machinery of a nucleated cell. This protein machinery is often hijacked by viruses. Consequently, infected non-nucleated cells pose less of a threat to the body, and perhaps this is a reason they do not require such close immune vetting.

3.1.2 Lymphocytes

There are two major classes of lymphocytes: the T lymphocyte (*T cell*) and the B lymphocyte (*B cell*). The origin of both cells is in the host's bone marrow, but it is the location of maturement that names the cells. The B cells remain in *B*one marrow to mature; the T cells migrate to the *T*hymus to mature. The B cells are responsible for *antibody* production, and undergo the evolutionary-like *clonal selection principle* [Burnet, 1959] to generate high quality antibodies. Since B cells play no part in the modelling in this chapter their discussion ends here.

On the surface of a T cell are of the order of 10^4 *T cell receptors* (TCRs), every TCR on the surface of a T cell is identical in conformation, and in general every T cell has a different TCR. Early in the development of a T cell the cells encoding the TCR are rearranged, the result is every T cell has a TCR with a effectively random conformation [1]. The TCRs bind

[1]For the discussions here it suffices to say that every TCR has a random conformation, counting the possible gene rearrangements without considering issues of protein folding there are approximately 10^{18}

to pMHC molecules on the surface of an APC. Depending on the strength of each TCR-pMHC bond or the *affinity* between each TCR and pMHC that bind in the population, the T cell may become activated. The affinity between a TCR and pMHC is dependent on the conformation of the TCR, the peptide and MHC. Its complexities are well abstracted to a single parameter: its average lifetime [Feinerman et al., 2008a]. If a T cell becomes activated it is said to have recognised the subset of the pMHC on the APC which were responsible for the T cell activation.

There are many classes of T cell, the two major classes being the *T helper* cell and the *cytotoxic* T cell. The T helper cell binds to pMHC class II on APC, if the T cell becomes activated it can then activate the APC. The meaning and implications of this APC activation relate to B cell activation [2] and are left out of this immunological review. The cytotoxic T cell binds class I pMHC on general APCs, if activated the cytotoxic T cell has the ability to kill its bound APC.

Clearly, the T cell has great power in the adaptive immune system and one of its major roles is in the defence against viruses. An APC infected with a virus will display viral peptides via the MHC on its surface, if cytotoxic T cell is able to recognise these viral peptides with its TCR then the T cell can kill the infected cell which contributes to halting the progression of the viral infection. A key topic in immunology is concerned with the generation of the T cell repertoire, that is that distribution of TCRs across all T cells. The repertoire must be such that it does not react with the host (i.e. no T cells recognise APCs displaying normal peptides) and that there is at least one T cell that is able to recognise any of the possible viral peptides. The fact that a working repertoire can be generated is remarkable, given that the details of possible viral peptides cannot be known in advance. Moreover, the conformation of a TCR is effectively randomly generated.

There are three mechanisms that allow a random repertoire to function appropriately. First, its size, as previously stated there are approximately 10^{18} different possible TCRs. Second, the maturation of T cells in the thymus includes a process known as *negative selection* or *clonal deletion* in which candidate T cells are tested against many APCs presenting combinations of normal peptides originating from the host. Any T cells which recognise these normal peptides are then deleted, resulting in a repertoire of T cells which under normal conditions is not activated by host peptides. The final mechanism lies in the specifics of the T cell recognition and activation process, which is the topic of the models here.

Before moving onto the specifics of the model it is worth making a brief mention of a few immunological theories which attempt to sum the behaviour of the immune system macroscopically; this is done in the next section.

3.2 Immunological Theories

There are many general theories of the immune system which attribute general processes to immune cells in an attempt to describe the governing principle of the immune system. The

different TCRs [Murphy et al., 2008]

[2] B cells, despite being part of the adaptive immune system are also professional antigen presenting cells

theories are often described as competing, however it seems highly unlikely (particularly given the points regarding degeneracy and evolution described in Chaper 1) that there is any one process or idea that "explains" immunology. Rather (as stated in [Cohen, 2000]), elements of all of these theories along with as yet undiscovered ideas contribute to explaining the immune response.

The most successful theory of immunology to date is that of *self-non-self discrimination*, which states that the immune system performs defence by discriminating between molecules derived from the host and molecules derived externally. Broadly, this theory requires that the negative selection of T cells is absolute, such that no T cell exists outside the thymus which interact with self molecules. Whilst, there is clear evidence that this is not the case [Germain and Stefanová, 1999; Vignali et al., 2008] (the conclusions of this chapter, Section 3.11, require some interaction with self) the self-non-self model has served immunology relatively well.

An alternative to the self-non-self model is *danger theory* [Matzinger, 2002], which proposes that professional APCs are able to detect damage in their local environment. Upon interaction with a lymphocyte, a professional APC can communicate an integrated signal of the recent damage via immune messenger molecules known as *cytokines*. The cytokines alter the probability of T cell activation, and so provide a *context* to the peptides presented by the APCs. An immune theory which has received recent popularity is that of the T regulatory cell [Vignali et al., 2008]. This is a class of T cell which specifically interacts with self in order to *regulate* an immune reaction.

The final theory is that of the *tuneable activation threshold* (TAT) hypothesis of lymphocytes [Grossman and Paul, 1992], this has some relevance to the work of the chapter and the following sub-section is devoted to its discussion.

3.2.1 The Tuneable Activation Threshold Hypothesis

First postulated by [Grossman and Paul, 1992], this states that T cells have an activation threshold that is dependent on their recent history of interactions. The T cell is said to be *excited* or *stimulated* by the interaction with an APC, the excitation level is then dependent on the affinity between all pMHC and the TCR. The activation threshold defines an excitation level which must be exceeded for the cell to become activated. This activation threshold is tuned by the recent interactions of the cell, such that continued interactions of comparable magnitude will not activate a cell. Consequently, T cells with a TCR with high affinity with normal host peptides ("self") will receive regular stimulation and so will have a high activation threshold. That is, a peptide with significantly higher affinity than the average host peptide is required to activate a cell. T cells with a low affinity for normal host peptides will have little stimulation and so have a low activation threshold. In [Grossman and Paul, 1992] a simple model of the TAT with sole purpose of explanation is presented. The model involves an excitation level $E(t)$ the stimulation received by the cell at time t; an excitation index $I(t)$ this is a time averaging of $E(t)$; and an activation threshold $A(t)$.

These are related as follows:

$$\dot{I}(t) = aE(t)(E(t) - I(t)), \qquad A(t) = I(t) + o, \tag{3.1}$$

a and o are positive constants. The activation condition for the cell is whenever $E(t) > A(t)$. The work of this chapter will be closely related to tunability. The word hypothesis in "tuneable activation threshold hypothesis" is key, it is backed by some experimental evidence, but largely it is an unconfirmed *general* theory. As discussed in the Chapter 1 the preference of this work is for models of immunology that are realistic. Spurred by the promise tuneability, the next section grounds discussion in actual T cell signalling events. Signalling events that have been implicated in tuning [Grossman and Paul, 2001]. The primary focus of this chapter will be in understanding T cell signalling events, a secondary focus will be the investigation for any inherent tunability.

3.3 T cell Signalling

The T cell must perform fine-grain discrimination between the abundant normal host pMHC, $99.9 - 99.99\%$ of all pMHC on an APC, and pMHC with peptides originating from foreign pathogens or non-normal conditions such as cancers. This second class comprises the other $0.01 - 0.1\%$ of the total pMHC expressed [Germain and Stefanová, 1999][3]. The difficulty is that the affinity of each class for the TCR does not differ by the same degree [Chan et al., 2001]. Moreover, the bond between the TCR and pMHC is low-affinity [Valitutti et al., 1995]; it is approximately 3-7 orders of magnitude weaker than between antibody and antigen. The lock-and-key paradigm that can be used with success to describe antibody recognition does not apply to the TCR. An explanation of the T cell's behaviour can be found in the complex and dynamic signalling cascades that arise from the internal component of the TCR. There are many T cell signalling pathways which offer a candidate explanation for T cell discrimination, [Germain and Stefanová, 1999] offers an excellent review of many of these potential signalling mechanisms. This chapter focuses on one in particular, which involves a set of signalling events that include *kinetic proofreading*, a negative influence from SH2 domain-containing phosphatase (SHP1) and a positive influence from extracellular signal regulated kinase (ERK).

Before discussing the details of these signalling events, it should be noted that they are all classed as *early signalling events*. That is, they occur immediately upon TCR pMHC engagement and they are responsible for determining how the APC–T cell interaction will proceed. The early signalling events precede *immunological synapse* formation [Bromley et al., 2001] which is a semi-permanent association between the T cell and APC. The signalling picture is complicated by synapse formation as each cell alters shape to generate a large area of adhesion between the two cells. However, the early signalling events occur on a timescale such that synapse formation can be (and is) ignored [Altan-Bonnet and Germain,

[3]It is this discrimination ability, combined with the potentially appealing properties of the TAT that makes this T cell signalling behaviour a desirable candidate for artificial immune systems.

2005]. In particular, the non-spatial assumptions used in [Altan-Bonnet and Germain, 2005] and this chapter only suitably hold on the timescale of early signalling events. It should be noted that early signalling events still are of importance to T cell discrimination; they may dictate whether synapse formation will proceed and in the case of the cytotoxic T cell (which is the variant considered here), the decision to kill a cell is made during these early events.

Returning to the specifics of the signalling process considered in this chapter, they are discussed in reference to tunability in [Germain and Stefanová, 1999; Grossman and Paul, 2001] and experimentally investigated in [Stefanová et al., 2003; Altan-Bonnet and Germain, 2005]. This signalling system has also received much modelling attention in [Chan et al., 2001; Altan-Bonnet and Germain, 2005; Artyomov et al., 2007; Wylie et al., 2007; Lipniacki et al., 2008]. The important quantitative and qualitative concepts that these signalling events impose on T cell discrimination are well summarised in [Chan et al., 2001]. Altan-Bonnet and Germain (ABG) [Altan-Bonnet and Germain, 2005] complement biological experiment with a detailed mathematical model of these signalling events. This ABG model is further discussed in [Feinerman et al., 2008a] and analysed in [Feinerman et al., 2008b]. The approach taken by [Altan-Bonnet and Germain, 2005] and [Feinerman et al., 2008b] is to model, as faithfully as possible, the bio-chemical reactions associated with the TCR, SHP1 and ERK and to demonstrate that their model replicates biological behaviour. To this end [Altan-Bonnet and Germain, 2005] achieves a model of 557 chemical reaction equations which are converted to 238 ordinary differential equations. Simplified versions of the concepts in the ABG model are produced in [Artyomov et al., 2007; Lipniacki et al., 2008] and stochastic analysis is performed. [Artyomov et al., 2007; Lipniacki et al., 2008] argue the importance of stochastic modelling particularly when small molecule numbers can produce bistabilities in a stochastic model that do not exist in the deterministic model.

The simplified models analysed in [Chan et al., 2001; Artyomov et al., 2007; Lipniacki et al., 2008] convey an understanding of the qualitative aspects of TCR, SHP1, ERK signalling. However, the exact connection between all the signalling detail in ABG model and the qualitative understanding gained from the simplified models is still not clear. The ABG model fits the criteria of being *realistic* and *precise*. This chapter takes the opportunity to dissect the biologically detailed ABG model with the objective of gaining clarified understanding of the biological processes that contribute to the qualitative behaviour. A stochastic version of the ABG model[4] is constructed which allows investigation at the single TCR level. The analysis starts with the behaviour of the TCR-pMHC bind and progressively composes all elements to re-build the ABG model. This approach provides understanding without the necessity of the simplifying assumptions such as the ones in [Lipniacki et al., 2008].

The analysis of the model proceeds as follows: Section 3.4 provides a biological background to the model; Section 3.5 justifies the stochastic approach and overviews the modelling methods used. The subsequent sections break the model and analyse sub-models.

[4]I am grateful to Gregoire Altan-Bonnet for providing the details of the ABG model

Section 3.6 investigates coreceptor CD8, TCR and pMHC binding; Section 3.7 kinetic proofreading; Section 3.8 negative feedback; Section 3.9 the mitogen-activated protein kinase (MAPK) MAPK cascade and protection of TCR; Section 3.10 then re-composes the entire model; finally Section 3.11 provides discussion and conclusions.

3.4 Model Biological Background

As noted earlier, the quality of a TCR pMHC association is well abstracted by its average lifetime [Feinerman et al., 2008a]. It may be classified by the response elicited in the cell, taking the definitions of [Germain and Stefanová, 1999]:

- *Agonist.* Will induce all possible activation signals within a cell. A TCR-pMHC lifetime of ≈ 18 seconds [Altan-Bonnet and Germain, 2005].
- *Antagonist.* Will actively inhibit activation signals within the cell.
- *Partial agonist.* Will induce a subset of all possible activation signals within a cell.
- *Null.* Will not have any effect, activatory or inhibitory.

A range of signal strengths exists within the first three classes. It is not the case that the signals induced by a weak agonist are necessarily stronger than those induced by a partial agonist. Normal-host-peptides fall into antagonist, partial agonist or null classes [Germain and Stefanová, 1999]. Three salient features of this TCR signalling model are: kinetic proofreading, negative feedback and a destruction of the negative feedback. Figure 3.2 depicts these features and provides reference to the appropriate section for analysis.

3.4.1 Kinetic Proofreading

Kinetic proofreading was first introduced to describe the accuracy of DNA replication and protein synthesis [Hopfield, 1974]. McKeithan [McKeithan, 1995] applied kinetic proofreading to T cell signalling and it is now a widely accepted model to account for ligand discrimination [Stefanová et al., 2003]. The process entails energy consuming steps that occur after association of pMHC to the TCR. The steps must be overcome for successful TCR signalling. With dissociation of pMHC from the TCR the steps are rapidly reversed. The result is step-like discrimination of pMHC ligands. Fast dissociating pMHC ligands fail to complete all steps, conversely all pMHC that bind long enough to complete all steps signal equally well. Kinetic proofreading provides a measure of the time the TCR and pMHC are associated. The steps involve phosphorylations[5] by Leukocyte-specific protein tyrosine kinase (Lck) of Immunoreceptor tyrosine-based activation motifs (ITAMs) on the TCR's internal ζ, ϵ and δ chains. Kinetic proofreading is, however, insufficient to explain antagonism [Feinerman et al., 2008a]. Further the model also fails in the high-density low-quality ligand case where stochastic fluctuations can allow poor quality ligands to overcome kinetic proofreading.

[5] A *phosphorylation* is the addition of a phosphate group to an already present molecule in the system.

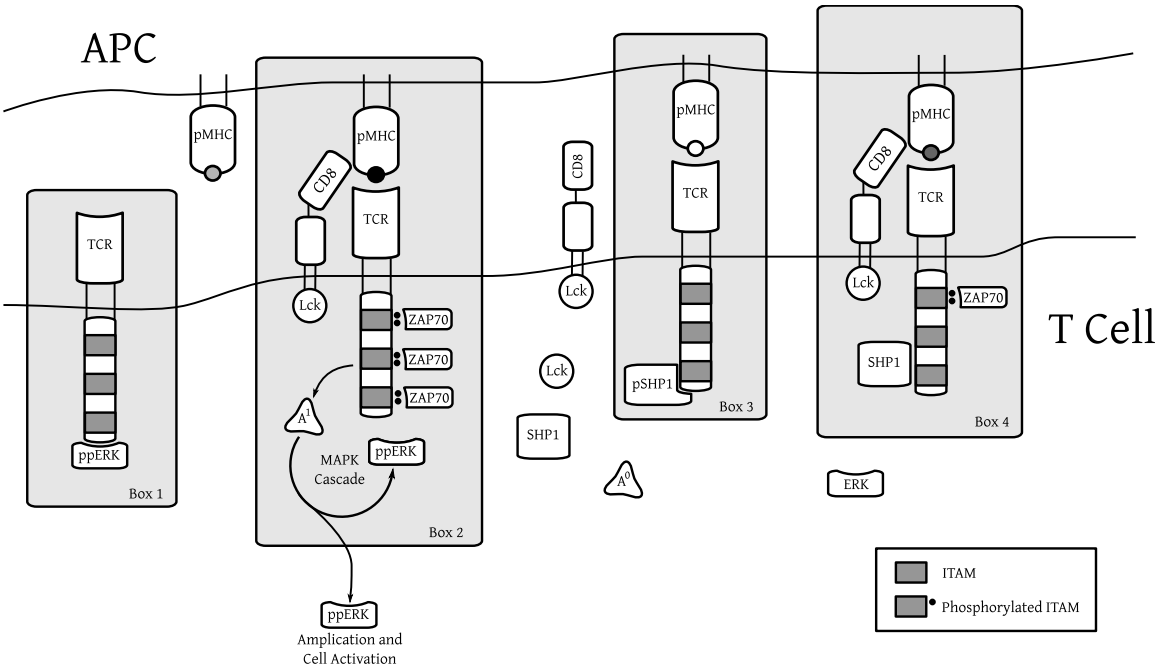

Figure 3.2: The TCR signalling processes. The TCR and pMHC may bind with co-receptor CD8, see *Box 2* and Section 3.6 for analysis. Upon association of pMHC to a TCR kinetic proofreading may start which involves phosphorylation of ITAMs and binding of ZAP-70 molecules. *Box 4* depicts a partially complete proofreading process; *Box 2* depicts a fully completed proofreading process. See Section 3.7 for kinetic proofreading analysis. A TCR internal chain with at least one ZAP-70 molecule may initiate negative feedback by binding and phosphorylating SHP1, see *Box 4*. Phosphorylated SHP1 (pSHP1) may then bind a TCR and upon further Lck action all phosphorylations will be lost, see *Box 3*. See Section 3.8 for analysis of negative feedback. A completed proofreading process may phosphorylate an adapter protein which initiates the MAPK cascade, resulting in high levels of ppERK, see *Box 2*. ppERK carries the activation signal of the cell and may bind a TCR internal chain and protect it from the action of pSHP1, see *Box 1*. See Section 3.9 for analysis of the MAPK cascade and TCR protection; see Section 3.10 for simulations of the entire system.

3.4.2 Negative Feedback

A negative feedback investigated experimentally in [Stefanová et al., 2003] may augment proofreading to explain antagonistic behaviour. The process is initiated by phosphorylation of SHP1 by Lck on the TCR internal complex. Phosphorylated SHP1 (pSHP1) may then associate to the TCR and dephosphorylate TCR internal chains. Thus, the process is initiated by proofreading steps and actively inhibits proofreading and so is a true negative feedback. The inclusion of negative feedback provides an explanation for antagonism and prevents large populations of low-quality ligands stochastically overcoming proofreading. A model of proofreading with negative feedback would suggest that the highest quality pMHC ligands would induce the largest negative feedback. However, this is not the case. There is a point at which ligand quality increases and the pSHP1 negative signal disappears [Stefanová et al., 2003].

3.4.3 Breaking the Negative Feedback

An explanation for the disappearance of the negative feedback is ascribed to a positive feedback through doubly phosphorylated ERK (ppERK) [Altan-Bonnet and Germain, 2005]. ppERK protects the TCR internal complexes from the action of pSHP1 by preventing pSHP1 from binding. Completion of kinetic proofreading initiates the MAPK cascade which results in the amplification of the proofreading activation signal by the production of large amounts of ppERK [Altan-Bonnet and Germain, 2005]. This acts dually to break the negative feedback and to carry the activation signal for the cell. How the activation signal determines cell fate is beyond the scope of this work.

The ppERK signal is often labelled as *digital* [Altan-Bonnet and Germain, 2005] in that it exhibits a step-like response. However, a clearer description is *binary*: it is either high or low and not found in discrete steps as the term digital would imply. Moreover, this signal is strictly not a positive feedback and it can be misleading to label it as one. The signal breaks the negative feedback allowing kinetic proofreading to continue with no inhibition, but only at the rate dictated by kinetic proofreading. One would expect this to be confirmed by experiment: a T cell with SHP1 removed would exhibit no negative feedback and allow observation of ppERK behaviour in isolation. Properties traditionally associated with positive feedback such as explosive amplification are present but are facets of the feed-forward MAPK cascade. A result found in Feinerman et al. [Feinerman et al., 2008b] can be interpreted to confirm this is not a positive feedback. Variation in the concentration of ppERK has no influence on the ability of a cell to appropriately signal. If ppERK was involved in a positive feedback its intra-cellular concentration represents a maximum bound on the positive signal. Thus variations in ppERK concentration would change the ability of the cell to signal. Interpretations of results in Lipniacki et al. [Lipniacki et al., 2008] give further weight to this argument. Changes in Lck concentration influence pSHP1 levels far more than ppERK levels; if ppERK were in positive feedback the action of Lck would be part of the feedback loop and so would influence ppERK levels. The behaviour of the MAPK cascade (analysed in Section 3.9) also contributes to these results. The ppERK signal is now referred to as a break in the negative feedback and not a positive feedback. As discussed in Chan et al. [Chan et al., 2001] the break in the negative feedback allows the T cell to remain sensitive to good agonists and helps to define a sharp discrimination threshold.

The signalling mechanisms may be mapped to TCR-pMHC bind classification:

1. Agonist – pMHC remains associated sufficiently long for a high probability of inducing MAPK cascade despite pSHP-1 dampening. The induction of the MAPK cascade will result in a high ppERK protection and activation signal.

2. Antagonist – pMHC associates sufficiently long to produce high levels of pSHP-1 but not to induce MAPK cascade.

3. Partial Agonist (Endogenous Ligand) – Does not induce high levels of pSHP-1, but may induce some partial phosphorylation of TCR internal chains.

The spreading of the pSHP-1 and ppERK signals is of interest; it is discussed in [Altan-Bonnet and Germain, 2005] and analysed directly in [Chan et al., 2001]. pSHP1 generated by a TCR will spread to dampen surrounding TCRs. The protection signal generated by a TCR will spread to protect surrounding TCRs. A hypothesis discussed in [Chan et al., 2001; Feinerman et al., 2008a] suggests that the protection signal allows endogenous ligands to "synergise" with agonist ligands and contribute to activation signalling. This suggestion is discussed in light of the findings of this chapter's analysis in Section 3.11 and a clarification to the argument is presented. The model of Chan et al. [Chan et al., 2001] investigates the spreading of the negative feedback and protection on a lattice. No investigation to the necessity of space to the model is given, the analysis performed here continues with the assumptions of [Altan-Bonnet and Germain, 2005] and analyses a non-spatial model.

3.5 Modelling Methods

In contrast to the deterministic approach of [Altan-Bonnet and Germain, 2005] and [Feinerman et al., 2008b] the models of [Chan et al., 2001; Artyomov et al., 2007; Wylie et al., 2007; Lipniacki et al., 2008] are stochastic in nature. In line with the arguments made in Chapter 2, and using the methods described in that chapter the ABG model is stochastically formulated. In particular, analysis is performed via the rate matrix of the master equation/continuous time Markov chain associated with the chemical reaction equations. Due to model size this is generally restricted to numerical solutions. As the full model is composed the analysis is only tractable through Gillespie Algorithm simulation.

The macroscopic rates of the ABG model are supplied in Table A.2, and converted to mesoscopic rates using the methods given in Chapter 2. The choice of three different ϵ used in this chapter are justified in Section 3.11.1 The definitions of names and rates are given in Appendix A.1; notation used in reaction equations is that described in Chapter 2.

3.6 TCR, pMHC, CD8 Binding

The ABG model enumerates the complexes nascent to the CD8-TCR-pMHC ternary complex. Reactions in and out of intermediate states TCR-pMHC, CD8-TCR, and CD8-pMHC are included. Other models that include co-receptor binding such as [Wylie et al., 2007] do not include CD8-TCR and CD8-MHC intermediates. The role of these intermediates is now investigated, particularly the influence of TCR-CD8. The reactions are depicted in Figure 3.3 and are as follows:

Reaction 1. *TCR pMHC CD8 Binding/Debinding*

$$\mathrm{T} + \mathrm{M} \underset{m_d}{\overset{m_a}{\rightleftharpoons}} \mathrm{B} \qquad \mathrm{T} + \mathrm{C} \underset{c_d}{\overset{c_t}{\rightleftharpoons}} \mathrm{TC}, \tag{3.2}$$

$$\mathcal{A} + \mathrm{C} \underset{c_d}{\overset{c_a}{\rightleftharpoons}} \mathcal{A}\mathrm{C} \qquad \mathcal{A} = \{\mathrm{M}, \mathrm{C}\}, \tag{3.3}$$

$$\mathrm{TC} + \mathrm{M} \underset{c_d \times c_s}{\overset{c_a}{\rightleftharpoons}} \mathrm{BC} \qquad \mathrm{MC} + \mathrm{T} \underset{c_d \times c_s}{\overset{c_a}{\rightleftharpoons}} \mathrm{BC}. \tag{3.4}$$

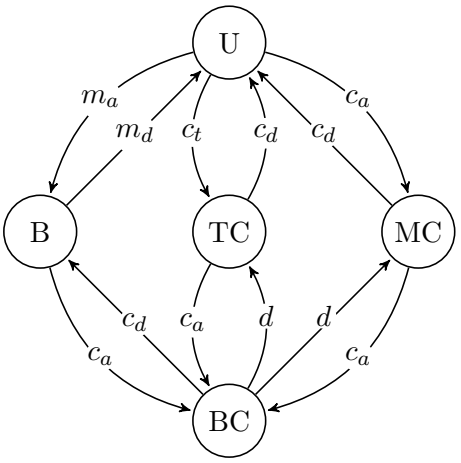

Figure 3.3: Diagrammatic representation of Reaction 1. A single TCR, pMHC and CD8 molecule are the reactants. U gives the completely unbound state, all other states are denoted by the bound product. Arcs are labelled with the reaction rate, d denotes reaction rate $m_d c_s$.

The reactions describe the following assumptions: The CD8 co-receptor acts to hook-in pMHC to the TCR increasing the on-rate; CD8 stabilises the TCR-pMHC by a multiplicative factor $c_d < 1$; CD8 may dissociate and re-associate with fast kinetics [Wyer et al., 1999], with the exception of state TC whose on rate c_t which is considerably smaller than c_a.

A volume restriction $\epsilon = N_T^{-1}$ is applied to scale the rates to the volume occupied by a single TCR. The stationary and transient distributions parameterised by $m_d \in [0.001, 1000]$ are given in Figure 3.4. The distributions show that the unbound state U and the bound state B have the highest probability. The probability of being in state TC is orders of magnitude lower than being in any other state nascent to BC. Write $\mathbb{P}_i(j)$ for the probability that given the current state is i, the next state is state j; then setting $m_d = 1/18$ and all other rates as Table A.2:

$$\mathbb{P}_{\mathrm{U}}(\mathrm{TC}) < \mathbb{P}_{\mathrm{U}}(\mathrm{B}) < \mathbb{P}_{\mathrm{U}}(\mathrm{MC}). \tag{3.5}$$

There is an order of magnitude difference between the three probabilities. In the opposite direction, the probabilities for leaving state BC:

$$\mathbb{P}_{\mathrm{BC}}(\mathrm{TC}) = \mathbb{P}_{\mathrm{BC}}(\mathrm{MC}) \ll \mathbb{P}_{\mathrm{BC}}(\mathrm{B}). \tag{3.6}$$

The inequality will hold if $m_d < c_d/c_s = 1000$ which is the case for realistic ligands. Then, if $\tau(i)$ is the holding time of state i (that is the expected time before leaving state i):

$$\tau(\mathrm{BC}) < \tau(\mathrm{TC}) = \tau(\mathrm{MC}) \ll \tau(\mathrm{U}) < \tau(\mathrm{B}) \quad \text{for } \epsilon = 1 \tag{3.7}$$

$$\tau(\mathrm{TC}) = \tau(\mathrm{BC}) < \tau(\mathrm{MC}) \ll \tau(\mathrm{B}) < \tau(\mathrm{U}) \quad \text{for } \epsilon = N_T^{-1}, \tag{3.8}$$

$\tau(\mathrm{U}) \approx \tau(\mathrm{B})$ and $\tau(\mathrm{BC}) \approx \tau(\mathrm{MC})$ for $\epsilon = N_T^{-1}$ and the inequalities hold for the majority of realistic ligands. The state TC is the least likely state to enter and has one of the shortest holding times, which explains its low probability at equilibrium.

The transient and stationary distributions are recalculated with the removal of state TC (the new stationary distribution is labelled $\boldsymbol{\pi}^{\mathrm{TC}}$) and they are shown in Figure 3.4. The stationary distributions and the trajectories are effectively identical. Under the assumption of approximately equal TCR, pMHC and CD8 concentrations the state TC may be removed

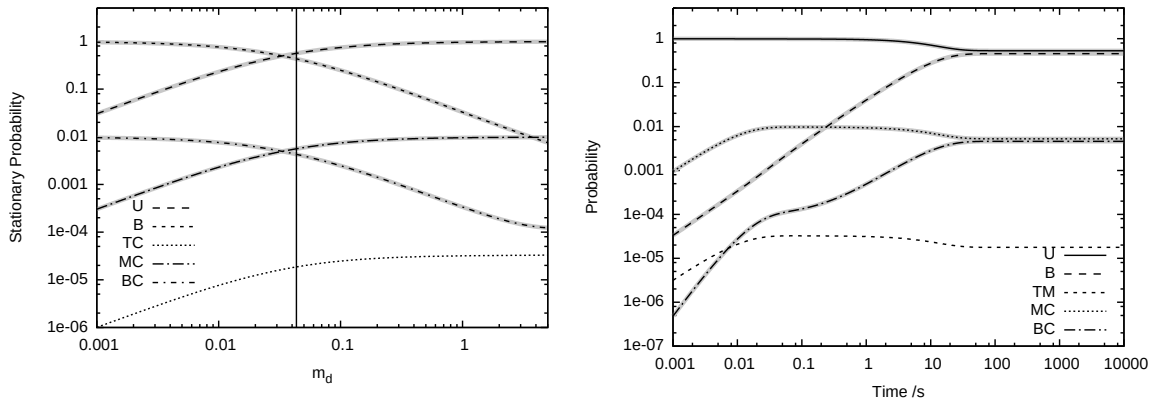

Figure 3.4: *Left*: Stationary distributions for $m_d \in [0.001, 5]$. Black dotted lines are the distribution $\boldsymbol{\pi}$, thick grey lines are the distribution $\boldsymbol{\pi}^{\mathrm{TC}}$. They are effectively identical behaviour across the range of m_d. The vertical line is at $m_d = 0.0434$ and represents the point where $\pi_U = \pi_B + \pi_{BC}$. *Right:* Transient distributions for time $t \in [10^{-3}, 10^4]$, starting in an unbound state U with $m_d = 0.05$. Black dotted lines are the distribution with all states, thick grey lines are the distribution with state TC removed. The probabilities follow the same trajectories.

from the model. This provides a prediction that may be validated against biology, should more than $2 \times 10^{-3}\%$ of all TCRs be found bound solely to the CD8 co-receptor then this is an incorrect abstraction of biology. The incorrect abstraction may not be in structure of reactions but in choice of rate. A clear example of this is state MC which is identical to TC except the rate from U to MC is two orders of magnitude greater than that of U to TC. This results in approximately a three orders of magnitude increased stationary probability in Figure 3.4.

The binding system may be analysed in general to gain an understanding of the parameter choices. The systems with and without the state TC are instances of a general binding system with n intermediate state separating an unbound and a bound state,

Definition 3.1. *A general binding system $\mathcal{B}_n$ has $n+2$ states, that is, n intermediates denoted by the natural numbers: $1, 2, \ldots, n$ and an unbound state U and a bound state V. The index i is used over any of the $n+2$ states, $i = U, 1, 2, \ldots, n, V$; the index k is over the n intermediate states, $k = 1, 2, \ldots, n$. A binding system $\mathcal{B}_n$ is defined by the rate matrix $\mathbf{G}_n$:*

$$\mathbf{G}_n = \begin{pmatrix} -s_u & g_{u1} & \cdots & g_{un} & 0 \\ g_{1u} & -s_1 & & & g_{1v} \\ \vdots & & \ddots & & \vdots \\ g_{nu} & & & -s_n & g_{nv} \\ 0 & g_{v1} & \cdots & g_{vn} & -s_v \end{pmatrix}, \qquad s_i = \sum_j g_{ij}, \tag{3.9}$$

g_{uk} is the rate from state U to the kth intermediate state, the rate g_{ku} is the rate back in the other direction. Rates g_{kv} and g_{vk} are defined similarly.

The binding system with TC is, then, a $\mathcal{B}_3$ binding system and without TC it is a $\mathcal{B}_2$ system.

The stationary distribution for $\mathcal{B}_n$ is $\pi = (\pi_u, \pi_1, \ldots, \pi_n, \pi_v)$ and is found by solving $\boldsymbol{\pi}\mathbf{G}_n = \mathbf{0}$, subject to $\sum_i \pi_i = 1$. Since the sum of every row of $\mathbf{G}_n$ is zero and there is one additional constraint there are $n+2$ equations:

$$-s_u\pi_u + \sum_k g_{ku}\pi_k = 0 \tag{3.10}$$

$$g_{uk}\pi_u - s_k\pi_k + g_{vk}\pi_v = 0, \qquad \text{for } k = 1, 2, \ldots n \tag{3.11}$$

$$\sum_i \pi_i = 1. \tag{3.12}$$

The k intermediate equations (3.11) may be used to eliminate all π_k in the probability constraint (3.12), resulting in an equation in π_u and π_v:

$$\left(1 + \sum_k \frac{g_{uk}}{s_k}\right)\pi_u + \left(1 + \sum_k \frac{g_{vk}}{s_k}\right)\pi_v = 1. \tag{3.13}$$

For convenience, write $\sigma_\alpha = 1 + \sum_k \frac{g_{\alpha k}}{s_k}$. Then,

$$\pi_v = \frac{1 - \pi_u\sigma_u}{\sigma_v}. \tag{3.14}$$

Substituting back into the each of the k intermediates gives π_k in terms of π_u:

$$\pi_k = \frac{g_{vk} + \pi_u(g_{uk}\sigma_v - g_{vk}\sigma_u)}{s_k\sigma_v}. \tag{3.15}$$

Then, substituting the π_k back into (3.10), gives π_u:

$$\pi_u = \frac{\sum_k g_{ku}g_{vk}s_k^{-1}}{s_u\sigma_v - \sum_k g_{ku}s_k^{-1}(g_{uk}\sigma_v - g_{vk}\sigma_u)}. \tag{3.16}$$

By symmetry, π_v is:

$$\pi_v = \frac{\sum_k g_{kv}g_{uk}s_k^{-1}}{s_v\sigma_u - \sum_k g_{kv}s_k^{-1}(g_{vk}\sigma_u - g_{uk}\sigma_v)}. \tag{3.17}$$

Returning to (3.15), it can be rearranged as follows:

$$\pi_k = \frac{g_{vk}}{s_k\sigma_v}(1 - \pi_u\sigma_u) + \frac{g_{uk}}{s_k}\pi_u. \tag{3.18}$$

The first term as σ_v in the denominator, for the choice of rates in Figure 3.3, σ_v is composed as follows,

$$\begin{array}{rcccccccc} \sigma_v & = & 1 & + & \dfrac{c_d}{c_a^\star + m_d} & + & \dfrac{m_d c_s}{c_a^\star + c_d} & + & \dfrac{m_d c_s}{c_a^\star + c_d} \\ & & & & \wr & & \wr & & \wr \\ 96 & \approx & 1 & + & 95 & + & 5\times 10^{-5} & + & 5\times 10^{-5}. \end{array} \tag{3.19}$$

The $\star$ superscript denotes the mesoscopic scaled association rate and $c_a^\star \approx 1$. Thus, the ratio $\frac{g_{vk}}{s_k\sigma_v} \approx 10^{-6}$ is very small, and the quantity g_{uk}/s_k gives the greatest contribution to (3.18). This quantity is given in Table 3.1 for the states B, TC and MC. It is both that

State	B	TC	MC
$\frac{g_{uk}}{s_k}$	$\frac{m_a^\star}{c_a^\star + m_d}$	$\frac{c_t^\star}{c_a^\star + c_d}$	$\frac{c_a^\star}{c_a^\star + c_d}$
Approximate Value	3×10^{-2}	3×10^{-5}	10^{-2}

Table 3.1: The quantity g_{uk}/s_k for parameters given in Table A.2

.

c_t is much slower than c_a and that c_d is so much faster than $m_d c_s$ that contributes to the low stationary probability of TC.

The system with TC removed, a $\mathcal{B}_2$ system, is now solved for m_d subject to the additional constraint:

$$\pi_U = \pi_B + \pi_{BC}. \tag{3.20}$$

The solution provides the dissociation rate at which there is equal probability of the TCR and pMHC being bound and unbound. This gives a quadratic in m_d:

$$\begin{aligned} m_d^2 c_d c_s + m_d(c_d^2 + c_d(c_a + c_a c_s - m_a c_s) - c_a^2) \\ -c_a^3 - c_a^2(m_a + c_d) - 2c_a c_d m_a - c_d^2 m_a = 0. \end{aligned} \tag{3.21}$$

There is a single positive solution $m_d = 0.0434$ which gives an average bind life time of 23.04 seconds, a strong agonist. The $m_d = 0.0434$ line is marked on Figure 3.4.

The analysis thus far does not apply if the concentrations of TCR, CD8 or MHC change. The concentration of CD8 is of interest as it has been shown to be a potential parameter for T cell tuning [Feinerman et al., 2008b; van den Berg et al., 2007]. So, define $v = |\mathrm{C}|/|\mathrm{T}|$ and then new rates: $c_t' = vc_t$ the association rate of T and C; $c_a' = vc_a$ the association of C to M or B. Then, recalculating the transient and stationary distributions reveals that a difference is only observed at very high densities $v > 100$, so the stationary analysis should hold for $v < 100$.

3.7 Kinetic Proofreading

The kinetic proofreading in the ABG model has added complexity over standard kinetic proofreading models [McKeithan, 1995]. It is described by the following reactions:

Reaction 2. *Soluble Lck Phosphorylation:*
Normal Phosphorylation:

$$\mathrm{B}_z^q + \mathrm{L} \underset{l_d}{\overset{l_a}{\rightleftharpoons}} \mathrm{B}_z^q\mathrm{L} \xrightarrow{l_c} \mathrm{B}_z^{q+1} + \mathrm{L}, \qquad q = 0, 1;\ z = 0, 1, 2 \tag{3.22}$$

$$\mathrm{B}_3^m + \mathrm{L} \underset{l_d}{\overset{l_a}{\rightleftharpoons}} \mathrm{B}_3^m\mathrm{L} \xrightarrow{l_c} \mathrm{B}_3^{m+1} + \mathrm{L}, \qquad m = 0, 1, 2. \tag{3.23}$$

Reaction 3. *CD8 association Lck Phosphorylation:*
Normal Phosphorylation:

$$\mathrm{B}_z^q\mathrm{C} \underset{l_{cd}}{\overset{l_{ca}}{\rightleftharpoons}} \mathrm{B}_z^q\mathrm{CL} \xrightarrow{l_{cc}} \mathrm{B}_z^{q+1}\mathrm{C}, \qquad q=0,1;\ z=0,1,2 \tag{3.24}$$

$$\mathrm{B}_3^m\mathrm{C} \underset{l_{cd}}{\overset{l_{ca}}{\rightleftharpoons}} \mathrm{B}_3^m\mathrm{CL} \xrightarrow{l_{cc}} \mathrm{B}_3^{m+1}\mathrm{C}, \qquad m=0,1,2, \tag{3.25}$$

Lck associated with CD8 is implicitly part of the C complex.

Reaction 4. *ZAP-70 Binding:*

$$\mathrm{B}_z^2 + \mathrm{Z} \underset{z_d}{\overset{z_a}{\rightleftharpoons}} \mathrm{B}_{z+1}^0, \qquad z=0,1,2. \tag{3.26}$$

ZAP-70 Debinding from TCR:

$$\mathrm{T}_z \xrightarrow{z_d} \mathrm{T}_{z-1} + \mathrm{Z}, \qquad z=1,2,3. \tag{3.27}$$

Reaction 5. *Dephosphorylation upon TCR binding/debinding:*

$$\mathcal{Q}_z^q \xrightarrow{\delta} \mathrm{T}_z, \qquad \mathrm{T}_z \xrightarrow{\alpha} \mathcal{Q}_z, \qquad z=0,1,2,3 \quad q>1. \tag{3.28}$$

Here if $\mathcal{Q}=\mathrm{B}$ then $\alpha=m_a, \delta=m_d$ and if $\mathcal{Q}=\mathrm{BC}$ then $\alpha=c_a, \delta=m_d c_s$

The reactions imply the following assumptions:

- Upon association of pMHC to the TCR internal chains undergo enzymatic Lck phosphorylation. This may occur via cytosolic Lck or Lck associated with CD8. The CD8 associated Lck has a greater rate of association to the TCR $l_{ca} \gg l_a$. The phosphorylations are restricted to occur in serial in order.
- ZAP-70 may bind and protect a double phosphorylated ITAM, 3 ZAP-70 molecules may bind in all. Three phosphorylations may occur once the third and final ZAP-70 molecule has bound.
- Dissociation of TCR and pMHC causes loss of phosphorylations not protected by ZAP-70 via a fast acting phosphatase. The fast-acting phosphatase is not modelled here, the phosphorylations are immediately lost upon dissociation. ZAP-70 and its protected phosphorylations may persist between TCR binds.
- The TCR internal chain may initiate activation signalling from a state B_3^q with $q>1$. The specific details of the activation signalling are not included until Section 3.9.

Note that the gain and loss of phosphorylations in the kinetic proofreading reactions appears to break conservation. This is notational convenience and the necessary phosphates are implicitly included in the reactants and products, this convention is used throughout this chapter.

m_d	$\max\{\mathbb{P}(\mathrm{B}_3^q,\ q>1)\}$	m_d	$\max\{\mathbb{P}(\mathrm{B}_3^q,\ q>1)\}$
0.03	1.6×10^{-1}	0.44	1.0×10^{-5}
0.06	0.3×10^{-1}	0.85	4.3×10^{-7}
0.12	4.5×10^{-3}	1.64	2.5×10^{-8}
0.23	2.6×10^{-3}	3.16	2.5×10^{-9}

Table 3.2: Maximum probability of signalling state for various m_d. This can approximate the expected number of ligands needed to achieve activation via $[\max\{\mathbb{P}(\mathrm{B}^q\mathrm{Z}_3,\ q>1)\}]^{-1}$ (the coupling between TCRs at this stage only involves competition for Lck and CD8, as such this approximation should be reasonable). For $m_d^{-1}\approx 30$s about 6 ligands are needed to guarantee a signalling state; for $m_d^{-1}\approx 16$s about 25 ligands are required to guarantee signalling state.

The kinetic proofreading reactions do not include the negative effects of pSHP1 and so describe the case when the TCR has been protected by ppERK. Write $\mathbf{K}$ for the rate matrix describing the kinetic proofreading reactions for single TCR, pMHC, CD8 molecules. To calculate the expected kinetic proofreading state $\mathbb{E}_K$, it is necessary to assign weights $\omega(s)$ linearly to all states $s\in\mathcal{K}$, the set of bound kinetic proofreading states:

$$\omega(\mathrm{B}_z^q)=3z+q+1, \tag{3.29}$$

that is, $\omega(\mathrm{B}_0^0)=1,\omega(\mathrm{B}_0^1)=2,\ldots,\omega(\mathrm{B}_3^3)=13$. Then $\mathbb{E}_K$ is given:

$$\mathbb{E}_K=\sum_{s\in\mathcal{K}}\omega(s)\mathbb{P}(s). \tag{3.30}$$

Figure 3.5 gives the transient probability for $\mathbb{E}_K$ over a range of m_d. The kinetic proofreading behaviour may be summarised by calculating the signalling state with the maximum probability, that is $\max\{\mathbb{P}(\mathrm{B}_3^q, q>1)\}$. The results are given in Table 3.2.

Figure 3.5 also gives the kinetic proofreading stationary distributions over m_d. The clumping together of probabilities of mid-kinetic proofreading states at low dissociation rates conveys the notion that all states are equally likely en route to the final state. Due to the high concentration of ZAP-70, the states in which a new ZAP-70 molecule may bind (B_0^2, B_1^2, B_2^2) have the lowest stationary probability ($\approx 10^{-6}$).

To further analyse the time taken by kinetic proofreading and particularly the influence of CD8 the expected hit times of TCR signalling state are calculated whilst varying v and m_d, see Figure 3.6. CD8 density v can modulate hitting times and bring them in line with the 1-5 minute timescales discussed in [Feinerman et al., 2008a]. An explanation as to why CD8 density is so successful at decreasing kinetic proofreading times is given in the following Section 3.7.1.

3.7.1 Proofreading Locking Mechanism

The kinetic proofreading mechanisms analysed here produce the step-like response prescribed by conventional kinetic proofreading models [McKeithan, 1995]. However the ability of ZAP-70 to "protect" TCR phosphorylations between TCR binds breaks McKeithan's

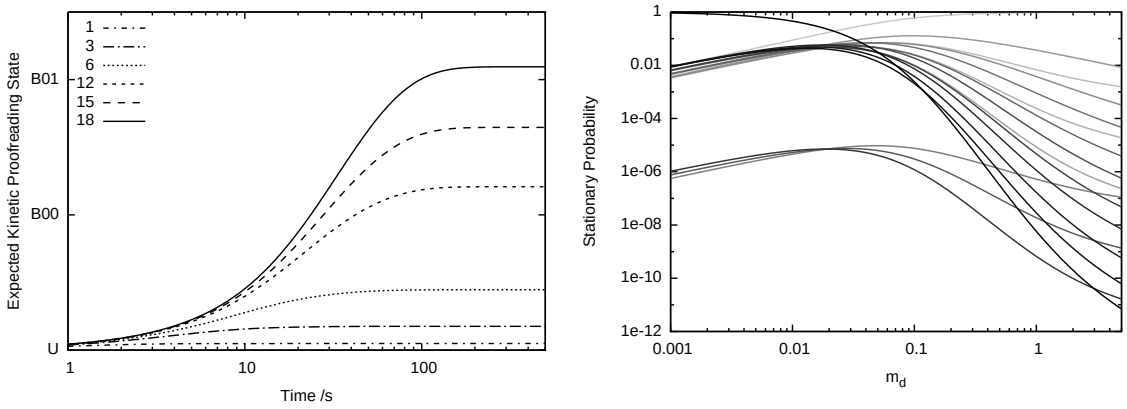

Figure 3.5: Kinetic proofreading transient and stationary behaviour. *Left:* The transient probability of $\mathbb{E}_K$ from initial unbound zero phosphorylation state, calculated by $e^{t\mathbf{K}}$, with $m_d^{-1} = 1, 3, 6, 12, 15, 18$. *Right* The stationary distribution. Lines represent $U_z \dots B_z^q$ and are coloured on a linear grey-scale: with U_0 lightest and B_3^3 darkest. $\mathbb{P}(U_0) \approx 1$ at large m_d, and $\mathbb{P}(B_3^3) \approx 1$ at small m_d. The inflexion between light and dark lines can be interpreted as the point where the latter half of the kinetic proofreading process becomes more probable.

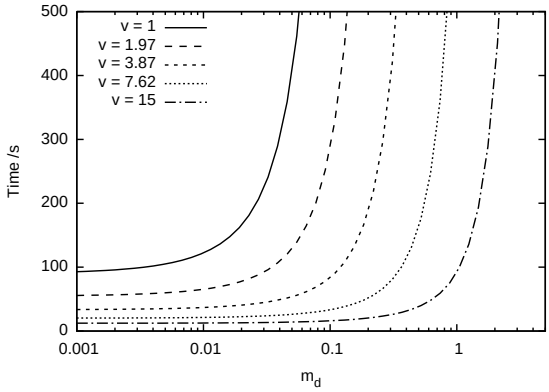

Figure 3.6: Expected hitting times of kinetic proofreading signalling state B_3^q, $q > 1$ varying m_d and v. Modulation of CD8 dictates the dissociation rate at which there is a steep increase in kinetic proofreading times.

model. The locking mechanism is now directly investigated. Removing the locking system amounts to changing Reaction 5 to the following:

Reaction 6. *Dephosphorylation and de-binding of ZAP-70 upon TCR debinding:*

$$\mathcal{Q}_z^q \xrightarrow{\delta} T_0^0, \qquad z = 0, 1, 2, 3 \quad q = 0, 1, 2, 3. \tag{3.31}$$

If $\mathcal{Q} = B$ then $\delta = m_d$ and if $\mathcal{Q} = BC$ then $\delta = m_d c_s$

Calculating non-locking kinetic proofreading transient and stationary distributions reveals a reduced probability of all later kinetic proofreading states. Further, the stationary distributions do not exhibit the inflexion of Figure 3.5.

The differences between non-locking and locking proofreading are investigated by varying pMHC and CD8 densities. Define $u = |M|/|T|$ and new rates: $m_a' = um_a$ the association rate of T and M; $c_m' = uvc_a$ the association rate of C and M; $m_c = uc_a$ the association rate of M to TC; $c_a' = vc_a$ the association rate of C to B.

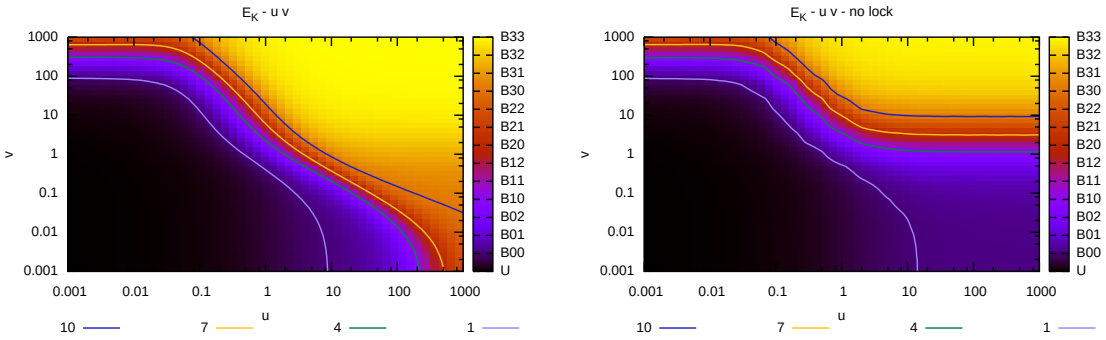

Figure 3.7: Stationary probabilities of $\mathbb{E}_K$ for $m_d = 1/18$, varying u and v *left* with ZAP-70 locking; *right* without ZAP-70 locking right. The locking mechanism greatly increases the area of the (u, v)-plane with a high expected proofreading state. This is due to u and v contributing to a greater probability of TCR-pMHC re-association before a ZAP-70 is lost. Without locking and $v < 1$ results in a low expected proofreading state.

The stationary distributions of $\mathbb{E}_K$ are recalculated using the new rates whilst varying u and v. Figure 3.7 gives these distributions [6] for a good agonist $m_d = 1/18$, this behaviour is found to be qualitatively similar for other values of m_d. High concentrations of CD8 and pMHC result in high $\mathbb{E}_K$. A high concentration of pMHC or CD8 increases TCR association rates and so increases the probability that the TCR and pMHC will re-associate before a ZAP-70 is lost. Mesoscopic rates for $\epsilon = N_T^{-1}$ are $m_a = 0.033$, $d_z = 0.11$. An increase in pMHC concentration of 3.33 will make re-association of the TCR and pMHC equally likely as the loss of ZAP-70. An increase in CD8 density v also contributes to the completion of kinetic proofreading steps due to efficient CD8-Lck. Thus changes to CD8 density alter the T cell's reactivity to a given pMHC ligand, and it is the locking mechanism which allows the CD8 co-receptor to be efficacious. This ZAP-70 hypothesis is confirmed in Figure 3.8 by varying ZAP-70 dissociation rate d_z with the inclusion of the locking mechanism. Decreases in d_z result in increases in $\mathbb{E}_K$ because slow d_z increases the probability that a ZAP-70 will remain associated between TCR-pMHC associations. Due to the high abundance of cytosolic ZAP-70 any ZAP-70 molecules lost during TCR-pMHC association will be rapidly replaced. Thus the periods when the TCR is dissociated benefit the most from decreases in d_z. Slow dissociations of ZAP-70 increase the efficacy of serial triggering.

3.8 Negative Feedback

The negative feedback signal is carried via cytosolic SHP-1, which may bind to a TCR internal chain with at least one ZAP-70 molecule. The following describes the negative

[6] The extremes of density here are larger than would occur naturally, they are included to demonstrate the range of behaviour, moreover the contribution to the parameters u and v could equally come from increased association rates.

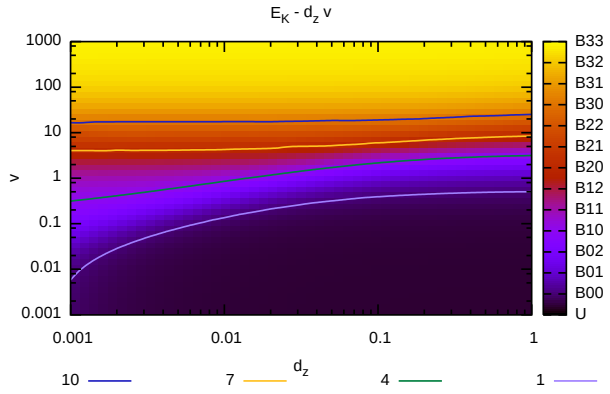

Figure 3.8: Stationary probabilities for $\mathbb{E}_K$ with $m_d = 1/18$ varying ZAP-70 dissociation rate d_z and CD8 density v with the locking mechanism. Decreases in d_z increase $\mathbb{E}_K$ because the probability that a ZAP-70 molecule will remain associated between TCR pMHC associations.

feedback process:

Reaction 7. *SHP-1 Binding, and phosphorylation to pSHP-1*

$$\mathrm{B}_z + \mathrm{S}^0 \underset{s_d}{\overset{s_a}{\rightleftharpoons}} \mathrm{B}_z\mathrm{S}^0 \xrightarrow{s_c} \mathrm{B}_z + \mathrm{S}^1 \qquad z = 1, 2, 3. \tag{3.32}$$

The phosphorylated pSHP-1 may load and unload from the TCR:

Reaction 8. *Binding of pSHP-1*

$$\mathcal{D} = \{\mathrm{T}_z,\, \mathrm{B}_z,\, \mathrm{B}_3^0\}, \qquad z = 0, 1, 2.$$

$$\mathcal{D} + \mathrm{S}^1 \underset{s_d}{\overset{s_a}{\rightleftharpoons}} \mathcal{D}\mathrm{S}^1. \tag{3.33}$$

This applies for $z < 3$ and $q > 0$ cases, for $z = 3$ there is one-step dephosphorylation:

$$\mathrm{B}_3^q + \mathrm{S}^1 \xrightarrow{s_a} \mathrm{B}_3^0 + \mathrm{S}^1, \qquad q = 1, 2, 3. \tag{3.34}$$

Upon Lck phosphorylation of a TCR internal chain with pSHP1 loaded all phosphorylations not protected by ZAP-70 are lost:

Reaction 9. pSHP1 Dephosphorylation through Lck:

$$\mathrm{B}_z^q\mathrm{S}^1 + \mathrm{L} \underset{l_d}{\overset{l_a}{\rightleftharpoons}} \mathrm{B}_z^q\mathrm{S}^1\mathrm{L} \xrightarrow{l_c} \mathrm{B}_z^0\mathrm{S}^1 + \mathrm{L}, \qquad q = 0, 1, 2;\ z = 0, 1, 2 \tag{3.35}$$

$$\mathrm{B}_3^0\mathrm{S}^1 + \mathrm{L} \underset{l_d}{\overset{l_a}{\rightleftharpoons}} \mathrm{B}_3^0\mathrm{S}^1\mathrm{L} \xrightarrow{l_c} \mathrm{B}_3^0\mathrm{S}^1 + \mathrm{L}. \tag{3.36}$$

CD8 association Lck Phosphorylation:

$$\mathrm{B}_z^q\mathrm{CS}^1 \underset{l_{cd}}{\overset{l_{ca}}{\rightleftharpoons}} \mathrm{B}_z^q\mathrm{CS}^1\mathrm{L} \xrightarrow{l_{cc}} \mathrm{B}_n^0\mathrm{CS}^1, \qquad q = 0, 1;\ n = 0, 1, 2 \tag{3.37}$$

$$\mathrm{B}_3^0\mathrm{CS}^1 \underset{l_{cd}}{\overset{l_{ca}}{\rightleftharpoons}} \mathrm{B}_3^0\mathrm{CS}^1\mathrm{L} \xrightarrow{l_{cc}} \mathrm{B}_3^0\mathrm{CS}^1. \tag{3.38}$$

The Lck associated with CD8 is implicitly part of the C complex.

The above reactions taken from [Altan-Bonnet and Germain, 2005] do not include a conversion of pSHP1 back to SHP1 and so the set of states with all SHP1 converted to pSHP1 are absorbing. To ensure that all states are irreducible and recurrent so the stationary distribution exists and is meaningful [Norris, 1998], the following reaction is added:

Reaction 10. Dephosphorylation of pSHP1.

$$\mathrm{S}^1 + \mathrm{P_S} \underset{q_d}{\overset{q_a}{\rightleftharpoons}} \mathrm{P_S S}^1 \xrightarrow{q_c} \mathrm{P_S} + \mathrm{S}^0. \tag{3.39}$$

The rates q_d, q_c are chosen in line with the phosphatases of the MAPK cascade; q_a is chosen similar to the association of Lck. The association rate q_a is rate-limiting and pSHP1 will persist on a timescale longer than a TCR-pMHC association. This choice of rate produces a quantitatively identical transient behaviour over the first 200 seconds with and without Reaction 10. As a less than maximal level of pSHP1 is necessary to contain the, subsequently discussed, rise of $\mathbb{E}_K$, the inclusion of this Reaction 10 should not significantly alter the activation threshold of the cell. Thus, the choice of rates are justified in the sense that they do not influence initial transient behaviour which is of importance to the concluding remarks of this model analysis, Section 3.11. Now, a potential stationary distribution for negative feedback is calculated.

There is a discrepancy between pSHP1 dephosphorylation in the B_3 case compared to B_q with $q = 0, 1, 2$. An investigation is provided in Appendix A.2 and now the discussions continue with *uniform negative feedback* as described in Appendix A.2. A volume restriction of $\epsilon = N_T^{-1}$ gives ≈ 26 SHP-1 molecules. The rate matrix for negative feedback may be written as a block tridiagonal matrix:

$$\mathbf{N} = \begin{pmatrix} \mathbf{Y_0} & \mathbf{A} & & \\ \mathbf{Q} & \mathbf{Y_1} & \ddots & \\ & \ddots & \ddots & \mathbf{A} \\ & & \mathbf{Q} & \mathbf{Y_m} \end{pmatrix} \qquad \mathbf{Y_i} = \begin{pmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{D} & \mathbf{K} \end{pmatrix}, \tag{3.40}$$

$\mathbf{Y_i}$ is the kinetic proofreading matrix with i pSHP1 molecules. $\mathbf{K}$ describes the basic kinetic proofreading reactions; $\mathbf{P}$ and $\mathbf{D}$ give the binding and debinding of the pSHP1 phosphatase respectively. $\mathbf{A}$ contains s_c from Reaction 7 on appropriate diagonal positions. $\mathbf{Q}$ contains rate q_c from Reaction 10 on all diagonal positions. $\mathbf{N}$ is 7420×7420 with only 50438 non-zeros (approximately 0.09% of $\mathbf{N}$); it is amenable to the analysis methods outlined in Section 2.2.

The influence of negative feedback will be overestimated in the stationary distribution without the inclusion of the protection from pSHP1 by ppERK (see Section 3.9). However, the TCR must generate an activation signal at least once without ppERK and so consideration of the non-protection case is necessary.

Similar to $\mathbb{E}_K$ the expected pSHP1 level $\mathbb{E}_S$ is calculated as follows:

$$\mathbb{E}_S = \sum_{i=1}^{\max|\mathrm{S}^1|} i\mathbb{P}(|\mathrm{S}^1| = i). \tag{3.41}$$

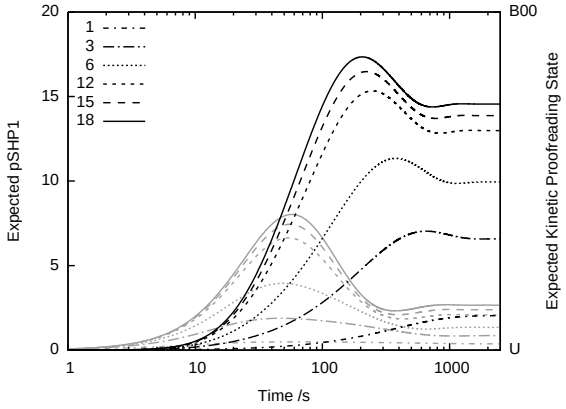

Figure 3.9: The transients of $\mathbb{E}_S$ (black lines, left axis) and $\mathbb{E}_K$ (grey lines, right axis) from initial unbound state U. With $m_d^{-1} = 1, 3, 6, 12, 15, 18$. Levels of pSHP1 lag the increase in kinetic proofreading state, resulting in an early hump in $\mathbb{E}_K$. Compare the heavily reduced values of $\mathbb{E}_K$ to kinetic proofreading without negative feedback in Figure 3.5

Figure 3.9 gives the transients of $\mathbb{E}_K$ and $\mathbb{E}_S$. The increase of $\mathbb{E}_S$ lags that of $\mathbb{E}_K$ because kinetic proofreading must reach state B_1^1 before pSHP1 can be produced. The level of pSHP1 will rise, arrest and reverse kinetic proofreading to a state $< \mathrm{B}_1^1$. With the conversion of pSHP1 back to SHP1 equilibrium is reached. The existence of a non-zero kinetic proofreading state that must be overcome before negative feedback is generated appears to be of great importance. This state is henceforth referred to as the *base negative feedback state*. If $\mathbb{E}_K$ and $\mathbb{E}_S$ are recalculated with base negative feedback state set at the zero kinetic proofreading state (B_0^0) then this hump in $\mathbb{E}_K$ and $\mathbb{E}_S$ is not observed. Moreover, the expected kinetic proofreading state and consequently expected pSHP1 levels are very low. Section 3.9 demonstrates that the existence of this base negative feedback state is of importance at the population level. The implications of this state are discussed in detail in the modelling conclusions (Section 3.11).

The stationary distributions of $\mathbf{N}$ and $\mathbb{E}_S$ (with the base negative feedback state reinstated) are given in Figure 3.10. States with zero phosphorylations not protected by ZAP-70 have the greatest probability. Write $\mathbb{E}_K(m_d)$ and $\mathbb{E}_S(m_d)$ for the expectations parameterised by m_d. For extremes of realistic ligands: $\mathbb{E}_S(0.05) \approx 15$ and $\mathbb{E}_S(1) \approx 2$. Kinetic proofreading of high quality pMHC ligands may be arrested by just over half-maximal pSHP1. One would expect this behaviour could be confirmed from experimentation with cells with ERK removed; one would look to confirm that pSHP1 will rise sufficiently to dampen kinetic proofreading of all ligands including agonists. Observations of the phosphorylation states of TCR internal chains would confirm the behaviour of the ZAP-70 locking mechanism.

3.8.1 Positive Tuning and Negative Feedback

In Section 3.7, particularly Section 3.7.1, it is demonstrated that CD8 and pMHC density parameters u and v may desensitise the cell. Now their influence is investigated with the inclusion of negative feedback. Figure 3.11 gives the stationary distributions for $\mathbb{E}_K$ and $\mathbb{E}_S$. Increases in u and v initially result in increases in $\mathbb{E}_K$ and $\mathbb{E}_S$, with u being slightly more efficacious. However, further increases in v result in a maximum in $\mathbb{E}_S$. This demonstrates that increases in v are able to maintain increased $\mathbb{E}_K$ despite also increasing $\mathbb{E}_S$. This also suggests that there is a v which produces maximum negative feedback. This hints towards

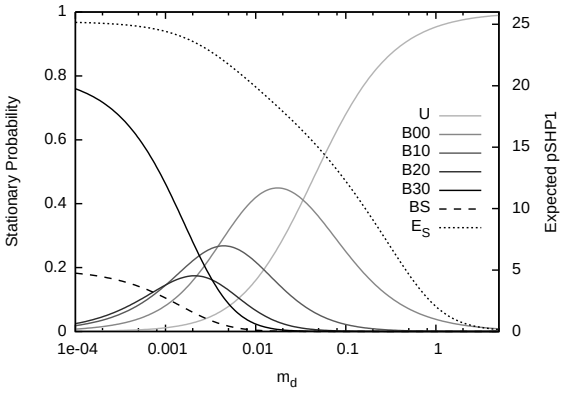

Figure 3.10: Stationary distribution for $\mathbf{N}$ with $m_d \in [10^{-4}, 5]$. Solid lines are high probability kinetic proofreading states. These are states of the form B_z^0, zero non-ZAP-70-protected phosphorylations. Dashed line "BS" is total probability of TCR signalling states. Fine dashed line $\mathbb{E}_S$ is read on right axis.

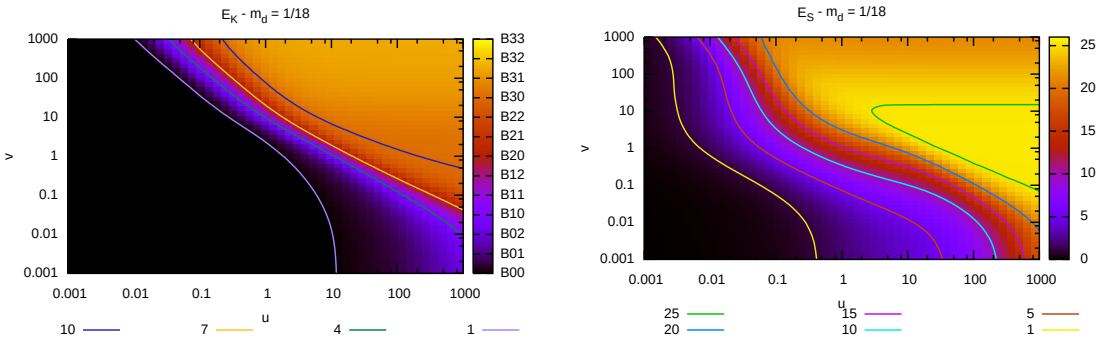

Figure 3.11: Stationary distributions for $\mathbb{E}_K$ (left) and $\mathbb{E}_S$ (right), varying $u, v \in [10^{-3}, 10^3]$, with $m_d = 1/18$. Contour lines $\mathbb{E}_K$ are $\omega(\mathrm{B}_0^0) = 1$, $\omega(\mathrm{B}_1^0) = 4$, $\omega(\mathrm{B}_2^0) = 7$ and $\omega(\mathrm{B}_3^0) = 10$. Contours for $\mathbb{E}_S$ are selected values of $|\mathrm{S}^1|$. Increases in u and v result in increases in both $\mathbb{E}_K$ and $\mathbb{E}_S$. However, there is a maximum in $\mathbb{E}_S$ in the v direction, which results in a decrease in $\mathbb{E}_S$ at high v.

the qualitative property to T cell tunability outlined in [van den Berg et al., 2007]: that a T cell is able to desensitise to a pMHC ligand μ_a whilst remaining sensitive to ligand μ_b with the dissociation rate of μ_a not necessarily faster than that of μ_b. Currently the results do not confirm or deny this behaviour; greater modelling and analysis is required.

3.9 MAPK Cascade and Negative feedback Destruction

The MAPK cascade may be initiated from a kinetic proofreading signalling state. The ABG model includes a generic adapter protein that connects the TCR to the MAPK cascade. The product of the MAPK cascade, ppERK, is able to bind to the TCR internal chain and protect it from the action of pSHP1. The reactions are as follows:

Reaction 11. *Phosphorylation of Adapter*

$$\mathrm{B}_3^q + \mathrm{A}^0 \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{B}_3^q\mathrm{A}^0 \xrightarrow{k_c} \mathrm{B}_3^q + \mathrm{A}^1 \quad q > 0. \tag{3.42}$$

These reactions do not occur for S^1 bound to B.

Reaction 12. *Dephosphorylation of Adapter*

$$\mathrm{A}^1 + \mathrm{P_A} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{A^1P_A} \xrightarrow{k_c} \mathrm{A}^0 + \mathrm{P_A}. \tag{3.43}$$

Reaction 13. *Phosphorylation of Raf1*

$$\mathrm{R}^0 + \mathrm{A}^1 \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{R^0A^1} \xrightarrow{k_c} \mathrm{R}^1 + \mathrm{A}^1. \tag{3.44}$$

Reaction 14. *Dephosphorylation of pRaf1*

$$\mathrm{R}^1 + \mathrm{P_R} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{R^1P_R} \xrightarrow{k_c} \mathrm{R}^0 + \mathrm{P_R}. \tag{3.45}$$

Reaction 15. *Phosphorylation of Mek/pMek*

$$\mathrm{K}^n + \mathrm{R}^1 \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{K^nR^1} \xrightarrow{k_c} \mathrm{K}^{n+1} + \mathrm{R}^1, \quad n = 0, 1. \tag{3.46}$$

Reaction 16. *Dephosphorylation of pMek/ppMek*

$$\mathrm{K}^n + \mathrm{P_K} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{K^nP_K} \xrightarrow{k_c} \mathrm{K}^{n-1} + \mathrm{P_K}, \quad n = 1, 2. \tag{3.47}$$

Reaction 17. *Phosphorylation of ERK/pErk*

$$\mathrm{E}^n + \mathrm{K}^2 \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{E^nK^2} \xrightarrow{k_c} \mathrm{E}^{n+1} + \mathrm{K}^2, \qquad n = 0, 1. \tag{3.48}$$

Reaction 18. *Dephosphorylation of pErk / ppERK*

$$\mathrm{E}^n + \mathrm{P_E} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{E^nP_E} \xrightarrow{k_c} \mathrm{E}^{n-1} + \mathrm{P_E}, \qquad n = 1, 2. \tag{3.49}$$

Reaction 19. *Protection of TCR by ppERK*

$$\mathcal{F} = \{\mathrm{T}, \mathrm{B}\}$$

$$\mathcal{F} + \mathrm{E}^2 \underset{e_d}{\overset{e_a}{\rightleftharpoons}} \mathcal{F}\mathrm{E}^2 \xrightarrow{e_c} \overline{\mathcal{F}} + \mathrm{E}^2. \tag{3.50}$$

The reaction does not occur for S^1 bound to $\mathcal{F}$

This instantiation of the MAPK cascade follows a familiar form modelled elsewhere [Huang and Ferrell, 1996]. It is necessary to understand its precise behaviour with the ABG model choice of parameters. The $n \times n$ rate matrix that describes the MAPK cascade with $\epsilon = N_T^{-1}$ has $n = 21004075008$ and approximately 4.3×10^{11} non-zero entries, which is beyond computational means with the current methods. Simulation offers a tractable solution. The MAPK cascade is simulated with a volume of restriction $\epsilon = 10^{-2}$; the volume of ≈ 300 TCRs. The molecule quantities are: $|\mathrm{R}| = 1000$, $|\mathrm{K}| = 4000$, $|\mathrm{E}| = 1000$, $|\mathrm{P_A}|, |\mathrm{P_R}|, |\mathrm{P_K}|, |\mathrm{P_E}| = 20$. Simulations are performed for 250 seconds and with initial $|\mathrm{A}^1| \in [1, 9]$ for 2000 repeats. A^1 is converted back to A^0 through the action of $\mathrm{P_A}$, however, there is no mechanism included to convert A^0 to A^1. Thus, the results give a

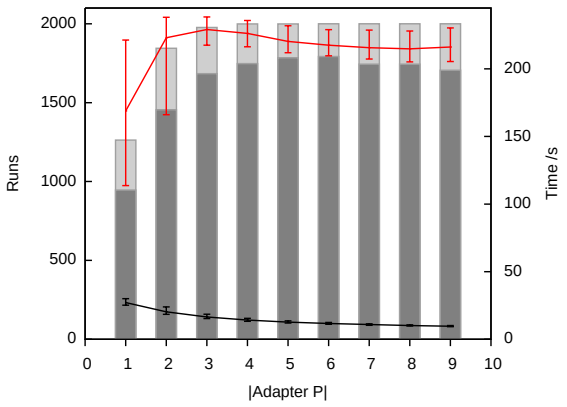

Figure 3.12: Half-maximal response of ppERK in MAPK cascade simulation results. Bars read from "Runs" axis: total bar height gives runs in which $|E^2| > 500$ occurred; dark grey lower bar gives the runs in which $|E^2| > 500$ occurred and then fell below 500 within 250 seconds. Lines and error bars are read on "Time" axis: Black line (bottom) gives the median time of first $|E^2| > 500$ occurrence; red line (top) gives the median time when $|E^2|$ fell back below 500. Error bars are the quartiles. Non-parametric statistics are given as they provide clearer representation of the distribution than parametric statistics.

lower bound to the behaviour of ppERK, as one would expect A^1 to be produced during the operation of the MAPK cascade. The results in Figure 3.12 are given in terms of half-maximal ppERK response ($|E^2| = 500$). A single A^1 molecule is sufficient for at least half the simulation runs to break the half-maximal E^2 barrier. With $|A^1| \geq 3$ the $|E^2| = 500$ barrier is broken on all runs within 25 seconds and remains broken until ≈ 220 seconds. During this period the mean $|E^2|$ is at a plateau very close to the maximum for all $|A^1|$ (not shown). With these parameters the MAPK cascade is sensitive, only requiring one or two initial A^1 molecules for long periods of near maximum ppERK signal. The results describe switch like behaviour of ppERK. This is potentially in line with the result of [Feinerman et al., 2008b] that variation in ppERK concentration has little influence on the ability of the cell to signal. If the variation is small in comparison to the "on" and "off" levels of the ppERK switch then the variation is unlikely to influence the designation of the switch. If small quantities of A^1 are sufficient to generate enough E^2 to be recognised as an activation signal, then a single TCR may be sufficient to activate the cell. Further, if the levels of E^2 are near maximal (or at least $|E^2|/|T| > 1$), as the simulations suggest then it is likely that all TCRs will receive ppERK protection. This has important implications for the spreading of the protection signal, particularly for the hypothesis discussed in [Feinerman et al., 2008a; Wylie et al., 2007] that endogenous ligands synergise with agonist ligands and contribute to the activation signal. These issues are discussed in detail in the concluding comments in Section 3.11.

Note that currently the MAPK cascade serves to guarantee a high ppERK signal if a kinetic proofreading state is reached. This behaviour of the MAPK cascade suggests that an extension to the experimental work of [Feinerman et al., 2008b] to investigate a variation in all MAPK components would be more revealing than the current investigation of ERK levels alone.

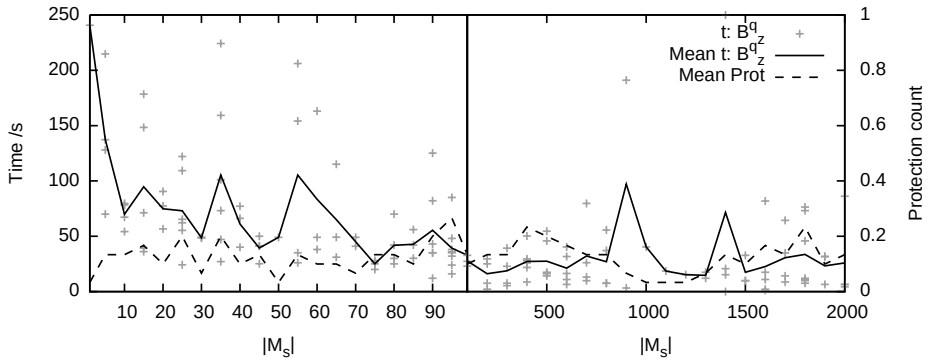

Figure 3.13: Simulations at $\epsilon = 10^{-1}$ with $|\mathrm{M_S}| \in [1, 2000]$ for 250 seconds and 20 repeats. Grey crosses are the times at which a signalling state is first reached, solid black line gives the mean of these times. Dashed black line (read on right axis) gives the number of times a protection occurred normalised by the number of runs.

3.10 The Entire system

The components analysed thus far are now combined in a single simulation, whose size is to $\epsilon = 10^{-1}$, ≈ 3000 TCRs [7].

The influence of increasing TCR stimulation to TCR activation is now directly tested. The pMHC is partitioned into: $\mathrm{M_S}$ a class of strong antigen ($m_d = 0.055$) and $\mathrm{M_E}$ a class of weak antigen ($m_d = 1.0$) with $|\mathrm{M}| = |\mathrm{M_S}| + |\mathrm{M_E}|$. $|\mathrm{M_S}|$ is varied in $[1, 2000]$ and the results are given in Figure 3.13. One or two strong ligands are as efficacious as thousands at producing a high ppERK signal in $10\% - 20\%$ of all runs. Increasing $\mathrm{M_S}$ results in quicker responses with the first signalling state being reached in the first 100 seconds.

An understanding of the activation behaviour can be gained by observing unbound pSHP1 over time. Figures 3.14 and 3.15 show 50 simulations of 250 seconds for $|\mathrm{M_S}| = 30$ and $|\mathrm{M_S}| = 3000$ respectively. The ppERK generated by a single TCR is sufficient to protect all TCRs from pSHP1 and in all cases the ppERK signal is close to maximal $\approx \epsilon N_{\mathrm{E}}$. When TCR protection occurs pSHP1 is unable to rebind the TCR resulting in a large spike in unbound pSHP1. For $|\mathrm{M_S}| = 30$, 5 activations occur throughout the first 250 seconds; for $|\mathrm{M_S}| = 300$, only 2 activations occur and they are within the first 50 seconds. The greater number of strong ligands produces a sharper increase in pSHP1. For $t > 50$ pSHP1 levels are so great that it is improbable that even a strong agonist will complete proofreading. Figure 3.15 suggests that this point occurs when $|\mathrm{S}^1| > |\mathrm{T}|$; there is at least one unbound pSHP1 molecule for every TCR. This is in agreement with Section 3.8 that a quantity of pSHP1 far less than maximal is sufficient to arrest the kinetic proofreading of a strong agonist.

Figure 3.16 plots all trajectories of the 50 runs with $|\mathrm{M_S}| = 3000$ whose unbound pSHP1

[7] In 1000 simulations with $\epsilon = N_T^{-1}$ with $m_d = 0.055$ no runs achieve a TCR signalling state due to the increased influence of stochastic fluctuations. However, qualitative similarities are observed between $\epsilon = N_T^{-1}$ and $\epsilon = 10^{-1}$ in Figures 3.19 and 3.16 respectively.

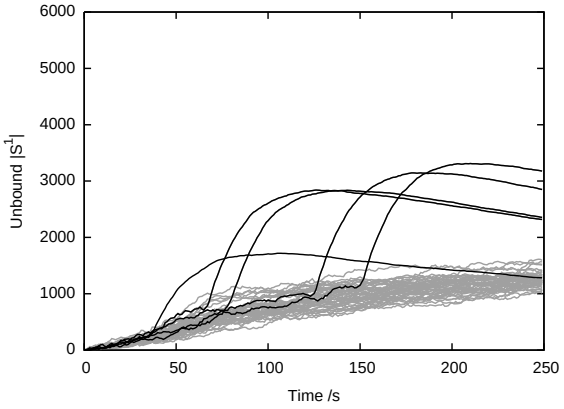

Figure 3.14: Simulation with $|\mathrm{M_S}| = 30$ for 250 seconds with 50 repeats plotting unbound pSHP1. Grey lines are simulations where no signalling state and no ppERK signal are achieved. Black lines are simulations where a ppERK protection of all TCR is achieved.

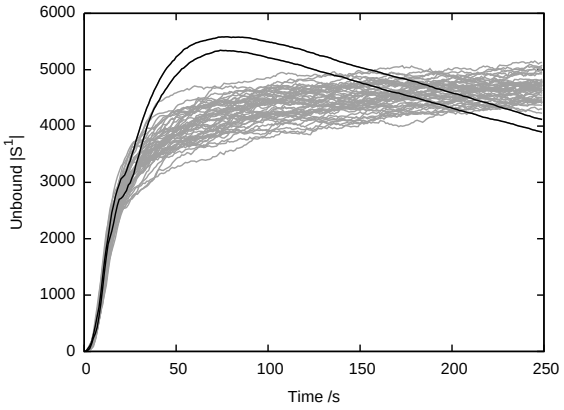

Figure 3.15: Simulation with $|\mathrm{M_S}| = 3000$ for 250 seconds with 50 repeats plotting unbound pSHP1. Grey lines are simulations where no signalling state and no ppERK signal are achieved. Black lines are simulations where a ppERK protection of all TCR is achieved. Two activations occur in the first 50 seconds, after this time unbound pSHP1 levels are sufficient that it is improbable that a TCR will reach a signalling state. These results are also plotted as a histogram plot in Figure 3.16.

is given in Figure 3.15. In agreement with small-ϵ Figures 3.9, 3.19 there is a greater occurrence in late kinetic proofreading states early in the simulation runs ($t < 50$). Also, note that simulation runs are generally confined to zero non-ZAP-70-protected phosphorylation states (B_z^0) further agreeing with small-ϵ stationary distributions given in Figure 3.10.

The agonist and antagonist roles of ligands are now compared. Simulations are performed varying ligand quality $m_d^{-1} \in \{1, 3, 6, 12, 16, 18\}$ for 250 seconds and 50 repeats. Agonist tests are performed with 30 pMHC at the given m_d, the other $|\mathrm{M}| - 30$ pMHC have $m_d = 1$. Antagonist tests are performed with 30 pMHC with $m_d = 1/18$ and then other $|\mathrm{M}| - 30$ pMHC are at given m_d. The results are summarised in Table 3.3.

The antagonist tests show higher number of activations. Low quality ligands generate little pSHP1 (Figure 3.10) allowing good ligands to signal uninhibited. This is unrelated to the synergy of [Feinerman et al., 2008a], the very low quality ligands are "synergising" by not inhibiting the high quality ligands. As pMHC ligand quality increases levels of pSHP1 increase resulting in antagonism. Activations are seen with further increases in ligand quality, all pMHC are now good agonists and have a higher probability of completing kinetic proofreading (Figure 3.10).

An explanation for poor quality ligands completing kinetic proofreading is due to initial $|\mathrm{S}^1| = 0$. Cells are expected to exhibit pSHP1 levels based on their recent interactions [Altan-Bonnet and Germain, 2005] [8] and particularly post-thymic T cells exhibit signifi-

[8] The rate at which pSHP1 is converted back to SHP1 is perhaps too great here, pSHP1 decays on the

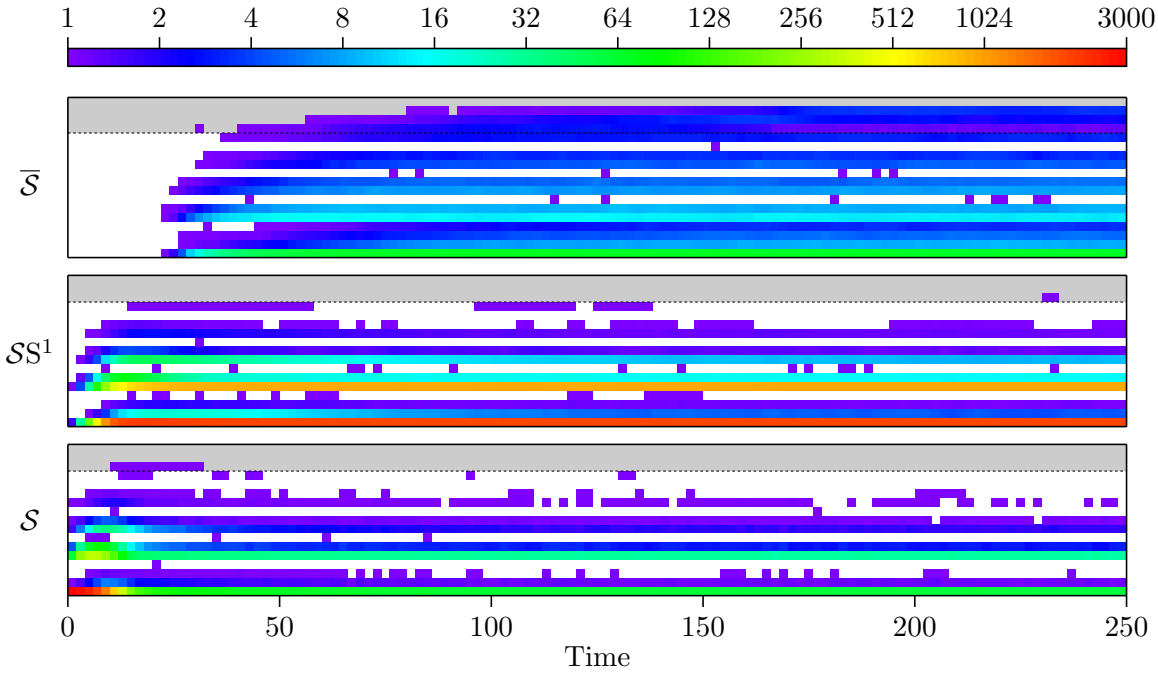

Figure 3.16: Histogram plot showing the occupation states $\mathcal{S}$, $\mathcal{S}S^1$ and $\overline{\mathcal{S}}$ in time (see Section A.3 for definition of $\mathcal{S}$ and advice on reading this plot), for simulations given in Figure 3.15, $|M_S| = 3000$ for 250 seconds with 50 repeats. Two activations occur, they can be seen early within the lower $\mathcal{S}$ section. Note the increased occupancy of the later kinetic proofreading states for $t < 50$ before the rise of pSHP1 has occurred. This is in agreement with small-ϵ results in Figures 3.9, 3.19.

m_d^{-1}	Agonist	Antagonist
1	0	5
3	1	1
6	1	0
12	3	4
16	2	4
18	5	2

Table 3.3: Agonist and antagonist results. Second column gives simulation runs in which a high ppERK signal occurs in the agonist tests; third column gives simulation runs in which a high ppERK signal occurs for the antagonist tests. There is an overlap between agonist and antagonists tests, simulations with 30 pMHC with $m_d = 1/18$ and $|M| - 30$ pMHC with $m_d = 1$ are performed only once and the results used for agonist and antagonist cases.

Initial $\lvert S^1 \rvert$	Activations	Within Time /s
0	7	100
2000	6	100
3000	4	3 in 100, 1 in 900
6000	0	–
12000	1	900
8×10^5	0	–

Table 3.4: Simulations increasing initial $\lvert S^1 \rvert$. The second column gives simulation runs in which a high ppERK signal occurs; the third column gives an approximate timescale in which all ppERK signals occur. The $\lvert S^1 \rvert = 0$ simulations are not repeated and the presented value is the result given in Figure 3.13, it should be noted that this result involves simulations of 250 seconds not 1000.

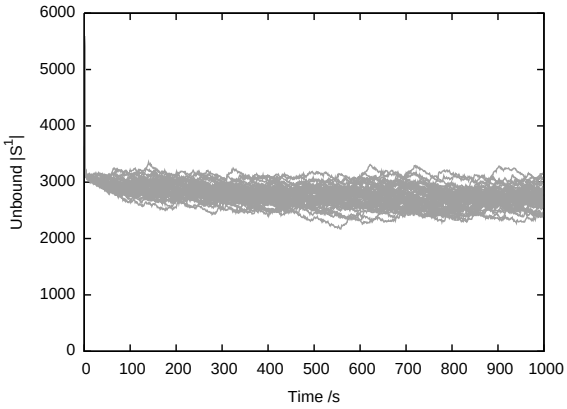

Figure 3.17: Simulations of $\lvert M_S \rvert = 400$, $\lvert M_E \rvert = 2600$, with initial $\lvert S^1 \rvert = 6000$ for 1000 seconds with 50 repeats. Grey lines are simulation runs in which no ppERK signal occurs. There is an immediate drop in unbound pSHP1 by $\approx \lvert T \rvert$ as it binds to all TCRs. No simulation runs achieve ppERK protection.

cantly increased pSHP1 [Germain and Stefanová, 1999]. Simulations are performed varying initial pSHP1 for 1000 seconds with 50 repeats, $\lvert M_S \rvert = 400$ agonists ligands ($m_d = 0.05$) and $\lvert M_E \rvert = 2600$ weak pMHC ligands ($m_d = 1.0$). The results are summarised in Table 3.4.

Increasing initial pSHP1 is able to reduce the probability of an activation signal for even strong agonist pMHC. However, when observed over a long enough timescale (900+ seconds) the initial pSHP1 decays to the point where stochastic fluctuations allow good agonists to produce activations. Figure 3.17 shows the unbound pSHP1 for initial $\lvert S^1 \rvert = 6000$.

The result given in Figure 3.11 is examined, that increases in CD8 density produce a maximum in $\mathbb{E}_S$ but not in $\mathbb{E}_K$. That is, that increasing the CD8 density is able to overcome increases in pSHP1. Figure 3.17 demonstrates zero activations in fifty with $\lvert S^1 \rvert = 6000$ and $v = 1$. Simulations are performed with $v = 2,5$ for 1000 seconds with 50 repeats, other simulation parameters are as Figure 3.17. The unbound pSHP1 for $v = 2,5$ is given in Figure 3.18. Despite initial pSHP1, increases in v is able to induce activations. Agreeing with Figure 3.11, increases in v result in greater pSHP1 but also an increased probability

minutes rather than hours timescale, this will only mean that the affect of pSHP1 could be underestimated here.

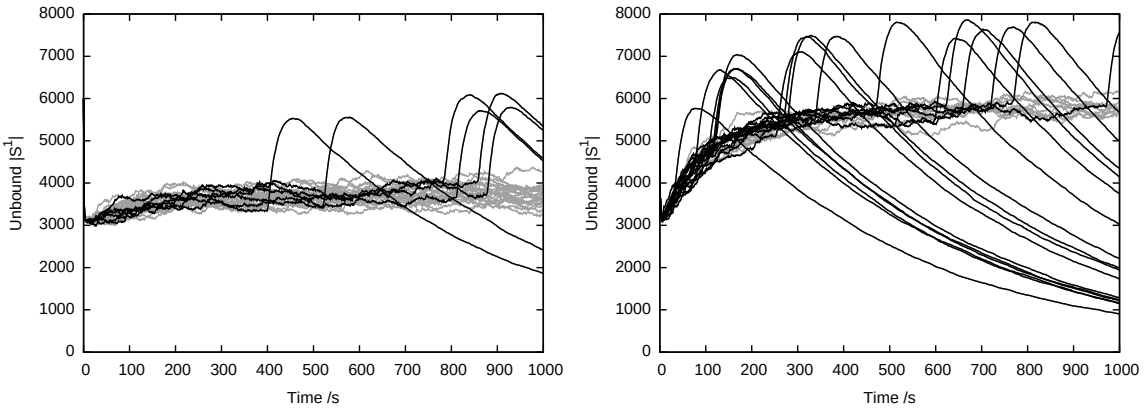

Figure 3.18: Simulations with $|M_S| = 400$; $|M_E| = 2600$; initial $|S^1| = 6000$; for with $v = 2$ ($|C| = 6000$) (left) and $v = 5$ ($|C| = 15000$) (right); for 1000 seconds with 25 repeats. Grey lines are simulation runs in which no ppERK signal occurs, black lines are simulation runs in which kinetic proofreading is successful. $v = 2$ results in 6 activations; $v = 5$ results in 16 activations, the most reliable activation seen thus far.

of reaching later kinetic proofreading states and so increased activations.

3.11 Discussion

This section discusses the findings of this chapter and culminates in a hypothesis for T cell activation in Section 3.11.2. A justification for the use of volume restriction ϵ is also discussed in Section 3.11.1.

A biologically detailed model originally presented in [Altan-Bonnet and Germain, 2005] which concerns TCR signalling with respect to the behaviour of cytosolic SHP1 and ERK has been analysed. This signalling system has received much modelling attention in [Chan et al., 2001; Wylie et al., 2007; Artyomov et al., 2007; Lipniacki et al., 2008; Feinerman et al., 2008b] who have often focused on the macroscopic or general qualitative behaviour. However, rather than make simplifying assumptions (such as those in [Artyomov et al., 2007; Lipniacki et al., 2008]) the analysis here has opted to retain the biological detail, to discover how this detail maps to the identified general qualitative cell behaviours. This has been achieved by dissecting, analysing and then re-composing the key components of the model. In contrast to the original presentation of the ABG model [Altan-Bonnet and Germain, 2005], the analysis has been stochastic. The stochastic approach looks quite appropriate given the small molecule numbers required to initiate the MAPK cascade (Section 3.9) and consequently the small numbers of signalling TCRs required to generate a high ppERK activation signal (Section 3.10). Further, [Artyomov et al., 2007; Lipniacki et al., 2008] have clearly highlighted that stochastic effects leading to bistabilities are of importance in TCR signalling models such as the one here. Certainly, some reactions considered fulfil the sufficiency conditions (branching, irreversibility and feedback) for a stochastic bistability given in [Artyomov et al., 2007]. It should be noted that the larger number of reactions in this model complicates the issue. For example, the switch like behaviour and large molecule

numbers of the MAPK cascade provides effective irreversibility. When observing the global cell response to the MAPK cascade, the deterministic approach could give reasonable results. However, the stochastic bistability conditions are effectively fulfilled by the reactions that initialise the MAPK cascade (Reaction 11). The reactions are reversible, but the forward rate is an order of magnitude greater than the reverse rate ($N_{\mathcal{P}} \ll N_A$). The feedback in the positive direction arises from many "cascading" steps of the MAPK process (see the well known depiction of reactions in [Huang and Ferrell, 1996; Altan-Bonnet and Germain, 2005]), there is no negative feedback. Consequently small initial molecule numbers of A^1 fulfil the bistability conditions. Further, with no negative feedback this stochastic switch is weighted toward the "on" position. The result is a stochastic MAPK cascade that is far more sensitive than the deterministic MAPK cascade (see Figure 3.12 compared in Figure S8A in [Altan-Bonnet and Germain, 2005]). As such, a smaller activation signal from the TCRs is required to produce a cell activation; the activations due to a single TCR in Section 3.10 are confirmation.

The results are in agreement with the qualitative behaviours discussed in [Chan et al., 2001; Altan-Bonnet and Germain, 2005; Lipniacki et al., 2008] as outlined in Section 3.4. Due to the approach, direct results comparisons are difficult as the focus is on differing levels of detail to the models in [Chan et al., 2001; Altan-Bonnet and Germain, 2005; Lipniacki et al., 2008]. However, the results are perhaps most readily comparable with those in [Altan-Bonnet and Germain, 2005] as the underlying reactions equations are identical. Certainly, there is agreement with the three predictions of the model given in [Altan-Bonnet and Germain, 2005]: lengthening of ppERK response time at low ligand densities (see Figures 3.12, 3.13, 3.14); hierarchy of antagonism, that is, superior sub-agonist ligands produce greater negative feedback (see Figure 3.10); flexibility in ligand discrimination undergoing differentiation, that is, the ability of a T cell to tune its response (see analysis involving u, v and Figures 3.17, 3.18). Here the analysis has refrained from stochastically running the numerical experiments carried out in [Altan-Bonnet and Germain, 2005] and comparing the quantitative results. This task has been performed on a simpler model in [Lipniacki et al., 2008]. Any quantitative differences are expected to arise at least where the bistability conditions apply. As discussed in the previous paragraph, this includes the sensitivity of the MAPK cascade and by a similar argument the strength of the pSHP1 negative feedback.

The key findings of the chapter are now summarised. Section 3.6 demonstrates that rate choices should be made in unison with model structure choices. This can reveal equivalently behaved simpler models and most importantly this provides an opportunity to validate against biology. What remains is to return to biology to discover the occurrence of the CD8-TCR complex and adjust the model accordingly. A solution could be a more complex model of CD8, TCR, pMHC binding similar to that in [van den Berg et al., 2007]. Section 3.7 investigates the ABG formulation of kinetic proofreading and the expected step-like discrimination is observed. The phosphorylation protection of ZAP-70 is shown to be of importance by aiding pMHC and CD8 density parameters u and v to sensitise the cell. This ability is a consequence of the relative stability of the ZAP-70 association

which allows ZAP-70 molecules to remain bound between TCR pMHC associations. Thus, the kinetic proofreading state may be preserved between TCR pMHC binding events. It is demonstrated that decreases in d_z can cause retention of kinetic proofreading state despite decreases in CD8 density v. The influence of this locking mechanism is largely undiscussed in [Altan-Bonnet and Germain, 2005].

The analysis of the negative feedback behaviour in Section 3.8 demonstrates a transient hump in expected kinetic proofreading state before a rise in expected pSHP1 level. This is due to a non-zero kinetic proofreading state which must be reached before the negative feedback can be generated; this is determined the *base negative feedback state*. Examining the response to changes in pMHC and CD8 density parameters shows that u and v are able to increase expected proofreading state and pSHP1 levels. However, there is a maximum in $\mathbb{E}_S$ in the v direction, as discussed in Section 3.8; this is an appropriate step towards the necessary condition for full tunability given in [van den Berg et al., 2007].

In Section 3.9 the MAPK cascade is shown to act as an ultra-sensitive amplification switch, amplifying small numbers of A^1 molecules to thousands of ppERK molecules. The ppERK signal is near maximal and remains so for long periods of time; with sustained A^1 production ppERK would remain near maximal permanently. As discussed in Section 3.9 this raises issues for the determination of the activation signal and the spreading of the protection signal. First, if a near maximal ppERK whose duration is > 200 seconds is sufficient for a cell to recognise an activation then a single TCR may be sufficient to activate the cell. Second, since numbers of near maximal ppERK are an order of magnitude greater than the number of TCRs then total TCR protection is expected (as is the case in the activations in Section 3.10) [9]. This has implications for the hypothesis presented in [Feinerman et al., 2008a] that suggests that protected endogenous ligands may contribute towards signalling. The issue is that these ligands must still overcome kinetic proofreading. As discussed in Section 3.4, consideration of the structure of the ppERK reactions reveal that it is not involved in a positive feedback and so ppERK protection does not aid the ability of any ligand to overcome kinetic proofreading. The discussion is eased by the following classification of ligands based on the induced signalling:

- *Class I:* Are able to complete kinetic proofreading despite negative feedback and generate an activation signal. Effectively agonists.
- *Class II:* Are unable to complete kinetic proofreading in the presence of negative feedback [10], but are able to complete kinetic proofreading if protected from negative feedback. Effectively antagonists.

[9] The TCR protection state is absorbing in this model and so all TCRs will receive protection with probability 1 over a suitable timescale. However, protection is expected to persist on a timescale longer than the signalling events considered here, if a loss of protection reaction were included very high levels of TCR protection would be observed, but not total protection.

[10] For the stochastic models considered here all ligands will be able to complete kinetic proofreading, with or without negative feedback, when considered over a long enough time period. When an inability to complete kinetic proofreading is discussed the meaning is that the probability a ligand will complete kinetic proofreading in a time period that is relevant to the cell is effectively nil.

- *Class III:* Are unable to complete kinetic proofreading even when protected. They are strong enough to achieve a kinetic proofreading state larger than the base negative feedback state, that is, they are able to generate negative feedback. Effectively antagonists/partial agonists.

- *Class IV:* Are unable to complete kinetic proofreading when protected. They are also unable to break the base negative feedback state. Effectively partial agonists/null ligands.

The agonist, antagonist etc. labels have not been used because the definitions of the classes differ from those given in Section 3.4. Within these classes only *I & II* may "synergise" with the signalling agonists, *III & IV* will never complete kinetic proofreading. Generally endogenous ligands are expected to be members of classes *II–IV*. A common hypothesis is that endogenous ligands are most often not antagonists [Altan-Bonnet and Germain, 2005; Feinerman et al., 2008a] and non-antagonist/non-agonist ligands are members of classes *III & IV*. Thus, the majority of endogenous ligands will not be able to synergise due to their inability to overcome proofreading. The expected hitting time of protected kinetic proofreading signalling state (Figure 3.6) shows a sharp increase with increases in m_d. This suggests that whilst members of class *II* can complete kinetic proofreading on a suitable timescale they may take substantially longer than members of class *I* and consequently generate substantially less A^1. Then, it is possible that a handful of class *I* ligands can generate far more A^1 than a population of class *II* ligands. The discussions of the behaviour of the MAPK cascade are concluded with two statements. First, if a single TCR is able to generate a near maximal ppERK signal then synergising ligands can only contribute by ensuring the ppERK signal remains near maximal over a longer period of time than the original ligand could achieve. Second, if the TCRs receive total protection this is not necessarily an indicator that all TCRs require protection, simply that this guarantees that the TCRs of interest are protected.

Section 3.10 re-composes the entire model to perform stochastic simulations on a system of many TCRs. The simulations demonstrate that a few agonists can be as efficacious as a population of thousands in producing activations. Moreover, the large-ϵ results are found to be in qualitative agreement with the small-ϵ single TCR predictions. The justification for ϵ is discussed in Section 3.11.1.

It is elucidating to observe the dynamics of unbound pSHP1. The pSHP1 generated is dependent on TCR stimulation, which is dependent on the quantity of pMHC and distribution of dissociation rates. As a consequence, simulations with zero initial pSHP1 consisting solely of weak ligands (normally considered to be members of class *II*) generate little pSHP1 and the weak ligands can induce activations. Increased TCR stimulation results in a sharper rise of pSHP1, the statistics of the sharp rise are related to the position of the base negative feedback state. Once the rise has occurred it becomes very unlikely that any TCR, even a good agonist, will complete kinetic proofreading. Performing simulations with ligands in agonist and antagonist roles shows the expected agonist and antagonist behaviour. Increases in initial pSHP1 are able to prevent weak ligands from inducing activations, with further

increases preventing agonist pMHC from inducing any activations. Finally, the small-ϵ v prediction is validated, that increased CD8 density is able to increase the expected kinetic proofreading state despite increasing the expected pSHP1.

3.11.1 Justification for ϵ

Three choices for ϵ are used in this chapter, as discussed in the Chapter 2 there is no guarantee that results calculated for an ϵ restricted volume will bare any resemblance to the true system. As a consequence, the small-ϵ results of this chapter have only been used to gain an intuition and understanding of the larger system, but then the behaviour must be confirmed in the larger system (this is the objective of any modelling, to make simplifying assumptions to gain understanding of a bigger system). The small-ϵ results guided the chosen analysis in Section 3.10, and as will now be discussed the small-ϵ intuition is in agreement with the large-ϵ results. In this sense the use of ϵ has been invaluable in this work.

In Sections 3.6, 3.7 and 3.8 the volume of a T cell is scaled to approximately the volume of a single TCR, so $\epsilon = N_T^{-1}$ with N_T the number of TCRs in the ABG model. There is qualitative agreement between small and large ϵ (see Figures 3.16 and 3.19). All small-ϵ calculations in Sections 3.6-3.8 are made solely with rate matrices and unlike simulation do not require many runs to view the entire distribution of behaviour. Assuming that the increased stochastic noise at $\epsilon = N_T^{-1}$ is suitably distributed then the transient and stationary expectations may be minimally affected. Rate matrix $\mathbf{N}$ is the most susceptible to noise as it analyses varying cytosolic molecule quantities. An increase of a single pSHP1 molecule at the single TCR level corresponds to a large increase in pSHP1 in the entire cell. However, the results matched well with results obtained with a real pSHP1 concentration parameter defined similarly to v defined in Section 3.6 (not shown).

In Section 3.9, $\epsilon = 10^{-2}$ is used for simulations of the MAPK cascade and in Section 3.10 $\epsilon = 10^{-1}$ for simulations of the entire system. The molecule numbers in these simulations were found to be sufficient that stochastic fluctuations do not dominate and results were found to be in good agreement with other large ϵ choices, including $\epsilon = 1$.

3.11.2 A Hypothesis for T cell Activation

Combining the results of the chapter a hypothesis for reliable T cell activation with respect to the base negative feedback state is now discussed. If the T cell is presented with a population comprising $0.01 - 0.1\%$ agonists and $99.9 - 99.99\%$ endogenous ligands and if the expected time for agonists to complete kinetic proofreading is less than the expected time for endogenous ligands to reach the base negative feedback state, then the T cell will be reliably activated. This hypothesis is depicted in Figure 3.20 along with validating results from the model. The results state that agonist ligands with $m_d = \mu_a$ are expected to reach a kinetic proofreading signalling state before endogenous ligands with $m_d = \mu_e$ reach the base negative feedback state. The expected times given in Figure 3.20 are calculated from the kinetic proofreading without negative feedback rate matrix $\mathbf{K}$ described in Section 3.7.

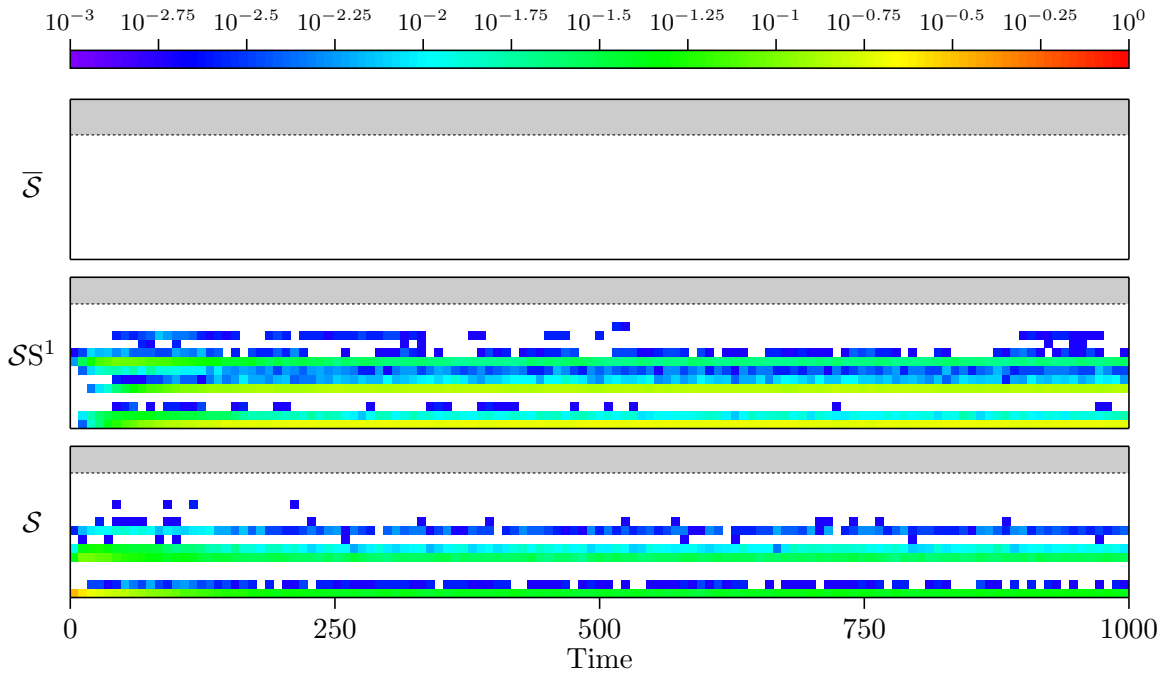

Figure 3.19: Histogram plot of states $\mathcal{S}$, $\mathcal{S}\mathrm{S}^1$ and $\overline{\mathcal{S}}$ for 1000 simulations restricted to a single TCR with $m_d = 0.054$ for 1000 seconds (see Section A.3 for definition of $\mathcal{S}$ and advice on reading this figure). No simulations reach the grey signalling region. Note the increased occupancy of later kinetic proofreading states within the first 500 seconds during the rise of pSHP1. This is in qualitative agreement with Figure 3.16.

As justification for calculating without negative feedback, it is noted that during the period in which agonist ligands generate negative feedback the endogenous ligands do not generate pSHP1. Since agonist ligands are in the vast minority and appealing to the assumption that pSHP1 is not diffusion-limited then the population of endogenous ligands "absorb" the pSHP1 generated by the agonist ligands. Further, this assumes that the endogenous ligands comprise a sufficiently large proportion of all pMHC, that the influence of agonist pSHP1 does not significantly alter their time to base negative feedback state given in Figure 3.20.

This hypothesis requires more detailed modelling and analysis to confirm; in reality, the negative feedback generated by the agonists will dampen their own proofreading as well that of the endogenous ligands. However, the arguments given here certainly demonstrate the need for further investigation.

This hypothesis suggests that a reason why the T cell can react to only 0.01–0.1% of the total pMHC expressed [Germain and Stefanová, 1999] is because this is a very effective way of producing T cell activation. Altan-Bonnet and Germain [Altan-Bonnet and Germain, 2005] suggest that the agonist ligands can quickly overcome the negative feedback due to its "limited nature". The results in this chapter suggest that the negative feedback is not limited in nature and is capable of arresting the proofreading of any realistic ligand. A clarification to their argument is given by adding that agonist ligands can induce activations by their ability to complete proofreading before weaker ligands can generate negative feedback.

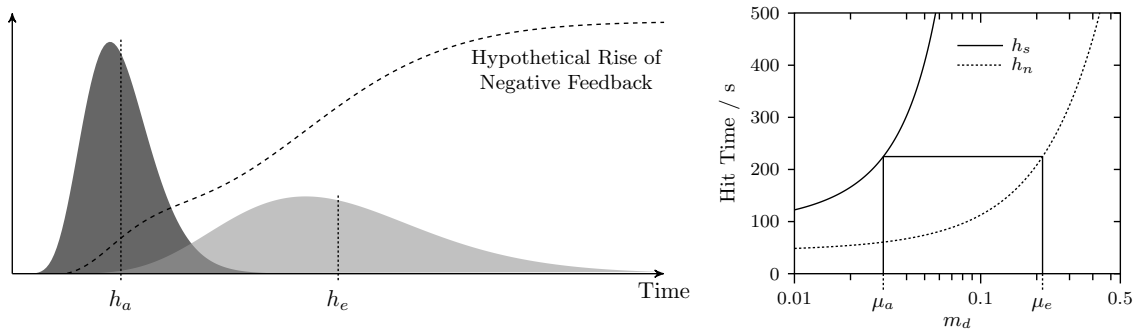

Figure 3.20: *Left*: Depiction of reliable activation hypothesis. The hypothetical distribution of agonist ($m_d = \mu_a$) ligands reaching signalling state is given on the left (dark grey), h_a is the expected hit time of signalling state. Hypothetical distribution of endogenous ligands ($m_d = \mu_e$) reaching base negative feedback state is given on the left (light grey), h_e is expected hit time of the base state. The distributions are purely illustrative and are sum of exponential distributions, specifically Erlang Distributions (Gamma distribution with integer shape parameter) and as such the mean is greater than the mode. Dotted line gives the hypothetical rise of negative feedback, again illustrative and calculated as a function of cumulative hit time distributions, expected agonist/endogenous pSHP1 levels and assumed agonist/endogenous populations ratio. *Right:* Validation of hypothesis within the model. Solid black line (h_s) is expected hit time of signalling state for agonist ligands ($m_d = \mu_a$) (taken from Figure 3.6), dotted line (h_n) is the expected hit time of base negative feedback state for endogenous ligands ($m_d = \mu_e$). All ligands with $m_d \leq \mu_a$ are expected to reach signalling state before ligands with $m_d > \mu_e$ reach base negative feedback state.

Note that this hypothesis does not prevent a population of agonists inducing an activation for two reasons. First, the larger population of agonists have a higher probability of sampling a proofreading time substantially faster than the mean. Second, the negative feedback parameters currently allow agonist ligands to complete kinetic proofreading despite high pSHP1 levels via the stochastic fluctuations that occur over an increased time period (see Table 3.4).

Biological experimentation which could test the results of the chapter is now suggested. If APCs can be prepared which present endogenous ligands in the majority and agonist ligands in the minority and if the concentration of agonist ligands can be incrementally increased, then the statistics of the rise time of pSHP1 and ppERK could confirm the details of the hypothesis in Figure 3.20. Further, the analysis suggests that the pSHP1 levels needed to dampen a strong agonist are an order of magnitude smaller than the total SHP1. Consequently a variation in pSHP1 less than an order of magnitude will not necessarily inhibit a cell in producing the required pSHP1 to dampen strong agonists. However, a reduction in SHP1 levels will reduce the rate at which pSHP1 can be produced and so increase the time taken to reach equilibrium. This should extend the time window in which activations can reliably occur and so cell activations will be observed over a longer time period. One could also look for the dynamics of unbound pSHP1 which would suggest its relationship to the protected TCR and particularly the spreading of the protection signal. Any investigations into the time taken for kinetic proofreading of varying strength ligands, particularly looking for the existence and position of a *base negative feedback state* would

prove very useful in validating the hypothesis presented in Figure 3.20.

This chapter has analysed a biological model for the purpose of modelling biology rather than the purpose of the construction of a bio-inspired algorithm. The approach has, hopefully, avoided biases that may have occurred if algorithmic development was in mind. In the next chapter, the detail of this model is greatly simplified in an abstracted model of T cell signalling. The abstracted model is then used to develop bio-inspired algorithm, in particular an anomaly detection system. The majority of the detail of this chapter is not carried through to the next. Apart from the contribution to biology, the benefit of the detail of this chapter is that it culminated in the negative feedback base state hypothesis. This feature will be pivotal in the behaviour of the algorithm in the next chapter.

CHAPTER 4

The Receptor Density Algorithm

This chapter takes abstractions from the previous chapter to construct an abstracted model of T cell signalling. This abstracted model in conjunction with inspiration and grounding in conventional machine learning approaches [Bishop, 1994] is used to develop an anomaly detection algorithm: The *Receptor Density Algorithm*. The algorithm shares structure in common with the T cell, in particular, with the internal component of the T cell receptor. The grounding with the machine learning techniques provides some justification in approach and understanding of the algorithm's anomaly detection mechanisms.

4.1 Biological Inspiration and Abstraction

The T cell signalling system is a suitable candidate for a bio-inspired algorithm, particularly due to the properties of T cell discrimination:

- The T cell is able to perform fine-grain[1] 2-class discrimination, where the second class is *unseen*, that is, the activating agonist pMHC is unknown a priori.
- The discrimination between the two classes is dependent on a history and based on the relative differences of examples from each class. That is, a new pMHC ligand will fail to activate a T cell if it has a similar TCR-pMHC bond lifetime to recently presented pMHC.
- The system fulfils the advice of Chapter 1. The structure of the system looks more important than its specifics, kinetic proofreading is example. There is the undesired molecular recognition in the TCR-pMHC bind, however the fact that this is not a lock-and-key molecular recognition and is combined with signalling processes to time this bind gives good potential for an algorithm.

In machine learning terminology, the above describes an *anomaly detection* system. Anomaly detection problems are two class classification problems, the system must decide whether samples (in this case pMHC) are members of *class 1* or *class 2*. The distinguishing feature

[1] i.e. small difference between agonist and antagonist pMHC

of anomaly detection problems is that only the details of *class 1* are known, and so, at the time of classification there is no description of *class 2*. Moreover, the anomaly detection system of the T cell is *on-line* in that instances of *class 1* are continually presented potentially altering the discrimination boundaries. As noted in Chapter 1, this is effectively "reasoning-by-metaphor". There looks to be a mapping to anomaly detection techniques, but it is necessary to become involved in specifics to confirm this suggestion. Thus, the T cell signalling system has properties to be a good candidate for a bio-inspired algorithm, and the *approximate application* (see Chapter 1) is anomaly detection. The first step is to create an abstracted model from the biological model. A recap of key biological concepts relevant to understanding the subsequent abstracted model (Section 4.1.1) is now given.

The hypothesis presented at the end of the last chapter found the *base negative feedback state* key in understanding T cell activation. This negative feedback barrier and other processes associated with the internal component of the TCR have the ability to evaluate the strength of the TCR-pMHC bond. Only a subset of the processes related to TCR signalling are utilised in this chapter, in summary:

- *Kinetic Proofreading.* These are the steps that must be completed before a TCR can generate an activation signal, the result is a step like discrimination between ligands which are able to complete kinetic proofreading and those that are not.

- *Negative Feedback & Base State.* Once the kinetic proofreading process is equal to or greater than the base negative feedback state, a negative feedback signal is generated. This signal is generated at a rate independent of the displacement above the base state, and so is not instantaneous, it takes time to build. The signal has the ability to inhibit and reverse the kinetic proofreading process.

- *Serial Triggering & Stimulation Spreading.* A single pMHC molecule may dissociate and re-associate to the same or a different TCR. Whilst this was not directly modelled in the last chapter, the constraints of the T cell surface increase the probability of re-association to the same or nearby TCRs.

- *Negative Feedback Spreading.* The negative feedback signal generated by a TCR will spread to dampen the surrounding TCRs.

4.1.1 The Generalised Receptor

This section presents a computationally relevant abstracted model of a T cell receptor. To avoid confusion, the acronym TCR refers to the biological entity and the word *receptor* refers to the TCR's computational abstraction:

Definition 4.1. *A receptor is a tuple* (p, n, β, ℓ, c) *with* $p, n, \beta, \ell \in \mathbb{R}$. *That is, a receptor has a position* $p \geq 0$*; a negative feedback* n*; a negative feedback barrier* $\beta > 0$ *and a length* $\ell > \beta$. *The state* $c \in C$ *represents the classification or "activation" state of a receptor, where* C *is the set of activation states of a receptor.*

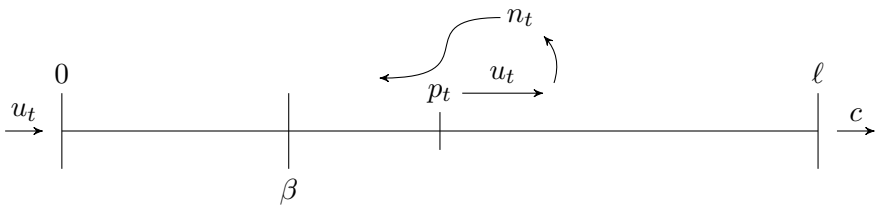

Figure 4.1: The receptor receives an input u_t pushes the receptor position p_{t+1} toward ℓ. The receptor will generate negative feedback if $p_t \geq \beta$. Should, $p_{t+1} \geq \ell$ then the receptor will successfully signal with $c = 1$.

The definition of a receptor is an abstraction of the internal component of the TCR. The mapping between the biology and the computational description is relatively obvious, but for clarity: p represents kinetic proofreading state (B_z^q in the terminology of the last chapter); n represents the pSHP1 level near the TCR; β is the base negative feedback state (B_1^0)[2]; ℓ is the maximum kinetic proofreading state (B_3^3); and c maps to the notion of TCR protection.

The following describes the behaviour of a receptor with the language of its definition, it is also depicted in Figure 4.1. A receptor receives a sequence of inputs $\{u_t\}$ with $u_t \in \mathbb{R}$ at discrete time points $t = 0, 1, 2, \ldots$ and $u_t \geq 0$. Then, a function f_t updates the receptor position and negative feedback as a function of u_t:

$$f_t : (p_t, n_t, u_t, \beta) \rightarrow (p_{t+1}, n_{t+1}). \tag{4.1}$$

If the receptor position is above the negative feedback barrier $p_t \geq \beta$, then negative feedback is generated. If the receptor position reaches or exceeds the receptor length ℓ, then a receptor activation or classification occurs. In this case, $c \in C = \{0, 1\}$ and $c = 1$ is $p_t \geq \ell$ and $c = 0$ otherwise. The receptor position and negative feedback will also decay and under repeated input $u_t = 0$ they will return to $p_t, n_t = 0$. A biological restriction is incorporated here: the decay rate of the receptor position is faster than that of the negative feedback. With this description of a receptor one could specify the update function f_t by application of recurrent equations as follows:

$$\begin{aligned} p_{t+1} &= bp_t + u_t - an_t \\ n_{t+1} &= \begin{cases} dn_t & \text{if } p_t < \beta \\ dn_t + g & \text{if } p_t \geq \beta \end{cases}, \end{aligned} \tag{4.2}$$

with $0 < b, d < 1$ the respective receptor position and negative feedback decay rates, and $b < d$; $a > 0$ controls the negative feedback influence; $g > 0$ is the negative feedback growth rate.

This definition of a receptor places much detail in assumption set A (Chapter 1). In particular, time is discrete and the kinetic proofreading state and negative feedback is

[2]Due to the continuous behaviour of the receptor position β is referred to as the base negative feedback *barrier* rather than state

continuous. Further, the model is now deterministic and not stochastic, though input u_t may still be stochastic. As a consequence, the action of the negative feedback on the kinetic proofreading locking mechanism is modelled as linear influence. Moreover, much detail has been discarded, such as the CD8 co-receptor, the MAPK cascade and the protection of TCRs for the pSHP1. The consequences of this will be discussed in the next section where a replication of biological behaviour is performed.

The above dynamics and slight variants are the main topic of this chapter and later will form the basis of an anomaly detection algorithm in Section 4.6. The behaviour of recurrences (4.2) is analysed later in Section 4.2 and again in Section 4.5. Before the analysis, a discussion of how many receptors may interact is given.

4.1.2 The Difficulty of Many Receptors

The model in the last chapter had well-mixed assumptions, which allowed the neglection of space. If the abstracted model followed suit, then one would collect many receptors together with a single uniting negative feedback. A collection of u_t for each t would be required as input (enough to stimulate each receptor), the receptors could be randomly matched to an input and all positions and global negative feedback updated accordingly. If this abstracted model behaved as the hypothesis of the last chapter suggested, then the resulting algorithm would detect outliers in the magnitudes of the u_t, but only in one direction (i.e. only significantly larger u_t than the mean would be detected). This does not seem to be of much computational value, conventional computers have many more direct ways of solving this problem. There are two factors at work here. First, the molecular-recognition of the TCR is missing, in biology this maps the information stored in the conformation and chemical composition of the peptide (and MHC) into information interpretable by the cell, i.e. *time*. Second, the biological substrate must work hard to perform comparisons between many different things, in this case the things are pMHC dissociation rates. Whereas, conventional computers find this sort of comparison relatively easy. The possibility that the biological substrate is performing comparisons has interesting implications. These are discussed in more detail in Section 6.2 and Chapter 6.

The following discussion returns to the computational abstraction and outlines the start of a solution to this problem which exploits the interesting behaviour of the TCR and so its computational counterpart the receptor.

First, the missing molecular recognition of the TCR is addressed. Consider that a pMHC complex is represented as a high dimension vector $\mathbf{m}$, for example the vector could specify the location of every atom of the complex. Then, consider the same for the TCR with resulting vector $\mathbf{t}$. Let κ be a binding function that calculates the dissociate rate m_d of that TCR and pMHC,

$$\kappa(\mathbf{m}, \mathbf{t}) \to m_d. \tag{4.3}$$

It would be possible to then use a function like κ to allow higher dimension inputs to the generalised receptors. This, more or less, is the approach of *shape-space* in theoretical immunology [Perelson and Oster, 1979] and many artificial immune systems [Timmis et al.,

2008a]. The approach sets the dimension by the problem, if input u_t is an n-dimensional input then the "TCR" is a n-dimensional vector and binding function κ is set as to a inverted standard distance metric. However, this is not a very immunological *realistic* model, [Carneiro and Stewart, 1994] notes that the true κ is highly non-linear and discontinuous. It certainly would not behave like a metric. From a computational perspective, comparisons such as κ behave non-intuitively as n increases. [Stibor, 2006] gives good discussion of the issues dimensionality in relation to AIS. Returning to the missing external component of the TCR, it is possible to use a shape-space approach to convert an n-dimensional input into a magnitude for use in the algorithm. Even if one ignores the difficulty of choosing an appropriate κ, this process shifts some of the computational difficulty out of the receptors and in to κ. This would slightly defeat the point of the whole modelling exercise. Moreover, even with a κ the issue of the simplistic outlier detection remains.

An alternative solution can be found by recalling that TCRs are arranged uniformly on the cell surface (at least by the assumptions of the last chapter and [Altan-Bonnet and Germain, 2005]). The same can be done for receptors, and in the next section receptors will be arranged in a uniform lattice. The dimension of the lattice is set to the same as the input, so an n-dimensional u_t is input to an n dimensional lattice. As is seen in the next section, each input must have a n-dimensional location and a magnitude; the input will influence receptors near its location. High dimension concerns still apply here, but in a slightly different manner to a function κ. An algorithm resulting from a lattice of receptors would currently be a 1-dimensional outlier detector with n-dimensional labels. Again, not very algorithmically useful. However, through consideration of the structure of the equations governing a lattice of receptors connections are found to a conventional computational approach to the n-dimensional outlier detection problem. This is shown in Section 4.3 where outlier detection is given a more formal description. In Section 4.4.2, this connection suggests an alteration to receptor interactions which will improve the richness of the information processing of a lattice of receptors.

4.1.3 Multiple receptors in a Lattice

Definition 4.2. *A n-dimensional lattice L of receptors is a set of uniformly spaced receptors in $\mathbb{R}^n$. Receptors in L are located at $(k_0\delta_0, k_1\delta_1, \ldots, k_{n-1}\delta_{n-1})$ and $k_i \in \mathbb{Z}$ and δ_i is the separation between receptors in the ith dimension. The lattice is restricted to be finite, so $k_i \in [k_{i_{min}}, k_{i_{max}}]$. The location of a receptor in $r \in L$ is written $r_{\mathbf{x}}$, and the receptor position and negative feedback for a receptor at point $\mathbf{x}$ are written $p_t(\mathbf{x})$ and $n_t(\mathbf{x})$ respectively.*

It is necessary for each input to have a *location* and a *strength* in analogy to the respective position of a pMHC ligand on the APC surface and the affinity of the pMHC ligand for the TCR. The input sequence to a receptor lattice is now $\{\boldsymbol{\gamma}_t\}$ with $\boldsymbol{\gamma}_t = (u_t, \mathbf{z}_t)$, $u_t \geq 0$ denotes the strength of the input and $\mathbf{z}_t \in \mathbb{R}^n$ is the location of the input. If $\mathbf{z}_t$ falls outside the boundaries of the lattice then the input $\boldsymbol{\gamma}_t$ has no influence on the receptors. The

recurrences (4.2) may now be augmented to describe the interaction between receptors:

$$\begin{aligned} p_{t+1}(\mathbf{x}) &= bp_t(\mathbf{x}) + K_S(\mathbf{x}, \boldsymbol{\gamma}_t) - an_t(\mathbf{x}) \\ n_{t+1}(\mathbf{x}) &= dK_N(\mathbf{x}, n_t) + gH(p_t(\mathbf{x}) - \beta), \end{aligned} \tag{4.4}$$

with $\mathbf{x} \in L$ a point in the lattice; $H(\cdot)$ the Heaviside step function $H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ otherwise [3]; $K_S(\cdot)$ and $K_N(\cdot)$ describe the interaction between neighbouring receptors. K_S is the stimulation kernel function, it is used to model the spreading of stimulation due to a single pMHC ligand, as the ligand may serially engage to the same or nearby TCRs. K_S is defined,

$$K_S(\mathbf{x}, \boldsymbol{\gamma}_t) = \frac{u_t}{h^n} K\left(\frac{\mathbf{x} - \mathbf{z}_t}{h}\right), \tag{4.5}$$

with $K(\cdot)$ is a kernel function with width h. All kernels considered in this thesis will have the constraints that,

$$K(\mathbf{x}) \geq 0, \qquad \text{and} \qquad \int_{\mathbb{R}^n} K(\mathbf{x}) d\mathbf{x} = 1. \tag{4.6}$$

Further discussions of kernel functions are given in Section 4.3. The function K_N similarly makes use of a kernel to describe the diffusion of negative feedback through the lattice:

$$K_N(\mathbf{x}, n_t) = \frac{1}{h^n} \sum_{r \in L} n_t(r_{\mathbf{x}}) K\left(\frac{\mathbf{x} - r_{\mathbf{x}}}{h}\right). \tag{4.7}$$

The function of the K_S and K_N is described in Figure 4.2. With regards to the boundaries of the lattice, later (Section 4.4 onwards) it will be seen that it is possible to make the lattice sufficiently large that all activity of interest will fall well within the confines of the lattice. This is due to a restriction of negative feedback diffusion, and it is then possible to set all boundary $p_t(\mathbf{x}), n_t(\mathbf{x}) = 0$. Until then, the boundaries of the lattice will be connected as an n-dimensional torus, and loosely speaking this models the lack of boundaries of a T cell surface.

The above describes the lattice of receptors receiving a single input per time step, it is a simple extension to describe a situation involving many inputs $\boldsymbol{\gamma}_{t_0}, \boldsymbol{\gamma}_{t_1}, \ldots, \boldsymbol{\gamma}_{t_m - 1}$ per time step. The input term in the receptor position recurrence in (4.4) is augmented to:

$$\sum_{i=0}^{m-1} K_S(\mathbf{x}, \boldsymbol{\gamma}_{t_i}). \tag{4.8}$$

These definitions add additional information to the assumption set A that connects the biological and abstracted models. Note that the generalised receptor, with additional justification, could be regarded as being situated on the *left* of Figure 1.1. Certainly, it is not a very realistic model and neither is it particularly general. So it is not a very good model of biology, but a model of biology nonetheless. Whereas, by definition the lattice

[3] Note that this definition sets $H(0) = 1$.

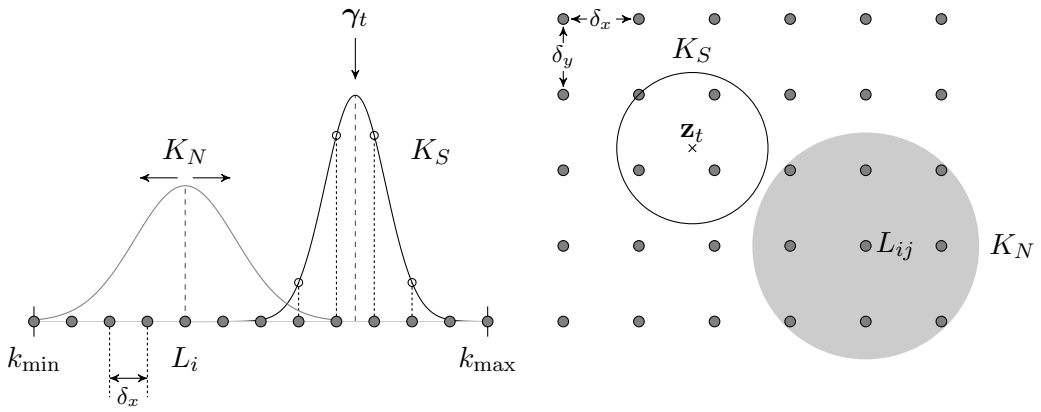

Figure 4.2: *Left:* A 1-dimensional lattice of receptors. K_S is shown for an input $\boldsymbol{\gamma}_t$ which falls between two receptors, note how K_S dictates the contribution of $\boldsymbol{\gamma}_t$ to the receptors surrounding $\mathbf{z}_t$. K_N describes diffusion of negative feedback from a receptor L_i. *Right:* A 2-dimensional lattice of receptors, the circles depict contours of K_N and K_S.

of receptors added detail to alleviate the computational concerns of replicating the TCR. However, the detail added was biological and with additional justification this too could be regarded as a purely biological model. Whether for biology or algorithms, it is necessary to examine whether this abstracted model can replicate the behaviour of the hypothesis of the last chapter. This is done in the next section.

4.1.4 A Comparison with the Biology

This section replicates the behaviour of the last chapter's hypothesis in the abstracted model. For the sake of example, take a 1-dimensional lattice L with $k_{\min} = 0$, $k_{\max} = 199$, $\delta = 1$ resulting in a lattice of 200 receptors. Each receptor is identical with parameters $b = 0.95$; $d = 0.99$; $g = 0.26$; $a = 1$; $\beta = 1$; $\ell = 4$. The kernels used in K_S and K_N are the standard normal kernel (4.38). For K_S, $h = 0.75$ so $\approx 99\%$ of an input's influence is spread over 4 receptors. For K_N, $h = 15$ so every time step $\approx 99\%$ of the negative feedback of a receptor diffuses to within 45 receptors distance. This choice for K_N reflects the well-mixed assumption used in the model in the last chapter.

Three input scenarios are chosen, in each the input is constant in time with $m = 200$ input points. Each input location is chosen uniformly at random within the range of the lattice $z_i \sim U(k_{\min}, k_{\max})$. The strength of each input point are chosen at random from differing distributions as follows:

$$\begin{aligned}
&\text{Scenario } a\text{:} \quad u_i \sim N(\mu_1, \sigma_1) \\
&\text{Scenario } b\text{:} \quad u_i \sim \begin{cases} N(\mu_1, \sigma_1) & \text{for } (1-\rho)m \text{ of the input.} \\ N(\mu_2, \sigma_2) & \text{for } \rho m \text{ of the input.} \end{cases} \\
&\text{Scenario } c\text{:} \quad u_i \sim N(\mu_2, \sigma_2),
\end{aligned}$$

with $\mu_1 = 1$, $\mu_2 = 2.75$ and $\sigma_1, \sigma_2 = \frac{1}{3}$ and $\rho = 0.05$. In scenario a and c the magnitudes

are drawn from the same respective distributions, in scenario b the magnitudes of 190 input points (also chosen uniformly at random) are drawn from the distribution with the lower mean, and 10 points are drawn from the distribution with the greater mean. The results are given in Figure 4.3. In scenario a and c the negative feedback uniformly grows to just above the mean of each input distribution and all $p_t(\mathbf{x}) < \ell$. In scenario b, the mean negative feedback is not enough to contain the receptors with input $N(\mu_2, \sigma_2)$ and receptor activation occurs. To map back to biology: a describes a T cell presented with a population of poor quality ligands pMHC ligands; b describes a T cell presented with a majority of poor quality pMHC and a minority of good pMHC; c describes a T cell presented with a population of solely high quality pMHC which pushes the cell into anergy. There are some discrepancies between this abstracted model and the biological model. In particular, the dynamics involving the time window before the rise of negative feedback are different. In part, this is related to continuous representations of receptor position and the linear application of negative feedback. If the objective was not an algorithm, it would be an interesting exercise to return to the definitions of the abstracted model and investigate how to achieve a better correspondence between both models. As algorithm development is the objective, the chapter continues with the current level of agreement between the biological and abstracted models.

The behaviour of the dynamics of a single receptor (4.2) are analysed in the following section.

4.2 The Single Receptor: Solutions and Equilibrium

It is convenient to calculate some basic properties of the single receptor dynamics described in (4.2). This will be of use when connecting the behaviour of a receptor to conventional anomaly detection approaches in Section 4.3. As the discussion is of a single receptor there is no concept of a lattice or interactions between neighbouring receptors, kernel based functions K_S and K_N do not apply. This is in analogy to the previous chapter, in that the isolated behaviour of a receptor is investigated.

The recurrences in (4.2) may be re-written in matrix form, with state variable $\boldsymbol{\sigma}_t = (p_t \; n_t)^T$:

$$\boldsymbol{\sigma}_{t+1} = A\boldsymbol{\sigma}_t + \mathbf{f}_t. \tag{4.9}$$

The 2×2 matrix A describes the interaction between p_t and n_t, the vector $\mathbf{f}_t$ describes the input at time t:

$$A = \begin{pmatrix} b & -a \\ 0 & d \end{pmatrix} \qquad \mathbf{f}_t = \begin{pmatrix} u_t \\ gH(p_t - \beta) \end{pmatrix}. \tag{4.10}$$

The solution to (4.9) is,

$$\boldsymbol{\sigma}_t = A^t\boldsymbol{\sigma}_0 + \sum_{k=0}^{t-1} A^{t-1-k}\mathbf{f}_k. \tag{4.11}$$

Calculating A^k is of relevance, and it may be found via the eigenvalues $\lambda = b, d$ of A and eigenvectors $\mathbf{e}_1 = (0 \; 1)^T$ and $\mathbf{e}_2 = (1 \; \frac{b-d}{a})^T$ of A. Then, $A^k = E\Lambda^k E^{-1}$ with $E = [\mathbf{e_1} \; \mathbf{e_2}]$

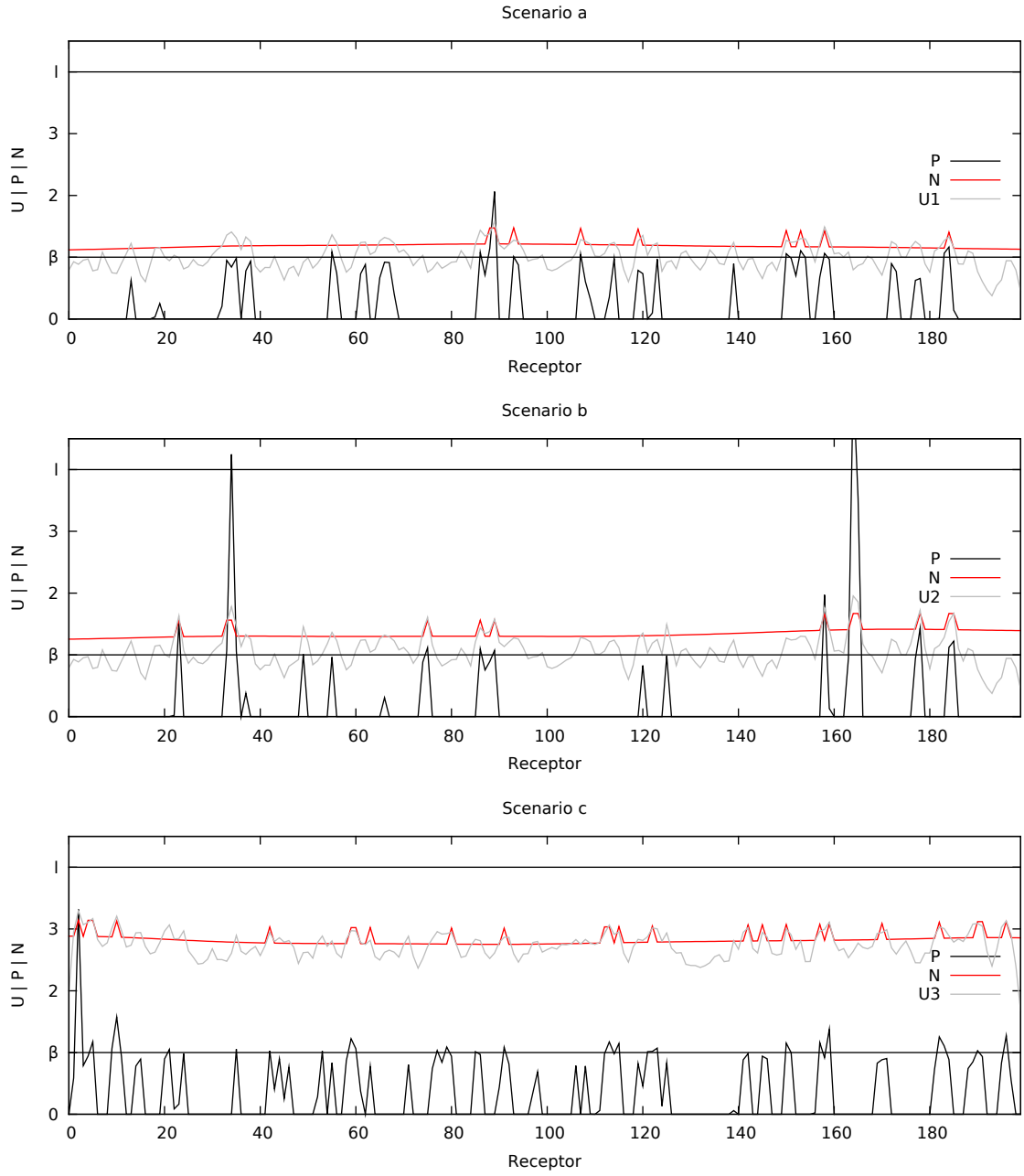

Figure 4.3: The input (grey), receptor positions (black) and negative feedback (red) for scenarios *a, b* and *c*. The lattice is drawn along the horizontal axis. See text for details and parameters. Note scenario *b* is the only scenario in which a receptor activation occurs.

and Λ the diagonal matrix of eigenvalues.

$$A^k = \begin{pmatrix} b^k & \frac{a(d^k - b^k)}{b-d} \\ 0 & d^k \end{pmatrix}. \tag{4.12}$$

Consequently the solution to (4.9) written for p_t and n_t is,

$$p_t = p_0 b^t + \frac{a(d^t - b^t)}{b-d} n_0 + \sum_{k=0}^{t-1} \left[u_k b^{t-1-k} + \frac{ag(d^{t-1-k} - b^{t-1-k})H(p_k - \beta)}{b-d} \right] \tag{4.13}$$

$$n_t = n_0 d^t + g \sum_{k=0}^{t-1} H(p_k - \beta) d^{t-1-k}. \tag{4.14}$$

Note that a only acts to scale n_t, the trajectory of p_t is unchanged if $a' = 1$ in (4.13) and $g' = ag$ in (4.14)[4]. Thus, a is eliminated from all further discussions.

The step in negative feedback presents some difficulties, however exact solutions for below and above the step can be given that are only valid whilst p_t remains in the relevant region. Let $\phi_\alpha^t \equiv \sum_{k=0}^t \alpha^k = \frac{\alpha^k - 1}{\alpha - 1}$, then,

$$\text{For } p_t < \beta \quad \begin{cases} p_t = p_0 b^t + \frac{d^t - b^t}{b-d} n_0 + \sum_{k=0}^{t-1} u^k b^{t-1-k} \\ n_t = d^t n_0 \end{cases} \tag{4.15}$$

$$\text{For } p_t \geq \beta \quad \begin{cases} p_t = p_0 b^t + \frac{d^t - b^t}{b-d} n_0 + \sum_{k=0}^{t-1} u^k b^{t-1-k} + \frac{g(\phi_d^t - \phi_b^t)}{b-d} \\ n_t = d^t n_0 + g\phi_d^t \end{cases}. \tag{4.16}$$

The equilibrium positions and fixed points of recurrence (4.9) are now discussed.

4.2.1 Fixed Points

The parameters b, d, g, β, ℓ define a scale with which to interpret the input u_t. Assume a constant input $u_t = u \geq 0$, then the summation involving u_t in (4.15) and (4.16) becomes $u\phi_b^t$. Then, four ranges of u: $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3, \mathcal{U}_4$ (Figure 4.4) define four regions of equilibrium for p_t and n_t. The receptor position at equilibrium is written p_e and similarly equilibrium negative feedback is written n_e.

- $\mathcal{U}_1 = \{u : u < \beta(1-b)\}$. For $u \in \mathcal{U}_1$ the negative feedback will not be broken, no negative feedback will be generated: $p_e = u/(1-b)$, $n_e = 0$.
- $\mathcal{U}_2 = \{u : \beta(1-b) \leq u < \beta(1-b) + g/(1-d)\}$. For $u \in \mathcal{U}_2$, negative feedback will be generated and it will grow and arrive at approximate equilibrium positions $p_e \approx \beta$, with $n_e \approx u - \beta(1-b)$. The details of this are discussed further in Section 4.2.2.
- $\mathcal{U}_3 = \{u : \beta(1-b) + g/(1-d) \leq u < \ell(1-b) + g/(1-d)\}$ For $u \in \mathcal{U}_3$ the negative feedback will not be able to hold $p_e \approx \beta$ and will rise to a maximum $n_e = g/(1-d)$. However, this maximum is enough to contain $p_e = u/(1-b) + g/(1-d)(1-b) < \ell$, and so changes in u to produce a $p > \ell$ are still of interest.

[4]The dashed notation denotes a new assignment of the variable

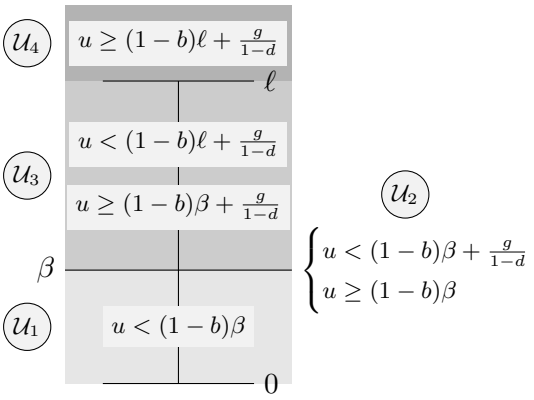

Figure 4.4: The equilibrium values of p_e and n_e for ranges of u defined by regions $\mathcal{U}_1$,$\mathcal{U}_2$,$\mathcal{U}_3$,$\mathcal{U}_4$, see text for details.

- $\mathcal{U}_4 = \{u : u \geq \ell(1-b) + g/(1-d)\}$ For $u \in \mathcal{U}_4$ the equilibrium positions are calculated identically as in region 3, but now $p > \ell$ and so this situation is not of interest.

Input $u \in \mathcal{U}_2$,$\mathcal{U}_3$ can "hold" the receptor position at β, with negative feedback directly related to u. It is only the lower boundary of $\mathcal{U}_3$ that can hold $p_e = \beta$ with maximum negative feedback, but because the negative feedback is constant for all $u \in \mathcal{U}_3$ the input range will be of no particular interest. Input range $\mathcal{U}_2$ will have the most interest for algorithmic purposes in Section 4.4.

For input in $\mathcal{U}_1$,$\mathcal{U}_3$,$\mathcal{U}_4$ the $(p_e, n_e)^T$ are fixed points of (4.9), and $(p_t, n_t)^T$ will asymptotically approach these points as $t \to \infty$. This can be established by noting that in these input regions the p and n recurrences only transiently experience the non-linearity of $H(\cdot)$ dependent on initial (p_0, n_0). After influence of the initial state has decayed the equations are linear with distinct eigenvalues $|\lambda_i| < 1$ and so the fixed points are stable. This asymptotic behaviour can also be seen with consideration of the difference operator:

$$\Delta\alpha_t = \alpha_{t+1} - \alpha_t. \tag{4.17}$$

Then, define $\delta_{\alpha t} = \alpha_e - \alpha_t$, as the distance of variable α from its equilibrium position. In regions $\mathcal{U}_1$,$\mathcal{U}_3$,$\mathcal{U}_4$ the p and n differences are all of the form $\Delta\alpha_t = \epsilon\delta_{\alpha t}$ with $\epsilon < 1$, since $\epsilon = 1 - b$ for p and $\epsilon = 1 - d$ for n. Each step p and n jump toward the equilibrium by a fraction of their distance from the equilibrium. In $\mathcal{U}_2$ the picture is quite different, p cannot escape the non-linearity and the system becomes periodic.

4.2.2 Region 2 Cycles

The only fixed point of the n-recurrence in (4.2) is the maximum value $n_e = g/(1-d)$. By definition an input $u \in \mathcal{U}_2$ cannot hold n_e at this point. So, p and n will cycle around $p_e = \beta$, $n_e = u - \beta(1-b)$ as p jumps above and below β. This behaviour is demonstrated in Figure 4.5. For diagrammatic clarity β is depicted as a horizontal barrier, consequently all discussions now regard points in the (n, p)-plane. The gradient $\Delta p_t/\Delta n_t$ describes the orbit around (n_e, p_e), and p and n differences are,

$$\Delta p_t = (b-1)p_t + u - n_t \tag{4.18}$$

$$\Delta n_t = (d-1)n_t + g. \tag{4.19}$$

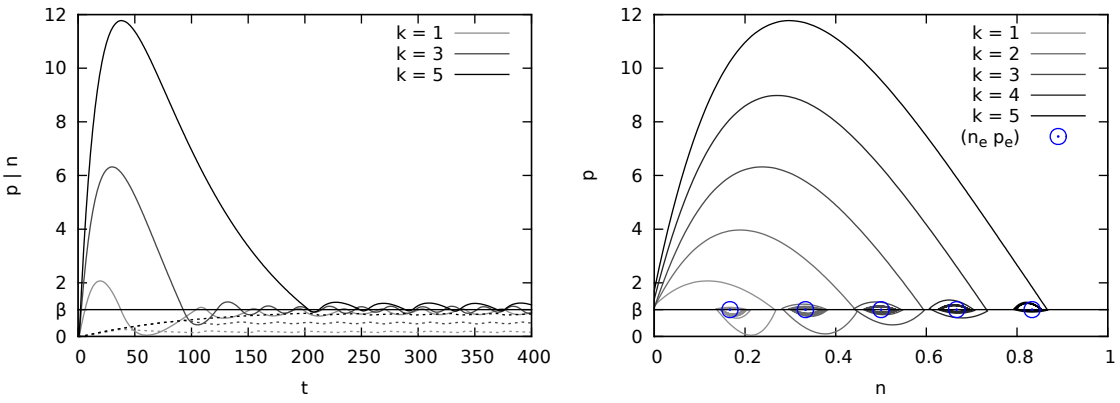

Figure 4.5: The oscillating behaviour of p_t and n_t for $u \in \mathcal{U}_2$, with parameters are $b = 0.95$; $d = 0.99$; $g = 0.01$; $\beta = 1$; $p_0, n_0 = 0$. $\mathcal{U}_2$ has $0.05 < u < 1.05$, the figures plot $u = k/6 + 0.05$. *Left:* plot p_t and n_t trajectories in time. *Right:* The trajectories are in the (n, p)-plane. Note that the magnitude of u changes the shape of the oscillation.

The signs of Δp_t and Δn_t partition the positive (n, p)-plane into four, clearly $\Delta n_t \geq 0$ when $p_t \geq \beta$ and $\Delta n_t \leq 0$ when $p_t < \beta$. The sign of Δp_t is defined by the $\Delta p_t = 0$ line: $p = (u - n)/(1 - b)$. Thus the four partitions are (arranged to match Figure 4.6),

$$\begin{aligned} \mathcal{R}_2 &= \left\{(n,p) : p \geq \beta \wedge p < \tfrac{u-n}{1-b}\right\} & \mathcal{R}_3 &= \left\{(n,p) : p \geq \beta \wedge p \geq \tfrac{u-n}{1-b}\right\} \\ \mathcal{R}_1 &= \left\{(n,p) : p < \beta \wedge p < \tfrac{u-n}{1-b}\right\} & \mathcal{R}_4 &= \left\{(n,p) : p < \beta \wedge p \geq \tfrac{u-n}{1-b}\right\} \end{aligned} \tag{4.20}$$

Figure 4.6 displays the four partitions along with a vector field for one choice of parameters. From any point $(n_t, p_t) \in \mathcal{R}_i$ an update may leave $(n_{t+1}, p_{t+1}) \in \mathcal{R}_i$. Note that the directions of the region boundaries oppose the gradients in each partition and so it is also possible to move to the next partition: $\mathcal{R}_{(i+1 \bmod 4)}$. These are the only possible transitions due to the following result:

Theorem 4.1. *A receptor under constant input $u \in \mathcal{U}_2$ will fall into the following cycle:*

$$\mathcal{R}_1 \to \mathcal{R}_2 \to \mathcal{R}_3 \to \mathcal{R}_4 \to \mathcal{R}_1. \tag{4.21}$$

Proof. First, consider the simple case of $\mathcal{R}_2$ and $\mathcal{R}_4$. In $\mathcal{R}_2$, by definition $\Delta p_t > 0$ and so it is only possible to leave $\mathcal{R}_2$ and enter $\mathcal{R}_3$. Then in $\mathcal{R}_4$, by definition $\Delta p_t < 0$ and since $p_t \geq 0$ it is only possible to leave $\mathcal{R}_4$ and enter $\mathcal{R}_1$.

In $\mathcal{R}_1$ and $\mathcal{R}_3$ it is necessary to show that (n_t, p_t) and (n_{t+1}, p_{t+1}) lie on the same side of line $p = (u - n)/(1 - b)$. In $\mathcal{R}_1$, the inequalities $p_t < (u - n)/(1 - b)$ and $p_t < \beta$ hold for a point (n_t, p_t), then $p_{t+1} < (u - n_{t+1})/(1 - b)$ holds since:

$$bp_t + u - n_t < \frac{u - dn_t}{1 - b}, \tag{4.22}$$

$$p_t < \frac{u - n_t}{1 - b} + \frac{n_t}{b}\left[\frac{1 - d}{1 - b}\right]. \tag{4.23}$$

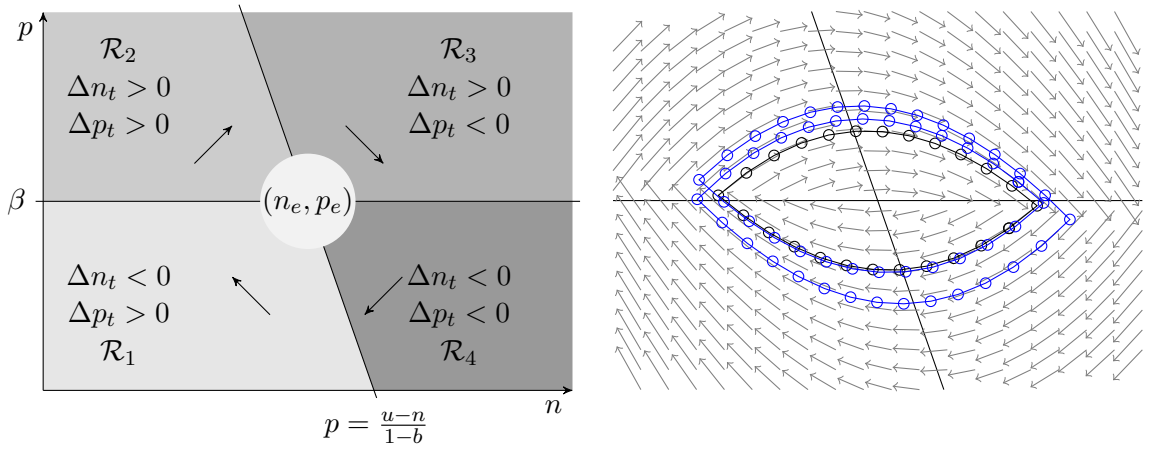

Figure 4.6: *Left:* Four partitions surrounding (n_e, p_e) defined by the sign of Δn_t and Δp_t which defines lines $p = \beta$ and $p = \frac{u-n}{1-b}$. *Right:* A vector field for parameters $b = 0.95$; $d = 0.99$; $g = 0.01$; $\beta = 1$; and $u = 0.55$, two cycles are shown which are periodic to 10^{-16}. The larger (blue) cycle has a period of 54 and is obtained from starting point $(0, 0)$; the smaller (black) is period 24 obtained from (n_e, p_e).

In $\mathcal{R}_3$, the inequalities $p_t \geq (u - n_t)/(1 - b)$ and $p_t \geq \beta$ hold for a point (n_t, p_t), then $p_{t+1} \geq (u - n_{t+1})/(1 - b)$ holds since:

$$bp_t + u - n_t \geq \frac{u - (dn_t + g)}{1 - b}, \tag{4.24}$$

$$p_t \geq \frac{u - n_t}{1 - b} + \frac{n_t}{b}\left[\frac{1 - d}{1 - b}\right] - \frac{g}{b(1 - b)}, \tag{4.25}$$

and note that:

$$\frac{n_t}{b}\left[\frac{1 - d}{1 - b}\right] - \frac{g}{b(1 - b)} \leq 0 \quad \text{holds since} \quad n_t \leq \frac{g}{1 - d}. \tag{4.26}$$

□

Thus, for $u \in \mathcal{U}_2$ n and p will cycle around the point (n_e, p_e). Two cycles are shown which are found starting from $(0, 0)$ and (n_e, p_e) with periods of 24 and 54 respectively. The start and end points of each cycle match to 15 decimal places. The non-uniqueness of the cycles is due to the discrete updates of the equations.

Due to the discrete equations governing p and n only a subset of the entire (n, p)-plane is accessible. This is of less relevance for p_t as the set of points it may occupy is a function of u_t. However, for n_t the set of accessible points is a function of g and d. Let τ_k denote a time step between 0 and t in which $p_{\tau_k} \geq \beta$, if this occurs n times then $\tau_1 < \tau_2 < \cdots < \tau_n$ and so n_t may only occupy points of the form,

$$n_t = \frac{g}{1 - d}\left[m - \sum_{k=1}^{n} d^{\tau_k}\right]. \tag{4.27}$$

This is not too concerning as t becomes large, however the one step difference $n_t \to \{dn_t, dn_t + g\}$ may be of concern. Thus for relatively smooth behaviour of n, d should be near 1 and g should be near 0.

The details of the possible behaviour of the cycles can be further understood by considering the points in the (n, p)-plane which are brought closer to the equilibrium by an iteration of the n and p recurrences. Consider the sum of the squared differences between a current point (n, p) and the equilibrium (n_e, p_e):

$$d_e(n, p) = (n - n_e)^2 + (p - p_e)^2. \tag{4.28}$$

An iteration takes $(n, p) \to (n_\star, p_\star)$ and the points for which $d_e(n_\star, p_\star) - d_e(n, p) = 0$ are the set points whose distance (but not position for $u \in \mathcal{U}_2$) from the equilibrium is left unchanged. It is necessary to calculate this distance above and below the line $p = \beta$, which gives:

$$\text{For } p < \beta, \qquad (dn - n_e)^2 - (n - n_e)^2 + (bp + u - n - p_e)^2 - (p - p_e)^2 = 0 \tag{4.29}$$

$$\text{For } p \geq \beta, \qquad (dn + g - n_e)^2 - (n - n_e)^2 + (bp + u - n - p_e)^2 - (p - p_e)^2 = 0. \tag{4.30}$$

Each of the above has the form of a conic, and both have discriminant $d^2(b^2-1)-b^2 < 0$ for $0 < b, d, < 1$ see (4.33), thus they are hyperbolas [Korn and Korn, 2000]. They divide the (n, p)-plane into regions which are attracted and repulsed from (n_e, p_e), see Figure 4.7. For small g the two hyperbolas are effectively identical, as g increases the hyperbola above moves away from the point (n_e, p_e). In Figure 4.7 (*right*) the large g prevents the trajectory from reaching the upper hyperbola, resulting in qualitatively different cycle behaviour. Since $p_t \geq 0$, if p is pushed to zero by a large negative feedback it will remain there until the negative feedback level has decayed.

The influence of g on the two hyperbolas can be clarified by a normalisation to a standard hyperbola form. In particular, the pn term in (4.30) results in a rotation of the hyperbola; a removal of this rotation along with a translation to the origin brings each hyperbola into a standard form. Hyperbola (4.30) may be expanded into the form:

$$a_{nn}n^2 + 2a_{np}np + a_{pp}p^2 + 2a_n n + 2a_p p + a = 0, \tag{4.31}$$

with the following parameters,

$$\begin{array}{llll} a_{nn} & = d^2 & a_n & = dg + n_e(1 - d) + p_e - u \\ a_{np} & = -b & a_p & = bu + p_e(1 - b) \\ a_{pp} & = b^2 - 1 & a & = g(g - 2n_e) + u(u - 2p_e). \end{array} \tag{4.32}$$

There are three invariants with respect to translation and rotation of a conic [Korn and Korn, 2000]:

$$Q = a_{nn} + a_{pp}, \qquad D = \begin{vmatrix} a_{nn} & a_{np} \\ a_{np} & a_{pp} \end{vmatrix}, \qquad A = \begin{vmatrix} a_{nn} & a_{np} & a_n \\ a_{np} & a_{pp} & a_p \\ a_n & a_p & a \end{vmatrix}, \tag{4.33}$$

D is the determinant of the conic, which is a hyperbola when $D < 0$ as is the case (4.30) and (4.29). The eigenvalues of the matrix whose determinant is D are also relevant and are

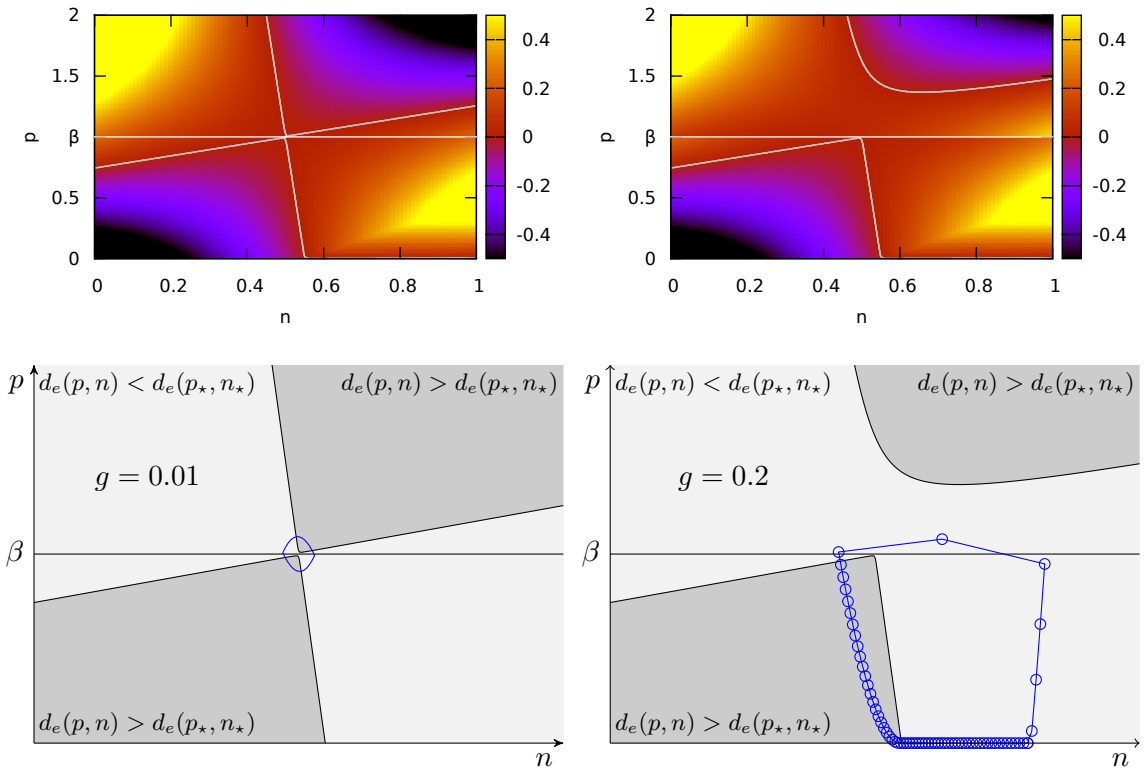

Figure 4.7: Regions where the iteration $(n,p) \to (n_\star,p_\star)$ brings $(n_\star,p_\star)$ closer to (n_e,p_e). All parameters are as Figure 4.5, except *Left:* $g = 0.01$ and *right:* $g = 0.2$. *Top:* The color maps show $d_e(n_\star,p_\star) - d_e(n,p)$ on the (n,p)-plane, marked are the hyperbolas of the zero contour. *Bottom:* Regions where the iteration results in $(n_\star,p_\star)$ closer to (n_e,p_e) than (n,p) (light grey) and further away (dark grey). Blue line shows the period 24 cycle from Figure 4.5 (left) a period 66 cycle obtained when $g = 0.2$ (right).

the roots of the equation $\lambda^2 - Q\lambda + D = 0$. Then the standard form for the hyperbola is calculated as follows:

$$\frac{x^2}{q_1^2} - \frac{y^2}{q_2^2} = 1, \qquad q_1^2 = -\frac{A}{\lambda_1^2\lambda_2}, \qquad q_2^2 = \frac{A}{\lambda_1\lambda_2^2}. \tag{4.34}$$

The standard form is given in (x,y) coordinate system which is the (n,p) coordinate system transformed via a translation of the centre of the conic to the origin and a rotation: $\tan 2\theta = a_{np}/(a_{nn} - a_{pp})$. Hyperbolas for a range of g are shown in Figure 4.8, note how the increase in g shifts the focus of the hyperbola away from the origin.

Finally, Figure 4.9 shows the cycles for inputs from $\mathcal{U}_2$. For parameters different to those used in Figure 4.9 the qualitative cycle behaviour is the same with the exception of large g results in p hitting its axis (see Figure 4.7).

The discussion now turns to conventional approaches to anomaly detection, through which connections will be made to the behaviour of a lattice of receptors.

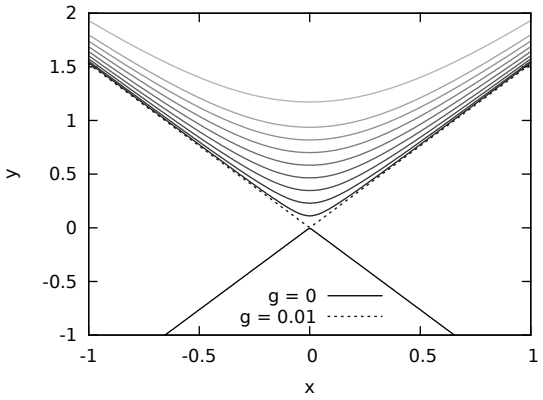

Figure 4.8: Hyperbolas in standard form. Parameters as Figure 4.5 and $g = 0, 0.01, k/10$ for $k = 1, \ldots, 10$. All have asymptotes $y = \pm\frac{q_2}{q_1}x = \pm 2.36x$. Only relevant halves of $g = 0$ and $g \neq 0$ hyperbolas shown. $g = k/10$ hyperbolas are plotted on a linear grey scale, $k = 1$ lightest and $k = 10$ darkest.

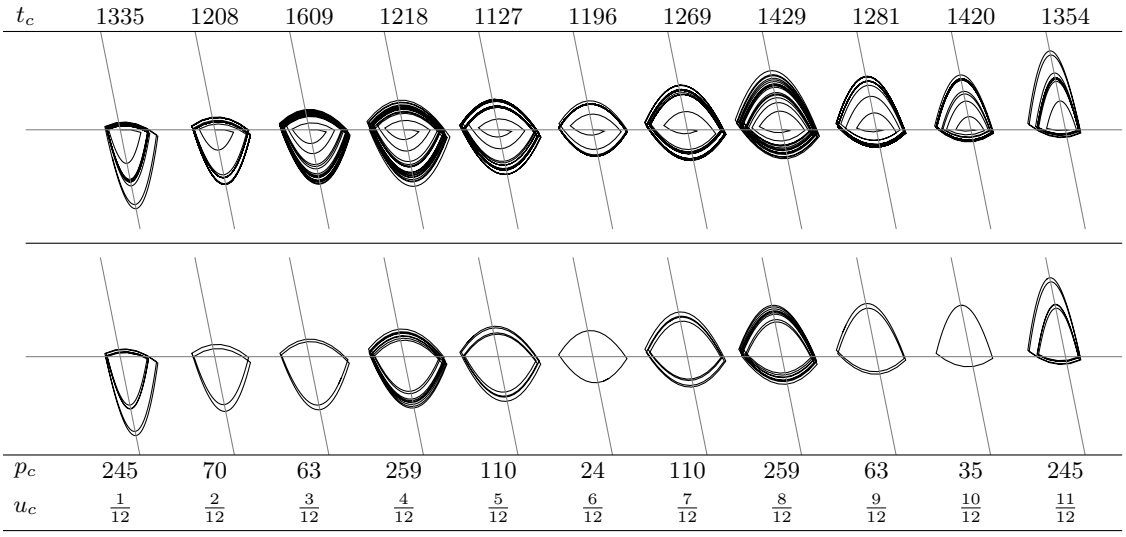

Figure 4.9: Cycles around (n_e, p_e) for parameters as Figure 4.5 and inputs $u = u_c + \beta(1 - b)$, with $u_c = \frac{k}{12}$. All from initial $n_0 = n_e$ and $p_0 = p_e$. *Bottom:* shows the final cycle correct to 1×10^{-8}. *Top:* gives the trajectories to the point the cycle starts. t_c gives the time step the cycle is first reached, p_c gives the period. Note the symmetry between $u_c = \frac{k}{12}$ and $u_c = 1 - \frac{k}{12}$, in part this is due to $\frac{g}{1-d} = 1$. However, it is still present to a lesser degree when $\frac{g}{1-d} \neq 1$.

4.3 Classical Anomaly Detection and Statistical Learning

This section provides a definition of anomaly detection and gives an overview of some conventional statistical anomaly detection approaches. For a superior and more comprehensive presentation of classical anomaly detection techniques the reader is directed to [Silverman, 1986; Bishop, 1995; Scott, 1992; Duda et al., 2001; Tsybakov, 2009]. The references provide a range of discussion on the definition of an "anomaly", for current purposes it is convenient to discuss anomalies in terms of a sequence of observations occurring sequentially in time. There is an interest in the following problems,

1. Given a sequence of m observations $\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_{m-1} \in \mathbb{R}^n$, classify a subsequent observation $\mathbf{v}$ normal or anomalous. This is denoted a *type 1* anomaly. This is also a manner in which one can formalise the outlier detection discussed in Section 4.1.2.

2. The classification of *type 1* anomalies has the assumption that the previous m observations are a good model with which to evaluate $\mathbf{v}$. This assumption can be addressed by observing changes in the properties of the m observations. If the properties change too quickly such that *type 1* anomaly detection cannot be performed with confidence, a *type 2* anomaly occurs.

The two types of anomaly can be formalised in terms of estimates of a probability distribution assumed to underlie the m observations. First, *type 1* anomaly detection is addressed with the assumption that underlying distribution is static in time. Second, *type 2* anomaly detection will be addressed in Section 4.3.2 at which point the distribution of concern will change in time.

A collection of m, $n-$dimensional observations $\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x}_{m-1} \in \mathbf{x} = \mathbb{R}^n$ are assumed to be independent and identically distributed (i.i.d.) with probability density function $p(\mathbf{x})$. Anomaly detection can be discussed in terms of an estimate of the density function $\hat{p}(\mathbf{x}|\mathbf{x}_0, \ldots, \mathbf{x}_{m-1})$.

With the estimation of distributions in mind, briefly return to the equations of the receptor and consider a n-dimensional lattice of receptors L in which there is no negative feedback ($a = 0$, or alternatively, $n_0 = 0$ and $g = 0$) with input $\boldsymbol{\gamma}_t = (u_t, \mathbf{x}_t)$. If the receptor position is written $\rho_t(\mathbf{x})$ then,

$$\rho_{t+1}(\mathbf{x}) = b\rho_t(\mathbf{x}) + K_S(\mathbf{x}, \boldsymbol{\gamma}_t), \tag{4.35}$$

the solution for ρ_t can be calculated, for $\rho_0(\mathbf{x}) = 0$,

$$\begin{aligned} \rho_t(\mathbf{x}) &= \sum_{i=0}^{t-1} b^{t-1-i} K_S(\mathbf{x}, \boldsymbol{\gamma}_t) \\ &= \frac{u_t}{h^n} \sum_{i=0}^{t-1} b^{t-1-i} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right). \end{aligned} \tag{4.36}$$

The above equation has the form of a weighted *kernel density estimate*. The weighting is by powers of b. Kernel density estimation [Silverman, 1986] (also known as Parzen Window

Estimation after [Parzen, 1962]) is a method of estimating a distribution that is assumed to underlie a sample. It is attractive due to its non-parametric nature and it will be the focus of the discussions on statistical anomaly detection. Section 4.4.1 will return to the above defined $\rho_t(\mathbf{x})$ and discuss its behaviour in detail. Kernel density estimation gives the following estimate, [Silverman, 1986],

$$\hat{p}(\mathbf{x}) = \frac{1}{mh^n} \sum_{i=0}^{m-1} K\left(\frac{\mathbf{x} - \mathbf{x_i}}{h}\right), \tag{4.37}$$

with $K(\cdot)$ a kernel function with the properties given in (4.6) with width h. Since $K(\cdot)$ is always positive and integrates to one, then the estimate $\hat{p}(\mathbf{x})$ is also a probability density function. Common choices for the kernel $K(\cdot)$ are the standard multivariate normal kernel:

$$K_n(\mathbf{x}) = (2\pi)^{-n/2} \exp\left(-\mathbf{x^T x}/2\right), \tag{4.38}$$

and the bounded multivariate Epanechnikov kernel:

$$K_e(\mathbf{x}) = \begin{cases} (2c_n)^{-1}(n+2)(1 - \mathbf{x^T x}) & \text{if } \mathbf{x^T x} < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{4.39}$$

$c_n = \pi^{\frac{n}{2}}/(\Gamma(\frac{n}{2}+1))$ is the volume of the unit $n-$dimensional sphere. Often, more important than choice of kernel function is choice of kernel width h [Silverman, 1986]. An understanding of this can be gained by a measure of accuracy of the kernel density estimate: the *mean squared error* (MSE) for a point $\mathbf{x}$,

$$\mathrm{MSE}_{\mathbf{x}}(\hat{p}) = \mathbb{E}\left[(\hat{p}(\mathbf{x}) - p(\mathbf{x}))^2\right]. \tag{4.40}$$

The expectation is with respect to the sample points $\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x_{m-1}}$. Since var $\hat{p}(\mathbf{x}) = \mathbb{E}[\hat{p}(\mathbf{x})^2] - \mathbb{E}[\hat{p}(\mathbf{x})]^2$, then expanding the $\mathrm{MSE}_{\mathbf{x}}$ gives,

$$\mathrm{MSE}_{\mathbf{x}}(\hat{p}) = \left(\mathbb{E}[\hat{p}(\mathbf{x})] - p(\mathbf{x})\right)^2 + \text{var } \hat{p}(\mathbf{x}). \tag{4.41}$$

The non-variance term is known as the *bias*, expressions for the bias and the variance for the kernel density estimator are given in [Silverman, 1986], and can be calculated from,

$$\mathbb{E}[\hat{p}(\mathbf{x})] = \frac{1}{h} \int_{\mathbb{R}^n} K\left(\frac{\mathbf{x} - \mathbf{y}}{h}\right) p(\mathbf{y}) d\mathbf{y} \tag{4.42}$$

$$\text{var } \hat{p}(\mathbf{x}) = \frac{1}{mh^2} \int_{\mathbb{R}^n} K\left(\frac{\mathbf{x} - \mathbf{y}}{h}\right)^2 p(\mathbf{y}) d\mathbf{y} - \frac{\mathbb{E}[\hat{p}(\mathbf{x})]^2}{m}. \tag{4.43}$$

Note that the bias is independent of the sample size, and that $\mathbb{E}[\hat{p}(\mathbf{x})]$ is the true density convolved with the width scaled kernel. The bias represents the systematic error caused by smoothing with the kernel, the variance represents the random error caused by sampling. A global measure of the accuracy of an estimator $\hat{p}$ is given by the *mean integrated square error* (MISE),

$$\mathrm{MISE}(\hat{p}) = \int_{\mathbb{R}^n} \mathrm{MSE}_{\mathbf{x}}(\hat{p}) d\mathbf{x}. \tag{4.44}$$

Minimising the MSE or the MISE by h involves a trade-off between the bias and variance, a small h reduces the bias but will give a large variance unless the sample size is also large. The influences of h on a kernel density estimation are shown in Figure 4.10. An approximation for the MSE and so the MISE is obtained in [Silverman, 1986] by truncating a Taylor series, which requires the following assumptions on the kernel function: first, that it is radially symmetric and then,

$$\int_{\mathbb{R}^n} \mathbf{x}K(\mathbf{x})d\mathbf{x} = 0 \qquad \text{and} \qquad \int_{\mathbb{R}^n} \mathbf{x}^2K(\mathbf{x})d\mathbf{x} = k_2 \neq 0, \tag{4.45}$$

for some constant k_2, then finally that p has continuous derivatives of all orders required. The approximation can be solved to find an approximation for the optimal kernel width $h_\star$, for the 1-dimensional case,

$$h_\star = k_2^{-\frac{2}{5}} \left[\int_{-\infty}^{\infty} K(y)^2 dy\right]^{\frac{1}{5}} \left[\int_{-\infty}^{\infty} p''(x)^2 dx\right]^{-\frac{1}{5}} n^{-\frac{1}{5}}. \tag{4.46}$$

When substituting $h_\star$ back into the MISE, [Hodges and Lehmann, 1956] demonstrates that it is the Epanechnikov kernel (4.39) that minimises the MISE expression. As such, the Epanechnikov kernel is optimal in terms of MISE in the class of kernel function with the assumptions given in (4.45). However, [Silverman, 1986] also points out that advantages of the Epanechnikov kernel in terms of MISE are sufficiently minor that it is legitimate to use alternative factors as the deciding criteria for a kernel. As stated above, choosing appropriate h dominates over the choice of kernel. Alternatively, computational requirements are often cited as the reason for choosing one kernel over another[5]. However, it is possible to calculate a kernel density estimation via a precomputed kernel [Scott, 1992]. For a 1-dimensional estimate, this can be done as follows,

$$e_p(x) = \frac{1}{m}\sum_{k=0}^{m-1} \delta(x - x_i) \tag{4.47}$$

$$\hat{p}(x) = e_p(x) * K_h(x) \quad = \int_{-\infty}^{\infty} \sum_{k=0}^{m-1} \delta(t - x_i)K_h(x - t)dt \tag{4.48}$$

$$= \frac{1}{m}\sum_{k=0}^{m-1} K_h(x - x_i), \tag{4.49}$$

with $e_p(x)$ is known as the empirical density function of the sample; $\delta(\cdot)$ is the Dirac delta; $K_h(\cdot)$ is a kernel function that has been appropriately scaled by h. Consequently, $\hat{p}$ can be computed by a convolution of K_h with e_p, and K_h can be precomputed and e_p can be approximated by a histogram with appropriate bin widths. As such, computation of $\hat{p}$ is relatively independent of the difficulty of computing K. In this thesis, due to the inspiration arising from diffusion and for its convolution properties the standard normal kernel is chosen.

[5]Kernel Density Estimation techniques have been present in the literature since the 1950s, during the last 60 years there has been a roughly exponential growth in computing power. Many discussions of computational tractability from just 10 years ago do not apply today.

An asymptotic relationship between sample size and kernel width dictates the convergence of $\hat{p}$. As described in [Parzen, 1962; Silverman, 1986; Duda et al., 2001], and under slightly milder assumptions on the kernel than (4.45) (see references for details of assumptions), the kernel width dependent on sample size is h_m and if it satisfies the following,

$$h_m \to 0 \qquad \text{and} \qquad mh_m \to \infty \qquad \text{as} \qquad m \to \infty, \tag{4.50}$$

then $\hat{p}(\mathbf{x})$ converges to $p(\mathbf{x})$ as $n \to \infty$ with probability 1. The condition essentially states that the kernel width must go to zero but not as rapidly as $1/m$ goes to zero.

The use of the MSE and MISE in choosing h is purely theoretical as both depend on the unknown density $p(\mathbf{x})$. In practice, one possibility is to calculate the MSE and MISE with respect to a reference distribution. Assuming that $p(\mathbf{x})$ is normal with variance σ^2 and assuming that a standard normal kernel is used, then the approximation for the MISE can be solved to give optimal $h_\star$,

$$h_\star = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma m^{-\frac{1}{5}}. \tag{4.51}$$

An estimate σ can be found by calculating the sample standard deviation. If the underlying distribution is not normal, and particularly if it is multimodal $h_\star$ may over-smooth. A heuristic alteration to (4.51) h is suggested by [Silverman, 1986],

$$h = 0.9Am^{\frac{1}{5}}, \qquad A = \min\{\sigma, \text{IQR}/1.34\}, \tag{4.52}$$

with IQR the interquartile range of the sample. In practice, h is usually chosen with respect to the sample $\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x_{m-1}}$, one approach is the leave-one-out likelihood cross-validation [Silverman, 1986]. This involves choosing h to maximise a score $C_v(h)$,

$$C_v(h) = \sum_{i=0}^{m-1} \log \left[\frac{1}{(m-1)h^n} \sum_{j \neq i}^{m-1} K\left(\frac{\mathbf{x_i} - \mathbf{x_j}}{h}\right) \right], \tag{4.53}$$

which should result in a density estimate $\hat{p}(x)$ that is close to the true density $p(x)$ in terms of the Kullback-Leibler divergence [Silverman, 1986; Kullback and Leibler, 1951][6]. This is a measure of closeness between two probability distributions in terms of information. For $p(\mathbf{x})$ and $\hat{p}(\mathbf{x})$, it is defined,

$$\text{KL}(p, \hat{p}) = \int_{\mathbb{R}^n} p(\mathbf{x}) \log \left[\frac{p(\mathbf{x})}{\hat{p}(\mathbf{x})}\right] d\mathbf{x}. \tag{4.54}$$

The result of choosing h by minimising KL is shown in Figure 4.10.

The discussions of this section have generally concerned vector $\mathbf{x}$ of n-dimensions. In small numbers of dimensions the kernel density estimate behaves well and quite intuitively. However, the intuition breaks down as the number of dimensions increases. Many authors note this point [Silverman, 1986; Scott, 1992; Bishop, 1995; Duda et al., 2001], the account

[6]The Kullback-Leibler divergence is called a divergence as it is not a true distance measure. It is, for example, not symmetric.

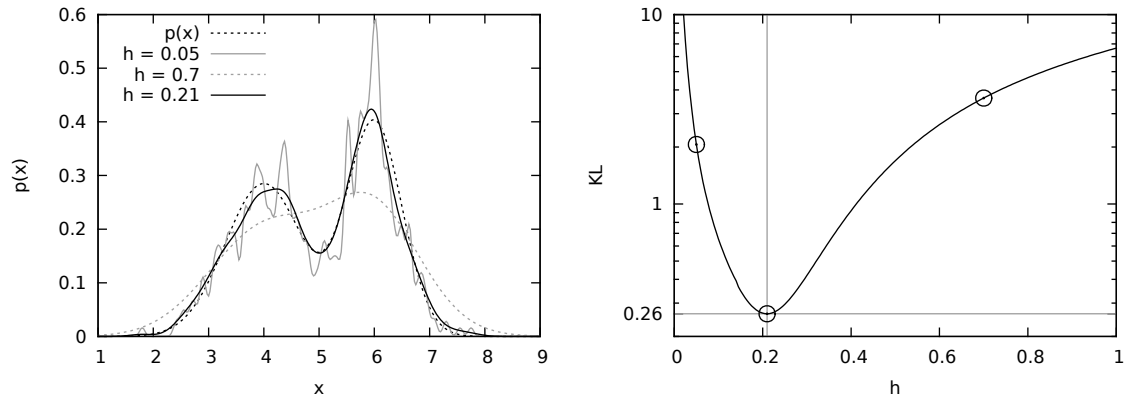

Figure 4.10: The bias and variance trade off in h. The sample is $m = 500$ points from $\frac{1}{2}(N(4, 0.7)+N(6, 0.5))$. *Left:* the distribution and kernel density estimates using a standard normal kernel, for high variance $h = 0.05$; high bias $h = 0.7$; and min $\mathrm{KL}(p, \hat{p})$ $h = 0.21$. *Right:* $\mathrm{KL}(p, \hat{p})$ for $h \in [0, 1]$, circles mark the Kullback-Leibler divergence of the three estimates. Note the log-scale on the vertical axis. The rise of KL to the left of the minimum is due to increasing variance of $\hat{p}(\mathbf{x})$, the rise to the right of the minimum is due to increasing bias of $\hat{p}(\mathbf{x})$.

of [Scott, 1992] is perhaps the most detailed. An example from [Scott, 1992] is replicated here, consider that the underlying distribution $p(\mathbf{x})$ is the standard n-dimensional normal distribution (4.38). The most probable point is clearly $\mathbf{x} = \mathbf{0}$, and the contours of equal probability are $(n-1)$-dimensional spheres. One can then ask, what is the probability that a point drawn randomly from $p(\mathbf{x})$ falls within the 1% contour. Then note,

$$\frac{p(\mathbf{x})}{p(\mathbf{0})} = e^{-\mathbf{x}^T\mathbf{x}/2}, \qquad \text{and so} \qquad -2\ln\frac{p(\mathbf{x})}{p(\mathbf{0})} = \sum_{i=0}^{n} x_i^2 \sim \chi^2(n), \tag{4.55}$$

and as a consequence the probability that a point lies within the 1% contour is given by,

$$\mathbb{P}\left(\frac{p(\mathbf{x})}{p(\mathbf{0})} \geq 1/100\right) = \mathbb{P}\left(\chi^2(n) \leq -2\ln(1/100)\right). \tag{4.56}$$

The above gives the probability that a random point drawn from $p(\mathbf{x})$ will not fall in the 1% tails. For $n = 1$ the probability is 0.9996; for two dimensions the probability is 0.9928; at 10 dimensions the probability drops to 0.36; and in twenty dimensions the probability is just 0.02. Thus, as the dimensions increase the majority of points lie in the tails of the distribution. The implication is that one requires exponentially larger samples to get good estimates in high dimensions. [Silverman, 1986] illustrates this by calculating the approximate number points required to achieve a relative mean square error at the origin of less than 0.1. Thus, [Silverman, 1986] calculates the sample size required for $\mathrm{MSE}_{\mathbf{0}}/p(\mathbf{0})^2 < 0.1$, for again $p(\mathbf{x})$ the standard multivariate normal and the kernel is also standard multivariate normal with optimally chosen $h_\star$. In 1 dimension only 4 samples are needed; 2 dimensions 19 samples; 5 dimensions 768 and in 10 dimensions 842000 samples are needed. This renders the kernel density estimator inappropriate in many dimensions unless one has a very large sample size.

4.3.1 Anomaly Classification Using Density Estimation

Methods described in [Duda et al., 2001; Bishop, 1994] can address *type 1* anomaly detection with kernel density estimation. The procedure is as follows: given training data $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_{m-1}} \in \mathbb{R}^n$ classify a new data point $\mathbf{v} \in \mathbb{R}^n$ to be in class $\mathcal{C}_1$ if the $\mathbf{v}$ is thought to come from the same distribution as the $\mathbf{x_i}$ or to be in class $\mathcal{C}_2$ if $\mathbf{v}$ is thought to be anomalous.

In [Duda et al., 2001; Bishop, 1994] this class assignment problem is formulated using Bayesian statistics. The new data point $\mathbf{v}$ may belong to either class with prior probabilities $\mathbb{P}(\mathcal{C}_1)$ and $\mathbb{P}(\mathcal{C}_2)$ and $\mathbb{P}(\mathcal{C}_1) + \mathbb{P}(\mathcal{C}_2) = 1$. To minimise the probability of misclassification, $\mathbf{v}$ is assigned to the class with the largest posterior probability $\mathbb{P}(\mathcal{C}_i|\mathbf{v})$. So, assign $\mathbf{v}$ to $\mathcal{C}_1$ if $\mathbb{P}(\mathcal{C}_1|\mathbf{v}) > \mathbb{P}(\mathcal{C}_2|\mathbf{v})$. Bayes theorem states:

$$\mathbb{P}(\mathcal{C}_i|\mathbf{v}) = \frac{p(\mathbf{v}|\mathcal{C}_i)\mathbb{P}(\mathcal{C}_i)}{p(\mathbf{v})}, \tag{4.57}$$

and can be used to calculate the posterior probabilities, so assign $\mathbf{v}$ to class $\mathcal{C}_1$ when:

$$p(\mathbf{v}|\mathcal{C}_1)\mathbb{P}(\mathcal{C}_1) > p(\mathbf{v}|\mathcal{C}_2)\mathbb{P}(\mathcal{C}_2). \tag{4.58}$$

This condition defines a *decision boundary* for point in $\mathbb{R}^n$. If the underlying distributions and priors are known, the condition is known as the *Bayes optimal decision boundary*, and it will give the lowest probability of error [Duda et al., 2001]. The task is then to estimate the quantities $p(\mathbf{v}|\mathcal{C}_1)$ and $p(\mathbf{v}|\mathcal{C}_2)$. The distribution involving class $\mathcal{C}_1$ may be modelled by a kernel density estimation of the training data. By definition nothing is known about the distribution of anomalous data and so the simplest approach is to assume a uniform distribution across a large region of input space. The prior probabilities $\mathbb{P}(\mathcal{C}_1)$ and $\mathbb{P}(\mathcal{C}_2)$ allow one to incorporate domain information regarding the occurrence of either class. However, generally in anomaly detection nothing is known about occurrence anomalies apart from the assumption that they are expected to occur at a lesser rate than non-anomalous data. A safe option, which weights towards false positives (classifying normal when actually anomalous) is to assume equal prior probability of $\mathcal{C}_1$ and $\mathcal{C}_2$. With this formulation condition (4.58) is equivalent to applying a threshold to the estimated probability density of the training data (Figure 4.11). If the distribution for the anomalous data is assumed to be uniform with probability α, then the classification decision for a point $\mathbf{v}$ is as follows:

$$\text{Classification}(\mathbf{v}) = \begin{cases} \text{Normal} & \text{if } \alpha < \frac{1}{mh^n} \sum_{i=0}^{m-1} K\left(\frac{\mathbf{v}-\mathbf{x_i}}{h}\right) \\ \text{Anomaly} & \text{otherwise.} \end{cases} \tag{4.59}$$

Since the kernel density estimator will converge on the underlying distribution, the decision boundary defined by the above condition will also converge on the Bayes optimal decision boundary. It may seem unjustified to assume a uniform distribution of anomalous data, certainly if the anomalous distribution was peaked in the centre of Figure 4.11 then this approach would fail. As a consequence, this approach necessarily requires some spatial difference between the normal and anomalous data and so the dimensions of the density

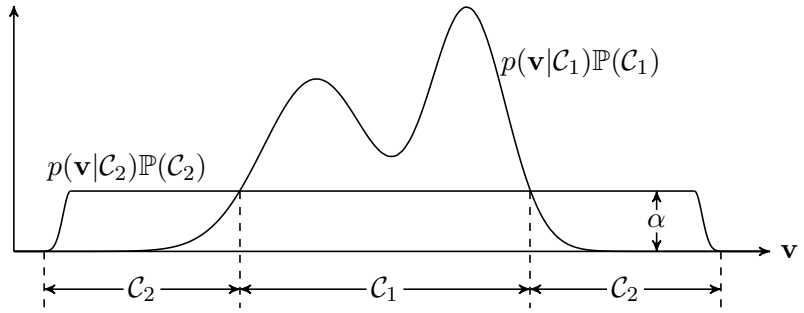

Figure 4.11: Formalism for anomaly classification of a new datapoint, adapted from [Bishop, 1994]. The input space for $\mathbf{v}$ is divided into regions which are classified as either class $\mathcal{C}_1$ or class $\mathcal{C}_2$, dependent on $p(\mathbf{v}|\mathcal{C}_1)\mathbb{P}(\mathcal{C}_1) > p(\mathbf{v}|\mathcal{C}_2)\mathbb{P}(\mathcal{C}_2)$. The threshold α defines the anomalous distribution. The dashed lines that divide the assignment of input space to either $\mathcal{C}_1$ or $\mathcal{C}_2$ are known as the *decision boundaries*.

estimate must be chosen appropriately. All the classification decision really states is that if a new point $\mathbf{v}$ falls in a region of low density in the density estimate (i.e. $\hat{p}(\mathbf{v}) < \alpha$), then it is likely to be anomalous. Stated in this manner the approach is simple but reasonable.

4.3.2 Distributions varying in Time

Regarding the second problem of *type 2* anomaly detection, first consider the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$, then let $\mathbf{x}_i$ occur at time t_i. Further, assume that no two samples occur at identical times and $t_i < t_j \Leftrightarrow i < j$ [7]. Finally, assume that the $\mathbf{x}_i$ are drawn from a potentially time varying distribution $p(\mathbf{x}, t)$. Define a sliding window ω_i of length w to be,

$$\omega_i = \{\mathbf{x}_{i-w+1}, \ldots, \mathbf{x}_i\} \tag{4.60}$$

Then it is possible to understand the rate of change of a distribution by comparing estimates $\hat{p}(\mathbf{x}|\omega_i)$ with $\hat{p}(\mathbf{x}|\omega_j)$. This may be done for example via the Kullback-Leibler divergence (4.54): $\mathrm{KL}(\hat{p}(\mathbf{x}|\omega_i), \hat{p}(\mathbf{x}|\omega_j))$. Then, a threshold on the rate of change of information divergence, ν, can be defined:

$$\frac{\mathrm{KL}(\hat{p}(\mathbf{x}|\omega_i), \hat{p}(\mathbf{x}|\omega_j))}{j-i} < \nu \qquad j > i. \tag{4.61}$$

Alternatively, the k-step difference operator can be applied:

$$\Delta_k \hat{p}(\mathbf{x}|\omega_i) = \hat{p}(\mathbf{x}|\omega_{i+k}) - \hat{p}(\mathbf{x}|\omega_i). \tag{4.62}$$

Then, the integrated square k-step difference ISD_k and the max square k-step difference MSD_k, are possible distance measures:

$$\mathrm{ISD}_k(i) = \int_{\mathbb{R}^n} \left[\Delta_k \hat{p}(\mathbf{x}|\omega_i)\right]^2 d\mathbf{x} \tag{4.63}$$

$$\mathrm{MSD}_k(i) = \max_{\mathbf{x}} \left[\Delta_k \hat{p}(\mathbf{x}|\omega_i)\right]^2. \tag{4.64}$$

[7] In a situation with many observations per time step then they may be combined in the manner of summation (4.8) and considered as a single observation

Then thresholds can be applied to the ISD_k and MSD_k. The algorithm developed in this chapter and described in Section 4.6 will perform *type 2* anomaly detection, however no explicit comparison is made to any existing technique at this task. As a consequence the definitions regarding detection of *type 2* anomalies are not developed any further here. However, density estimates on the above presented sliding window form a key part of this chapter and are essential to the final discussions of *type 2* detection given in Section 4.7.

4.4 The Receptor and Classical Anomaly Detection

The beginning of the previous section demonstrated that the solution to $\rho_t(\mathbf{x})$ recurrence (4.35) had the form of a kernel density estimate. This section first develops these ideas further, and then draws analogies between the negative feedback barrier β (Figure 4.1) and the anomalous distribution threshold α (Figure 4.11)

4.4.1 Kernel Density Estimation and the Receptor

Consider $\rho_t(\mathbf{x})$ defined in (4.35), its solution may converted into a probability density function. First, take the lattice $\delta_i \to 0$, so that $\rho_t(\mathbf{x})$ is evaluated at every point in $\mathbb{R}^n$. Then, choose all u_t identical to scale $\rho_t(\mathbf{x})$ to sum to 1,

$$\rho_t(\mathbf{x}) = \frac{b-1}{(b^t-1)h^n} \sum_{i=0}^{t-1} b^{t-1-i} K\left(\frac{\mathbf{x}-\mathbf{x}_t}{h}\right), \tag{4.65}$$

The above is a probability density function and it is a weighted kernel density estimate. The point u_{t-1} has the largest weight $\frac{b-1}{(b^t-1)h^n}$ then each previous sample point is weighted b times the last. The estimate $\rho_t(\mathbf{x})$ is denoted the *decay density estimate*, and has the advantage that it may be evaluated in an "on-line" manner via recursion (4.35). The concept of "on-line" or recursive density estimators is not new, for example see [Wolverton and Wagner, 1969; Hall and Patil, 1994] which use a uniform weighting scheme. The geometric weighting in the decay density estimate requires further investigation, note the following limits,

$$\lim_{b\to 1} \frac{b-1}{b^t-1} = \frac{1}{t}, \qquad \lim_{t\to\infty} \frac{b-1}{b^t-1} = 1-b. \tag{4.66}$$

The first limit shows that the standard kernel density estimated is retrieved when $b \to 1$. The second limit is the infinite sum of the sequence $\{b^t\}$, the relevance can be seen if this weighting is directly applied to (4.35) an setting $u_t = 1-b$:

$$\rho_{t+1}(\mathbf{x}) = b\rho_t(\mathbf{x}) + \frac{(1-b)}{h^n} K\left(\frac{\mathbf{x}-\mathbf{x}_t}{h}\right). \tag{4.67}$$

The weighting scheme is known as an *exponentially weighted moving average* or an *exponential smoothing* of $\{\mathbf{x}_t\}$. It is a common concept in the analysis of time series [Kendall and Ord, 1990] and particularly in the analysis of financial time series [Tsay, 2002]. There are a wide range of well developed exponential smoothing techniques, [Gardner, 2006] reviews the state of the art and provides a classification of fifteen exponential smoothing variants with

respect to underlying trends and seasonality. The smoothing in the decay density estimate is the simplest class discussed in [Gardner, 2006]; certainly assumptions regarding trending or seasonality can not be made about the sequence $\{\mathbf{x}_t\}$.

A comparison is made between the decay density estimate and a standard kernel density estimate. To focus on the influence of exponential smoothing the discussions regard the point-wise expectations and variances of each estimate. Let $\mu_t = K\left(\frac{\mathbf{x}-\mathbf{x}_t}{h}\right)$, thus μ_t contribution of the sample to point $\mathbf{x}$ after being passed through the kernel. As such, μ_t has an expectation $\mathbb{E}\mu_t$ and a variance var μ_t which are dependent on the sample and the kernel in a manner discussed in Section 4.3. It is assumed that the kernel does not alter the independence of the sample, all μ_t are independent. The weighting techniques in the two methods of density estimation can be compared on sequence $\{\mu_t\}$. Let ρ_t be the exponential weighted smoothing of μ_t, and w_t be the equally weighted smoothing of μ_t by the use of a sliding window ω_t, and so,

$$\rho_{t+1} = b\rho_t + (1-b)\mu_t \tag{4.68}$$

$$w_{t+1} = \frac{1}{w}\sum_{k=t-w+1}^{t}\mu_k. \tag{4.69}$$

The equal weighting recursive density estimates presented in [Wolverton and Wagner, 1969; Hall and Patil, 1994] do not require a sliding window, and are generalisations of the form,

$$\hat{p}_{t+1}(\mathbf{x}) = \frac{t}{t+1}p_t(\mathbf{x}) + \frac{1}{t+1}K\left(\frac{\mathbf{x}-\mathbf{x}_t}{h}\right), \tag{4.70}$$

and so at time t all previous samples are equally weighted. Since the concern here regards a potentially infinite sequence of samples $\ldots \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1} \ldots$ it is necessary for an estimator to be able to "forget" the details of old samples and so a sliding window is necessary for w_t. The expectations and variances for ρ_t and w_t can be calculated, first asymptotically as $t \to \infty$, such that initial ρ_0 can be ignored,

$$\text{As } t \to \infty \begin{cases} \mathbb{E}\rho_t = \mathbb{E}\mu_t & \text{var } \rho_t = \left[\frac{1-b}{1+b}\right] \text{var } \mu_t \\ \mathbb{E}w_t = \mathbb{E}\mu_t & \text{var } w_t = \frac{\text{var } \mu_t}{w} \end{cases}. \tag{4.71}$$

Thus, both ρ_t and μ_t are unbiased estimators of the expectation of μ_t, and have equal variance when $\frac{w-1}{w+1} = b$. Regarding convergence to the above asymptotic values, one must wait for w time steps for the sliding window to fill before the results for w_t are valid. And, in the case of ρ_t one must wait of the order $\log \epsilon / \log b$ for ρ_t to come within ϵ of $\mathbb{E}\mu_t$ dependent on initial ρ_0, which in general will be a significantly longer wait than the time window that set b. An obvious solution is to re-weight such that all weights sum to 1 for all t,

$$\rho^\star_{t+1} = b\left[\frac{b^t - 1}{b^{t+1} - 1}\right]\rho^\star_t + \left[\frac{b-1}{b^{t+1}-1}\right]\mu_t \tag{4.72}$$

$$w^\star_{t+1} = \frac{1}{t - \max\{1, t-w\}}\sum_{k=\min\{0,t-w+1\}}^{t}\mu_k. \tag{4.73}$$

$\rho_t^\star$ is weighted to ensure the geometric series with terms b^t sums to 1. Note there is no need to set $\rho_0^\star$, the calculation of $\rho_1^\star$ involves $(b^0-1)\rho_0^\star$, as can be seen from the solution,

$$\rho_t^\star = \frac{b-1}{b^t-1}\sum_{k=0}^{t-1} b^{t-1-k}\mu_k. \tag{4.74}$$

The expectation and variance for the newly weighted $\rho_t^\star$ and $w_t^\star$, are as follows,

$$\text{For } t \geq 1 \begin{cases} \mathbb{E}\rho_t^\star = \mathbb{E}\mu_t & \text{var } \rho_t^\star = \left[\frac{(1-b)(1+b^t)}{(1+b)(1-b^t)}\right] \text{var } \mu_t \\ \mathbb{E}w_t^\star = \mathbb{E}\mu_t & \text{var } w_t^\star = \frac{\text{var } \mu_t}{\max\{t,w\}} \end{cases}. \tag{4.75}$$

Both var $\rho_t^\star$ and var $w_t^\star$ decay with time and are asymptotically equivalent when $b = (w-1)/(w+1)$ as before. The convergence of variances of var $\rho_t^\star$ and var $w_t^\star$ are very comparable with var $\rho_t^\star < 1.3131(\text{var } w_t^\star)$ for all $t \geq 1$. This is demonstrated by Theorem 4.2 which compares var $\rho_t^\star$ and $w_t^\star$ through the following definition:

$$\xi_t = \text{var } \rho_t^\star / \text{var } w_t^\star. \tag{4.76}$$

Theorem 4.2. *Moving averages $\rho_t^\star$ and $w_t^\star$ with $b = (w-1)/(w+1)$ and $w \geq 2$ for all $t \geq 1$ are such that var $\rho_t^\star \geq$ var $w_t^\star$ and $\xi_t < (e^2+1)/(e^2-1)$.*

Proof. Consider the two stages of ξ_t:

$$\xi_t = \begin{cases} (t/w)[(1+b^t)/(1-b^t)] & \text{for } t < w \\ (1+b^t)/(1-b^t) & \text{for } t \geq w \end{cases}. \tag{4.77}$$

Demonstrating that $\Delta\xi_t > 0$ for $1 \leq t < w$ and noting that $\Delta\xi_t < 0$ for $t \geq w$ will yield the desired result with the maximum value of ξ_t occurring at $t = w$. First, consider $\Delta\xi_t$ with $1 < t < w$:

$$\Delta\xi_t = \frac{t+1}{w}\left[\frac{2b^{t+1}}{1-b^{t+1}}+1\right] - \frac{t}{w}\left[\frac{2b^t}{1-b^t}+1\right] \tag{4.78}$$

$$= \frac{2tb^t}{w}\left[\frac{b}{1-b^{t+1}} - \frac{1}{1-b^t}\right] + \frac{1}{w}\left[\frac{2b^{t+1}}{1-b^{t+1}}+1\right]. \tag{4.79}$$

The requirement is $\Delta\xi_t > 0$ and so it is necessary to show:

$$2tb^t\left[\frac{1-b}{(1-b^t)(1-b^{t+1})}\right] < \frac{2b^{t+1}}{1-b^{t+1}}+1, \tag{4.80}$$

$$t < \left[\frac{1-b^t}{1-b}\right]\frac{1+b^{t+1}}{2b^t}. \tag{4.81}$$

Translating $b = (w-1)/(w+1)$, the condition for $\Delta\xi_t > 0$ becomes:

$$t < \frac{1}{4}\left[\frac{(w+1)^{2t+1}-(w-1)^{2t+1}}{(w^2-1)^t}\right] - \frac{1}{2}. \tag{4.82}$$

Then, note the following:

$$\frac{(w+1)^{2t+1}-(w-1)^{2t+1}}{(w^2-1)^t} > \frac{(w+1)^{2t+1}-(w-1)^{2t+1}}{w^{2t}} \tag{4.83}$$

$$= \frac{1}{w^{2t}}\left[\sum_{k=0}^{2t+1}\binom{2t+1}{k}w^k - \sum_{i=0}^{2t+1}\binom{2t+1}{i}w^i(-1)^{2t+1-i}\right] \tag{4.84}$$

$$= \frac{1}{w^{2t}}\sum_{k=0}^{2t+1}\binom{2t+1}{k}w^k(1-(-1)^{2t+1-k}) \tag{4.85}$$

$$= \frac{2}{w^{2t}}\sum_{k=0}^{t}\binom{2t+1}{2k}w^{2k} = 2\sum_{k=0}^{t}\binom{2t+1}{2k}w^{2k-2t}. \tag{4.86}$$

Combining the above result with the condition for $\Delta\xi_t > 0$ in inequality (4.82) produces the following:

$$2t+1 < \sum_{k=0}^{t}\binom{2t+1}{2k}w^{2k-2t}, \tag{4.87}$$

which can be seen to hold by consideration of $k=t$ term in the summation, and so $\Delta\xi_t > 0$ for $t < w$. As $\Delta\xi_t < 0$ for $t \geq w$ the maximum value of ξ_t occurs at $t = w$. Consideration of the behaviour of b^w as w grows provides a bounding for ξ_w. First, note that b^w is increasing as $\Delta b^w > 0$:

$$\left[\frac{w-1}{w+1}\right]^w < \left[\frac{w}{w+2}\right]^{w+1}. \tag{4.88}$$

Second, b^w increases to the following asymptote:

$$\lim_{w\to\infty} b^w = \lim_{w\to\infty}\left(1-\frac{2}{w+1}\right)^w = e^{-2}. \tag{4.89}$$

Thus, $b^w < e^{-2}$ for $2 \leq w < \infty$ and this bounding implies the following bound on ξ_w:

$$\xi_w = \frac{1+b^w}{1-b^w} < \frac{(e^2+1)}{(e^2-1)} \approx 1.3130. \tag{4.90}$$

Finally, noting that the variances are equal: var $\rho_1^\star$ = var $w_1^\star$ = var μ_t at $t = 1$ and again as $t \to \infty$ then the behaviour of $\Delta\xi_t$ shows that var $\rho_t^\star \geq$ var $w_t^\star$. □

A demonstration of the decay density estimate in comparison with the kernel density estimate is given in Figure 4.12. Shown are decay density estimates $\rho_a(\mathbf{x})$ and $\rho_b(\mathbf{x})$, and a kernel density estimate on a sliding window $\hat{p}(x|\omega_t)$ with window $w = 1000$. Estimates ρ_a and $\hat{p}$ are set to have equal variance, so for ρ_a, $b = (w-1)/(w+1)$. The estimate ρ_b is set to have b much closer to 1 and has equivalent variance to a windowed estimate with $w = 3 \times 10^4$. The Kullback-Leibler divergence of the standard estimate contains large discontinuous jumps (vertical lines in Figure 4.12), these occur when samples which happen to be correlated in time and space drop in and out of the sliding window. The estimate ρ_b requires a smaller h to achieve the lower Kullback-Leibler divergence, this is due to a large sample size contributing to the estimate. As such, it seems sensible to adjust h

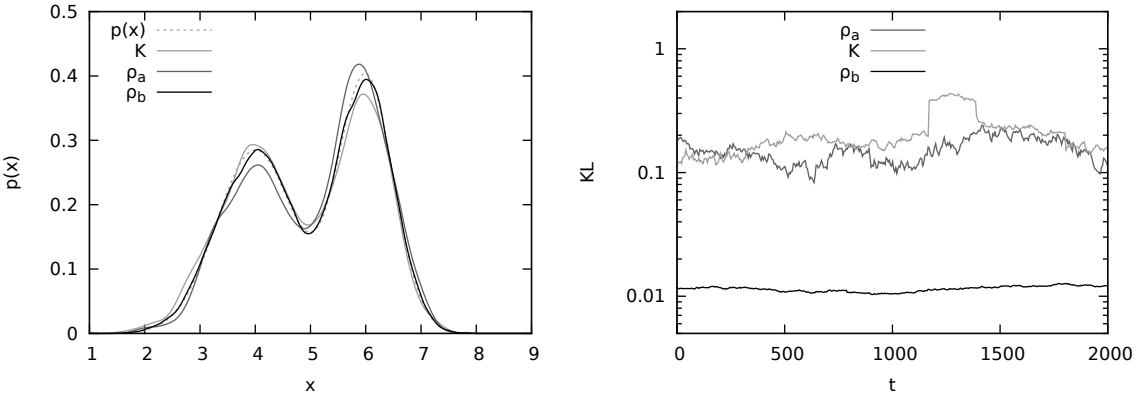

Figure 4.12: A comparison between the decay density estimates $\rho_a(x)$, $\rho_b(x)$ and a kernel density estimate on a sliding window $\hat{p}(x|\omega_t)$ length $w = 1000$. Samples are drawn sequentially from $\frac{1}{2}(N(4, 0.7) + N(6, 0.5))$ (the same as Figure 4.10). All estimates use a standard normal kernel with $h = 0.18$ for ρ_a and $\hat{p}$ and $h = 0.085$ for ρ_b. K (light grey) marks $\hat{p}(x|\omega_{t,w})$; ρ_a (grey) has $b = 0.9980$; ρ_b (black) has $b = 0.9999$. Results are taken after a settle time of 3×10^4 steps. *Left:* The estimates at graph time $t = 2000$. *Right:* The Kullback-Leibler divergence between the estimates and true distribution at each time step. Note the discontinuous jumps in the $KL(p, \hat{p}(x|\omega_{t,w}))$ occur when a succession samples correlated in space drop in and out of the sliding window.

during the recursive evaluation of the density estimate. [Hall and Patil, 1994] discusses an on-line method for adjusting the h: the approach is to estimate the standard deviation of a streaming sample $\ldots \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1}, \ldots$ and to use this in conjunction with the "normal-reference" method from [Silverman, 1986] given in (4.52). There are many well known algorithms for computing the standard deviation or variance in an on-line manner (e.g. see [West, 1979; Knuth, 1979]), however these algorithms require the sample size in advance. An alternative is to note that the recurrence $\rho_t^\star$ computes an expectation, thus to compute the variance one can use $\rho_t^\star$ to compute $\mathbb{E}\mathbf{x}$ and $\mathbb{E}\mathbf{x}^2$ and so compute the variance. A further advantage is that this variance will have been computed with the appropriate weightings for the decay density estimate. [Bruce, 1969] takes a similar approach to recursively evaluate the estimate but with a recurrence of the form ρ_t in (4.68).

Some discussions of exponential smoothing state that the weightings are of benefit when there are underlying shifts or trends in the data. This is usually based on the intuition that the most recent data points have the greatest weight and so track changes better than an equal weighting scheme. To investigate this effect, consider that $\mu_t \sim \text{Bernoulli}(q_t)$, i.e. $\mu_t = 1$ with probability q_t and $\mu_t = 0$ with probability $1 - q_t$, and so, $\mathbb{E}\mu_t = q_t$ and $\text{var}\ \mu_t = q_t(1 - q_t)$. In general, this variance will be greater than when μ_t was set by sample and kernel, and it represents the case when $h \to 0$ and the kernel becomes like a delta function. Then q_t is set as follows,

$$q_t = \begin{cases} u & \text{for } t < t_c \\ a(t - t_c) + u & \text{for } t \geq t_c \end{cases}, \tag{4.91}$$

so q_t changes from a constant u to linearly increasing at rate a at time t_c. The expectation

of $w_t^\star$ can be calculated before, during and after the transition from constant to linear q_t, assume that $t_c > w$, then,

$$\mathbb{E}w_t^\star = \begin{cases} u & \text{for } t < t_c \\ u + \frac{a}{2w}\left((t-t_c)^2 - (t-t_c)\right) & \text{for } t_c \leq t < t_c + w \\ u + a\left(t - tc - \frac{w+1}{2}\right) & \text{for } t \geq t_c + w \end{cases}. \tag{4.92}$$

During the transition the expectation of $w_t^\star$ quadratically increases toward the target expectation, after the transition the expectation is simply the $\mathbb{E}\mu_t$ at the mid point of the window. The expectation of $\rho_t^\star$ can be calculated in a similar manner, first for $t \geq t_c$,

$$\mathbb{E}\rho_t^\star = \frac{b-1}{b^t - 1}\sum_{k=0}^{t-1}\mathbb{E}\mu_k b^{t-1-k} = u + \frac{a}{\phi_b^t}\left[\left(\sum_{k=t_c}^{t-1} k b^{t-1-k}\right) - t_c\phi_b^{t-t_c}\right] \tag{4.93}$$

$$= u + \frac{a}{1-b^t}\left[\frac{b^{t-t_c} - b}{1-b} + (t-1) - t_c\right]. \tag{4.94}$$

The above summation is the discrete convolution (see (4.122) for details) between the sequence $\{t\}$ and $\{b^t\}$ and is readily computed by application of the Z-transform [Elaydi, 1995]. Then combining with $t < t_c$ gives,

$$\mathbb{E}\rho_t^\star = \begin{cases} u & \text{for } t < t_c \\ u + \frac{a}{1-b^t}\left[\frac{b^{t-t_c}-b}{1-b} + (t-1) - t_c\right] & \text{for } t \geq t_c \end{cases}. \tag{4.95}$$

Once the transition has occurred $\rho_t^\star$ and $w_t^\star$ are no longer unbiased estimators of μ_t. The expectations of $\rho_t^\star$ and $w_t^\star$ may be more readily compared by calculating an approximation of a Taylor series of $\mathbb{E}\rho_t^\star$ around $t = t_c$. To the second order, with additional approximation that $\ln(b) \approx b - 1$, this gives,

$$\mathbb{E}\rho_t^\star \approx u + \frac{a(1-b)}{2(1-b^{tc})}(t-t_c)^2. \tag{4.96}$$

This is of similar form to $\mathbb{E}w_t^\star$ during the transition, but without the linear term which acts to arrest the increase of $\mathbb{E}w_t^\star$. The exponential smoothing of $\rho_t^\star$ is able to react quicker to changes in μ_t. However, this approximation for $\rho_t^\star$ is only valid near t_c, and in reality $\rho_t^\star$ and $\mathbb{E}w_t^\star$ track very similarly. This is especially so when b and w are chosen to give equivalent variance to $\rho_t^\star$ and $w_t^\star$ during the constant phase of q_t. Figure 4.13 demonstrates this point with a numerical comparison of $\mathbb{E}\rho_t^\star$ and $\mathbb{E}w_t^\star$.

To summarise, the decay density estimate can have very comparable performance to the standard kernel density estimate on a sliding window in terms of expectation and convergence of variance. The decay density estimate has the advantage that it can be applied recursively with no requirement to store a window of data. Moreover, the decay density estimate does not suffer from the discontinuous influences of a sliding window (see Figure 4.12).

In the following section, negative feedback is returned to the picture and analogies between the negative feedback barrier β and the anomaly threshold α are presented.

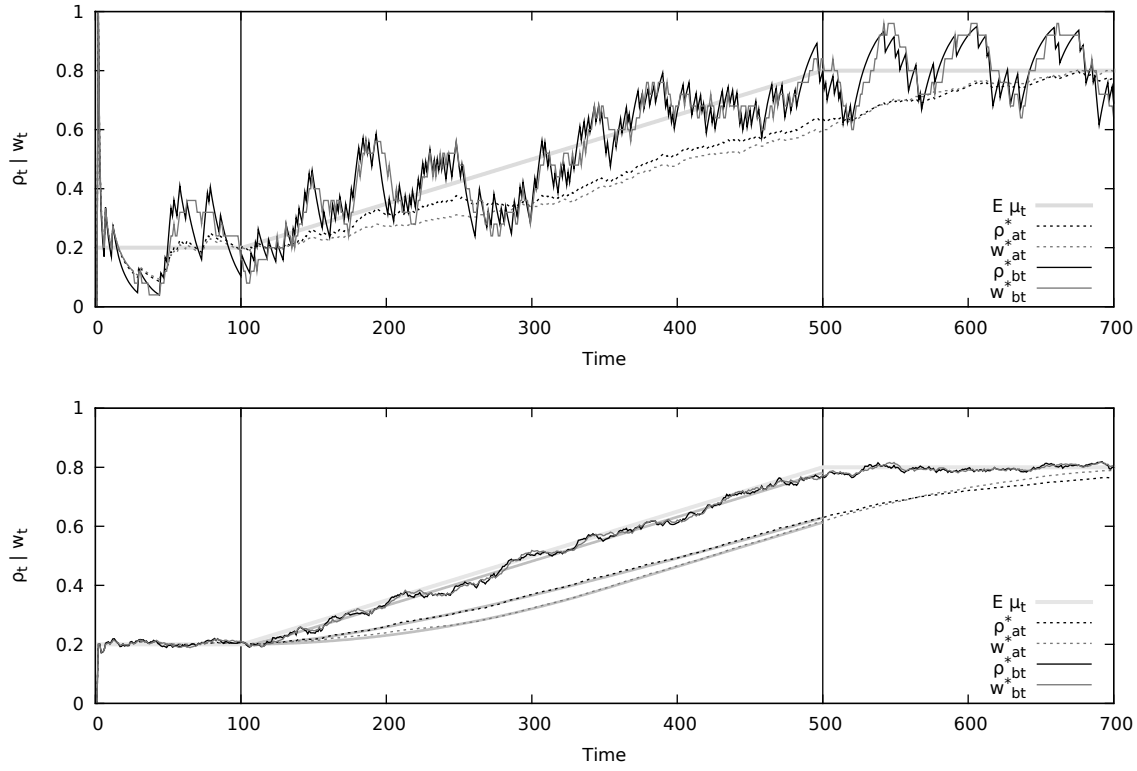

Figure 4.13: Demonstration of $w_t^\star$ and $\rho_t^\star$ when tracking μ_t with q_t. With $u = 0.2$; $a = 0.0015$; $t_c = 100$. q_t is set a constant 0.8 for $t > 500$. Comparisons are made between a long ($w_{at}^\star$) and a short window ($w_{bt}^\star$), length $w_a = 250$ and $w_b = 25$ respectively. The two exponentially weighted $\rho_{at}^\star$ and $\rho_{bt}^\star$ have $b_a = \frac{w_a - 1}{w_a + 1}$ and $b_b = \frac{w_b - 1}{w_b + 1}$ set accordingly. *Top:* Shows a single trajectory of $w_{at}^\star, w_{bt}^\star, \rho_{at}^\star, \rho_{bt}^\star$, note the trajectories display the $1/t$ decrease in variance. *Bottom:* Gives the mean over 50 runs, the thick lightest grey line marks $\mathbb{E}\mu_t$, the dashed lines are the mean $\rho_{at}^\star$ and $w_{at}^\star$, the high variance line as $\rho_{bt}^\star$ and $w_{bt}^\star$ which still exhibit the stochastic noise. Solid grey lines beneath give the theoretical expectations. The equivalent variance $w_t^\star$ and $\rho_t^\star$ have similar performance.

4.4.2 Bayesian Anomaly Classification and the Receptor

Without negative feedback, the receptor positions are a density estimate of the input to the receptor lattice. The inclusion of negative feedback would result in a negative feedback level that is relatively constant across the lattice and that reflects the mass of the density estimate above β. The density estimate would be lost, though some information regarding the peaks of the underlying distribution would remain (Figure 4.14). However, if the diffusion rate of the negative feedback was reduced the density estimate could be retained. Figure 4.14 demonstrates this point by plotting the recurrences given in (4.4) at $t = 100$ for two choices of K_N width: $h = 100$ and $h = 1$. The reduced kernel width represents diffusion limited negative feedback. The equilibrium conditions calculated in Section 4.2 now apply in the diffusion limited case. Thus, with appropriate parameters $p_t(\mathbf{x})$ will oscillate around β and,

$$n_t(\mathbf{x}) \approx \rho(\mathbf{x}) - \beta(1 - b), \tag{4.97}$$

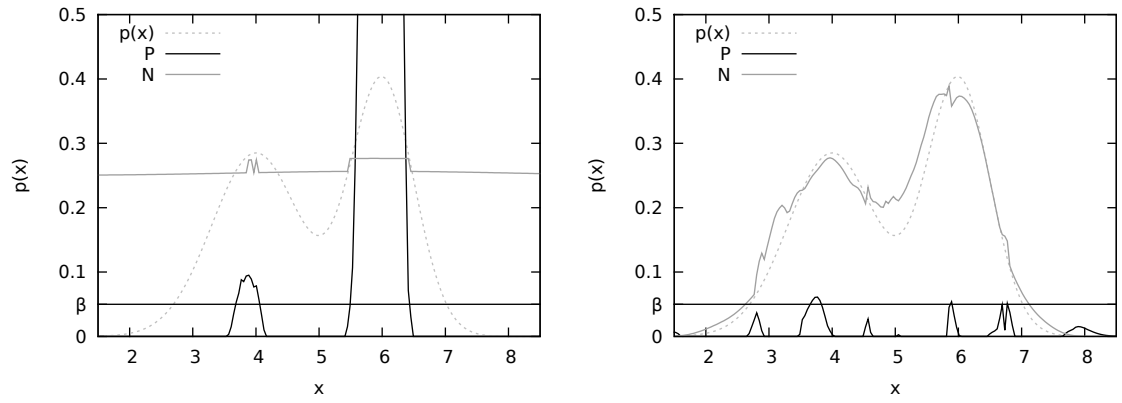

Figure 4.14: Retaining the density estimate in a lattice of receptors. The recurrences in (4.4) are iterated from $p_0(\mathbf{x}), n_0(\mathbf{x}) = 0$ till $t = 100$. With parameters: $b = 0.95$; $d = 0.99$; $a = 1$; $g = 0.02$; $\beta = 0.05$. K_S is calculated with a standard normal kernel with $h = 0.21$. The input u_t is drawn independently of time from $\frac{1}{2}(N(4, 0.7) + N(6, 0.5))$, 500 input points are drawn per time step. K_N is calculated with a standard normal kernel for high diffusion $h = 100$ (left) and low diffusion $h = 1$ (right). Note how the low diffusion case retains the information from the distribution.

$\rho(\mathbf{x})$ is the decay density estimate calculated with a kernel that is the convolution of the stimulation and negative feedback kernels. The density estimate is now retained in the negative feedback. For $n_t(\mathbf{x})$ to be a good quality estimate the diffusion kernel must be small so as to reduce the bias. Noting the convolution of two normal distributions is also normal, the negative feedback kernel may be dropped altogether. The width of the K_S kernel can then be set directly by the input $\{\boldsymbol{\gamma}_t\}$.

The thresholding anomaly condition of (4.59) can be reconstructed through a receptor framework. For the sake of motivating argument, the receptor positions and negative feedback are set statically but in agreement with equilibrium positions given Section 4.2. To reiterate, the problem is to classify a point $\mathbf{v}$ normal or anomalous given a training sample $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{m-1}$. This may be addressed by a *training phase* in which the receptor positions and negative feedback are initialised, then a test phase which evaluates $\mathbf{v}$. Again, consider the lattice of receptors in which $\delta_i \to 0$. The receptor position at point $\mathbf{x}$ is written $r_p(\mathbf{x})$ and similarly the negative feedback $r_n(\mathbf{x})$.

Training Phase

Assume the training sample is in input regions $\mathcal{U}_1$ and $\mathcal{U}_2$ (Figure 4.4). The $r_p(\mathbf{x})$ and $r_n(\mathbf{x})$ are set as,

$$r_p(\mathbf{x}) = \begin{cases} \beta & \text{for } \rho(\mathbf{x}) \geq \beta \\ \rho(\mathbf{x}) & \text{for } \rho(\mathbf{x}) < \beta \end{cases} \qquad r_n(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}) - \beta(1-b) & \text{for } \rho(\mathbf{x}) \geq \beta \\ 0 & \text{for } \rho(\mathbf{x}) < \beta \end{cases}, \tag{4.98}$$

with $\rho(\mathbf{x})$ the decay density estimate of the training sample.

Test Phase

In the test phase the receptor position recurrence is applied with test point $\mathbf{v}$ as follows,

$$r_{pt}(\mathbf{x}) = br_p(\mathbf{x}) + K_S(\mathbf{x}, \mathbf{v}) - r_n(\mathbf{x}). \tag{4.99}$$

Then the classification of $\mathbf{v}$ is,

$$\text{Classification}(\mathbf{v}) = \begin{cases} \text{Normal} & \text{if for all } \mathbf{x},\ r_{pt}(\mathbf{x}) < \ell \\ \text{Anomaly} & \text{if there exists } \mathbf{x},\ r_{pt}(\mathbf{x}) \geq \ell. \end{cases} \tag{4.100}$$

To clarify the above condition, note the outcome of (4.99), when $\beta = \alpha$ and $\ell = \max\{K_S\}$.

$$r_{pt}(\mathbf{x}) = \begin{cases} b\rho(\mathbf{x}) + K_S(\mathbf{x}, \mathbf{v}) & \text{for } \rho(\mathbf{x}) < \alpha \\ K_S(\mathbf{x}, \mathbf{v}) & \text{for } \rho(\mathbf{x}) = \alpha \\ K_S(\mathbf{x}, \mathbf{v}) - \rho(\mathbf{x}) - \alpha & \text{for } \rho(\mathbf{x}) > \alpha \end{cases} \tag{4.101}$$

Thus, $\max\{r_{pt}(\mathbf{v})\} = \max\{K_S\} = \ell$ when $\rho(\mathbf{v}) \leq \alpha$ and $\max\{r_{pt}(\mathbf{v})\} < \ell$ otherwise, which describes the thresholding of (4.59). Further, if $b = \frac{m-1}{m+1}$ then the expectations of (4.100) and (4.59) will be identical and the variances related by Theorem 4.2.

The removal of negative feedback diffusion solves the second issue raised in Section 4.1.2, the lattice of receptors now does more than 1-dimensional outlier detection. The removal of negative feedback diffusion must be added to the set A which connects the biological model and this abstracted model. Ignoring the differences in substrate (i.e. the implementation details) it is important to note that the biological models and the qualitative model differ. The biological model evaluates TCR-pMHC bond strength, the average bond strength is communicated to all TCRs via the high diffusion negative feedback. Those TCRs associated with pMHC ligands with a significantly stronger bond than the mean are able to complete kinetic proofreading and produce activations ($r_{pt}(\mathbf{x}) \geq \ell$). Here, TCR-pMHC bond strength is exchanged with the probability of a sample point being drawn at $\mathbf{x}$. With the negative feedback diffusion restricted the receptors now make local comparisons to decide if the sample point at $\mathbf{x}$ is sufficiently improbable to cause $r_{pt}(\mathbf{x}) \geq \ell$. This point is given further discussion in the concluding Section 4.7 and the concluding chapter of this thesis.

With connections and analogies to classical anomaly detection methods established, the chapter returns to analysis of a single receptor. Particularly, to understand how the receptor may dynamically break ℓ.

4.5 The single receptor breaking ℓ

This extends the results of Section 4.2 and calculates conditions for $p_t \geq \ell$. Consequently the input $\{u_t\}$ is no longer constant. First, step increases in $\{u_t\}$ are considered.

4.5.1 Step increases in u_t

This scenario describes a receptor which is at the cyclic equilibrium under $u \in \mathcal{U}_2$ which then undergoes an increase in input to $u_\star > u$ the input then remains at $u_\star$. The details of the cycle are ignored for simplicity, and the scenario is modelled by constant input $u_\star$ and initial conditions $p_0 = \beta$ and $n_0 = u - \beta(1-b)$ Then, the first receptor position delta is,

$$\Delta p_t = p_e(b-1) + u_\star - n_e = u_\star - u, \tag{4.102}$$

and ℓ can be broken in one time step if,

$$u_\star - u \geq \ell - \beta. \tag{4.103}$$

If ℓ is not broken on the first time step the question of whether it will ever be broken can be answered by the location of the maximum in p_t. This requires further consideration of the difference operator:

$$\Delta p_t = p_{t+1} - p_t = b\Delta p_{t-1} - \Delta n_{t-1} \tag{4.104}$$

$$\Delta n_t = n_{t+1} - n_t = d\Delta n_{t-1}. \tag{4.105}$$

The differences are recurrences with interaction matrix (4.10), and have solution:

$$\Delta p_t = b^t \Delta p_0 + \frac{d^t - b^t}{b-d}\Delta n_0 \tag{4.106}$$

$$\Delta n_t = d^t \Delta n_0. \tag{4.107}$$

The time of the maximum in receptor position, $t_\star$, can be found by,

$$\Delta p_t = 0 \qquad \text{which gives} \qquad \frac{\Delta p_0(d-b)}{\Delta n_0} = \frac{d^t - b^t}{b^t}. \tag{4.108}$$

Taking logs:

$$t_\star = \left\lceil \frac{1}{\log d - \log b} \log\left(\frac{\Delta p_0(d-b)}{\Delta n_0} + 1\right) \right\rceil. \tag{4.109}$$

Due to the discrete time steps, the maximum occurs after p_t jumps over the line $p = \frac{u-n}{1-b}$ and so $t_\star$ is the ceiling of the the solution to equation (4.108). The first deltas are:

$$\Delta p_0 = u_\star - u \qquad \Delta n_0 = (u - \beta(1-b))(d-1) + g. \tag{4.110}$$

Then $t_\star$ can be used to locate the maximum $p_\star$. Recall that the maximum occurs when p_t crosses the line $p = (u_\star - n)/(1-b)$. Thus, $p_\star$ can be found by calculating the negative feedback at maximum $n_\star$ using (4.16). As $t_\star$ (4.109) is a logarithm to the base $\frac{d}{b}$ it is then convenient to write $\mathcal{L}(x) \equiv \log_{\frac{d}{b}}(x)$. So, clearly $\mathcal{L}(d) = \mathcal{L}(b) + 1$ and there is the following logarithmic identity, $y^{\mathcal{L}(x)} = x^{\mathcal{L}(y)}$. The identity may be used to swap $u_\star$ and u out of the exponent of d in the solution to n_t. Consequently, the value of p_t at maximum is,

$$p_\star = \frac{1}{1-b}\left[u_\star - \frac{g}{1-d} - \left(u - \beta(1-b) - \frac{g}{1-d}\right)\left(\frac{\Delta p_0(d-b)}{\Delta n_0} + 1\right)^{\mathcal{L}(d)}\right] \tag{4.111}$$

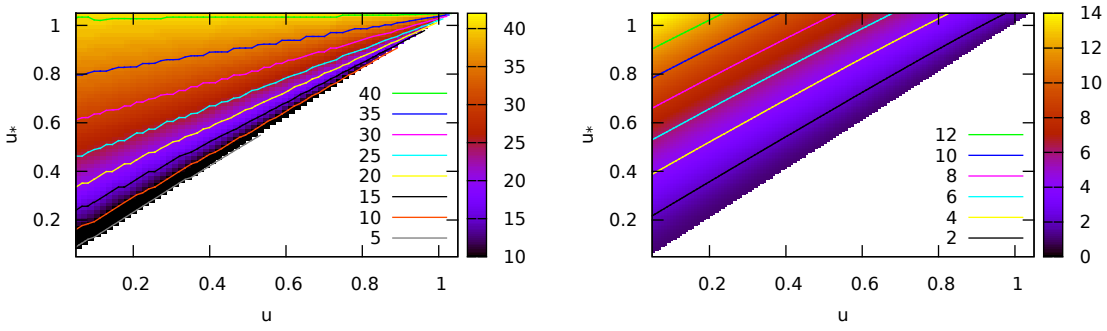

Figure 4.15: The time $t_\star$ (*left*) and location $p_\star$ (*right*) of the receptor position maximum when input is increased from u to $u_\star$. Parameters are set as examples in Section 4.2 ($b = 0.95$; $d = 0.99$; $g = 0.01$; $\beta = 1$). Note the effectively linear contours in both plots, and note the increase in $t_\star$ with increase in $u_\star$.

Note the occurrence of term $u - \beta(1-b) - g/(1-d)$ which describes the difference between u and the upper boundary of $\mathcal{U}_2$, which scales the negative feedback accordingly. Thus, $u_\star$ causes ℓ to be broken if $p_\star \geq \ell$. The time and value of the maximum in p_t are numerically calculated for $u, u_\star \in \mathcal{U}_2$, see Figure 4.15. Note the lengthening of $t_\star$ with increasing $u_\star$, for $u_\star$ on the upper boundary of $\mathcal{U}_2$ the peak is only reached asymptotically as $t \to \infty$. The linear relationship between u, $u_\star$ and $p_\star$ is due to $p_\star = (u_\star - n_\star)/(1-b)$ and the simple relationship $n_0 = u - \beta(1-b)$.

Approximations are now calculated for this step increase in input scenario. The approximations are useful rules of thumb when applying the algorithm, they are certainly of use in the next chapter. An approximation for $t_\star$ can be found by returning to the peak time equation (4.108) and truncating the Taylor series in t around $t = 0$, for $\left(\frac{d}{b}\right)^t$. A second order approximation gives a quadratic in t, since the $t_\star$ must be positive the relevant solution is,

$$t_\star \approx \frac{-1 + \sqrt{1 + 2\frac{\Delta p_0 (d-b)}{\Delta n_0}}}{\log d - \log b}. \tag{4.112}$$

This qualitatively shows the correct behaviour for all t, and gives a good approximation for $t_\star < 30$ for the parameters in Figure 4.15, beyond this point the approximation loses accuracy. Note that the first order approximation $t_\star \approx \frac{\Delta p_0 (d-b)}{\Delta n_0 \log(d/b)}$, still gives good results for $t_\star < 10$. Rather than pursue an approximation for $p_\star$, it is more useful to calculate an approximation for p_t when $p_t \geq \beta$. The Taylor series around $t = 0$, or the Maclaurin series of t, in p_t (4.16) is then,

$$p_t = p_0 + \sum_{k=0}^{\infty} \frac{t^k}{k!} \left[(p_0 + u_\star) \ln(b)^k + \frac{1}{b-d} \left(n_0 \left(\ln(d)^k - \ln(b)^k \right) + g \left[\frac{\ln(d)^k}{d-1} - \frac{\ln(b)^k}{b-1} \right] \right) \right]. \tag{4.113}$$

Taking the first order approximation with additional approximations that $\ln(b) \approx b - 1$ and

$\ln(d) \approx d - 1$ for b and d close to 1, gives the form of the p-recurrence:

$$p_t \approx p_0 + [(p_0 + u_\star)(b-1) - n_0]\, t \tag{4.114}$$
$$= \beta + (u - u_\star)t \tag{4.115}$$

The second order approximation similarly has the form of two applications of the p-recurrence:

$$p_t \approx p_0 + [(p_0 + u_\star)(b-1) - n_0]\, t + \left[(p_0 + u_\star)(b-1)^2 - n_0(b+d-2) - g\right] \frac{t^2}{2} \tag{4.116}$$
$$= \beta + (u_\star - u)t + \left[(\beta + u_\star)(b-1)^2 - (u - \beta(1-b))(b+d-2) - g\right] \frac{t^2}{2} \tag{4.117}$$

If $P_n(t)$ is the degree n polynomial obtained by truncating the Taylor series of $f(t)$, then $R_n(t)$ is the associated error: $f(t) = P_n(t) + R_n(t)$. The error term $R_n(t)$ is known as the Lagrange Remainder [Adams, 1991] and is given by:

$$R_n(t) = \frac{f^{(n+1)}(t_c)}{(n+1)!}(t-c)^{n+1}, \tag{4.118}$$

for some $t_c \in (c, t)$. Here, $\left|p_t^{(n)}\right|$ is maximised at $t = 0$, thus R_n calculated with $t_c = 0$ bounds the error of the approximation. The first and second order errors for parameters $b = 0.95$; $d = 0.99$; $g = 0.01$; $\beta = 1$; $u = 0.25$; $u_\star = 0.75$ at $t = 5$ are,

$$|R_1(5)| < 0.4352 \qquad\qquad |R_2(5)| < 0.0389. \tag{4.119}$$

Neither are ideal, but the method demonstrates that the terms $\ln(b)^{n+1}$ and $\ln(d)^{n+1}$ in $|R_n^{(n+1)}|$ must be sufficiently small to combat t^{n+1} for good approximations as t increases. Since $b < d$, if $t^{n+1}(d-1)^{n+1} \ll 1$ the nth degree approximation should give good results up to t. Further, the error scaled by accumulated input: $R_n(t)(b-1)/(u_\star(b^t - 1))$ may give a more practical measure of the early error. Under this scaled measure, the first and second order approximations errors are reduced for large $u_\star$ (potentially $u_\star \in \mathcal{U}_3, \mathcal{U}_4$), which conveys the notion that the large input dominates over the increase in n_t. This concept of large $u_\star$ will be of interest in the next chapter. The first and second order approximations can be solved to calculate the first time step that $p_t \geq \ell$, denoted t_ℓ. Note that the solutions are only valid for t_ℓ in which the approximations are valid, i.e. $|R_n(t_\ell)| < \epsilon$ for some acceptable margin of error ϵ. For the first order approximation from initial $p_0 = \beta$ and $n_0 = u - \beta(1-b)$,

$$t_\ell \approx \frac{\ell - \beta}{u_\star - u}, \tag{4.120}$$

which exactly describes the one-step breaking condition (4.103). The second order approximation gives a quadratic in t (see Equation 4.116), if the approximation is valid for both solutions then they are the times that p_t breaks ℓ and then falls below ℓ respectively.

4.5.2 Varying input u_t

A fuller picture can be obtained by considering varying $\{u_t\}$, and the assumptions that $p_0 = p_e$, $n_0 = n_e$ and $p_t \geq \beta$ are temporarily dropped. Recall the solution for p_t (4.13), and

writing without the specifics of n_t,

$$p_t = p_0 b^t + \sum_{k=0}^{t-1} u_k b^{t-1-k} - \sum_{k=0}^{t-1} n_k b^{t-1-k}. \tag{4.121}$$

The summations involving u and n are a discrete convolution with the sequence $\varphi_t = \{b^t\}$, so for u,

$$\varphi_t * u_t = \sum_{k=0}^{t} b^{t-k} u_k. \tag{4.122}$$

The condition for $p_t \geq \ell$ can then be restated, if there is a t such that,

$$p_0 b^t + \varphi_{t-1} * u_{t-1} \geq \varphi_{t-1} * n_{t-1} + \ell, \tag{4.123}$$

then $p_t \geq \ell$. A receptor performs a comparison between the growth of $\{u_t\}$ and the growth of $\{n_t\}$, this is depicted in Figure 4.16. The convolution equation describes the anomaly condition for any choice of the dynamics of n_t. The condition will be of use in Section 4.6.2 and the next chapter when the receptors are applied to type 2 anomaly detection. If the summation in the convolution between $\{u_t\}$ and $\{\varphi_t\}$ is made, again, explicit:

$$\sum_{k=0}^{t-1} b^{t-1-k} u_k \geq \varphi_{t-1} * n_{t-1} + \ell - p_0 b^t, \tag{4.124}$$

then $\{u_t\}$ may be taken out of the convolution by the Cauchy-Schwartz inequality,

$$\left(\sum_{k=0}^{t-1} b^{2(t-1-k)}\right)\left(\sum_{k=0}^{t-1} u_k^2\right) \geq \left(\varphi_{t-1} * n_{t-1} + \ell - p_0 b^t\right)^2 \tag{4.125}$$

$$\sum_{k=0}^{t-1} u_k^2 \geq \frac{1}{\phi_{b^2}^t}\left(\varphi_{t-1} * n_{t-1} + \ell - p_0 b^t\right)^2. \tag{4.126}$$

Equality will hold between the left hand sides of (4.124) and (4.125) if $u_k = c b^{t-1-k}$, for a constant c and a particular t. Although, in general one must expect,

$$\left(\sum_{k=0}^{t-1} b^{2(t-1-k)}\right)\left(\sum_{k=0}^{t-1} u_k^2\right) > \left(\sum_{k=0}^{t-1} b^{t-1-k} u_k\right)^2, \tag{4.127}$$

and so the anomaly condition is weakened as there are more $\{u_t\}$ for which (4.125) is true than (4.124). If both conditions hold, then (4.125) will hold for earlier t than (4.124). The difference between the left hand sides of (4.124) and (4.125), gives the error E,

$$E = \frac{1}{2}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\left(u_k b^{t-1-j} - u_j b^{t-1-k}\right)^2. \tag{4.128}$$

If (4.125) holds at time $t = \tau$ and $\sum_{k=\tau+1}^{t^\star} u_k^2 < E$, then (4.125) will have erroneously predicted an anomaly. Here, $t^\star$ represents some time in the future when either u_t has gone to zero or $\varphi_{t-1} * n_{t-1}$ has increased to invalidate the condition of (4.126). Despite weakening

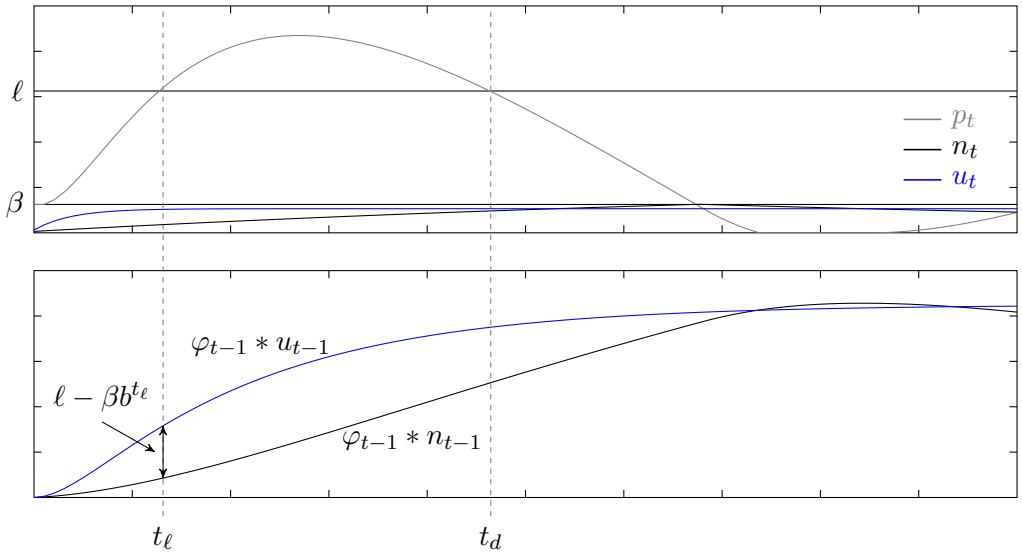

Figure 4.16: The condition for $p_t \geq \ell$. *Top:* p_t, n_t u_t, with parameters $b = 0.95$; $d = 0.99$; $g = 0.02$; $\beta = 1$; $p_0 = \beta$; $n_0 = u_0 - \beta(1-b)$. u_t is defined by (4.129) with $\mu = 0.764$, $\nu = 0.2$ and $u_0 = 1$. $l = 5$ and t_ℓ and t_d are the times when p_t breaks ℓ and falls back down below ℓ respectively. *Bottom:* The convolutions between the sequence ϕ_t, u_t and n_t. The separation between the two at time t_ℓ is $\ell - \beta b^{t_\ell}$

the anomaly condition (4.126) provides a good and useful approximation for the time when $p_t \geq \ell$, this will be demonstrated in Section 4.6.2. Conditions (4.123) and (4.126) will be respectively referred to as the convolution and summation conditions of *type 2* anomaly, as they describe the breaking of ℓ by a variable input.

It should be noted that with the current choice of negative feedback starting from $p_0 < \beta$ the left-hand and right-hand sides of (4.126) are not independent. With the assumption of the scenario in Section 4.5.1 that $p_0 = p_e = \beta$ and $n_0 = n_e = u_0 - \beta(1-b)$ for $u_0 \in \mathcal{U}_2$ and the additional assumption that $u_t \geq u_0$ for all t then the two sides of (4.126) become independent of u_t whilst $p_t \geq \beta$.

Finally, consider the case where $\{u_t\}$ is controlled by a recurrence with the same form as the negative feedback:

$$u_{t+1} = \mu u_t + \eta \tag{4.129}$$

with $0 < \mu < 1$ and $\eta > 0$. Clearly, if $\frac{\eta}{1-\mu} \geq \beta(1-b) + \frac{g}{1-d}$ then an anomaly will eventually occur as $\max\{u_t\}$ is not in $\mathcal{U}_2$. Assuming that $\{u_t\}$ remains in region 2 and η is small enough that it will not cause an anomaly on the first timestep, so $\eta < \ell - \beta - u_0(\mu - 1)$, then an anomaly will never occur if $\eta < g$. However, an anomaly will only occur if $\eta > g$ and (4.123) holds. In general the metric of the growth of $\{u_t\}$ being greater than that of $\{n_t\}$ is a good indicator of an anomaly. Thus, if $u_t \in \mathcal{U}_2$ for all t and $\Delta u_t < g$ then an ℓ will never be broken.

4.6 The Receptor Density Algorithm

Combining the ideas of the chapter thus far a lattice of receptors can now be described as an anomaly detection algorithm:

Definition 4.3. *The Receptor Density Algorithm (RDA) is a n-dimensional lattice of receptors L, all with identical parameters b, d, g, β, ℓ and the stimulation kernel function K_S with width h. The lattice is presented with input $\{\boldsymbol{\gamma}_t\} = \{(u_t, \mathbf{z}_t)\}$ with $u_t \geq 0$ and $\mathbf{z}_t \in \mathbb{R}^n$ is updated according to,*

$$p_{t+1}(\mathbf{x}) = bp_t(\mathbf{x}) + K_S(\mathbf{x}, \boldsymbol{\gamma}_t) - n_t(\mathbf{x}) \tag{4.130}$$

$$n_{t+1}(\mathbf{x}) = dn_t(\mathbf{x}) + gH(p_t(\mathbf{x}) - \beta), \tag{4.131}$$

and will detect an anomaly in the input at time t if for any $\mathbf{x}$: $p_t(\mathbf{x}) \geq \ell$.

Note that the input to the RDA is a $(n+1)$-dimensional vector $\boldsymbol{\gamma}_t$. This does not imply that if anomaly detection is desired on y-dimensional data $\mathbf{D}_t \in \mathbb{R}^y$ then the RDA should be applied with a $(y-1)$-dimensional lattice. The first dimension of $\boldsymbol{\gamma}_t$ has a special meaning that weights each input point. In general $\boldsymbol{\gamma}_t = \mathbf{D}_t$ and $\boldsymbol{\gamma}_t = (1, \mathbf{D}_t)$ will produce different anomaly detection behaviour. Thus, y-dimensional data may still require a y-dimensional receptor lattice.

A discussion is now given relating the anomaly detection of the RDA to anomaly detection problems *type 1* and *type 2*.

4.6.1 The RDA and *type 1* anomaly detection

Recall that type 1 anomaly detection involves the classification of a new sample point $\mathbf{v}$ given a training sample $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$. The RDA is able to address this problem via the argument outlined in Section 4.4.2. The sample is presented and the negative feedback only builds in regions where the decay density estimate is above threshold β, and so if $\mathbf{v}$ falls in a low probability region of the lattice the receptor position will immediately break ℓ due to the lack of sufficient negative feedback. This is a one-step breaking of ℓ from $p_t \leq \beta$. The condition given in (4.103) describes one-step breaking from β with negative feedback and should not be satisfied. The requirement is that the presence of negative feedback prevents one-step breaking but without negative feedback one-step breaking always occurs. It is required that ℓ is broken from a point on the lattice with $p(\mathbf{x}) = 0$ and $n(\mathbf{x}) = 0$, so $\max\{K_S\} \geq \ell$ must hold. As condition 4.103 must not be satisfied, then there is the following bound on ℓ,

$$\max\{K_S\} - \beta b < \ell \leq \max\{K_S\}, \tag{4.132}$$

for type 1 anomaly detection to occur. By the arguments of the chapter, $b < d$ should both be set near 1 for a good quality density estimate and g should be set near 0 to minimise jumps in $n_t(\mathbf{x})$. However, in practice this scheme fails of the majority of $\mathbf{v}$ close to the decision boundary. The issue is that the approach is based on the static somewhat qualitative argument of Section 4.4.2. By virtue of the recurrences, the state of $n_t(\mathbf{x})$ is disturbed by

the anomaly and one must wait for $n_t(\mathbf{x})$ to decay before a subsequent anomaly in the same region can be detected. This draws attention to the fact that no discussion examining the stochastic input $\{u_t\}$ in conjunction with the negative feedback and β has been developed in this chapter. In reality, since $\beta < \ell \leq \max\{K_S\}$ sample points near the anomaly detection boundaries will generate a non-zero negative feedback in regions where $p(\mathbf{x}) < \beta$. Despite this, the arguments of the chapter give the correct intuition to understand the behaviour, and without making any formal arguments this problem may be solved with relative ease. The parameters should be set such that there is non-zero but stable negative feedback over the tails of $p(\mathbf{x})$. Then ℓ may be lowered such that one-step breaking of ℓ is possible from a region of non-zero negative feedback, and that it is not possible for the level of negative feedback just inside the decision boundary. The stable negative feedback provides a clean anomaly detection platform. This can be achieved by setting d close to 1, and then set g so that maximum negative feedback $g/(1-d) > 1$.

A simple test is performed to evaluate the performance of the RDA at type 1 anomaly detection in one dimension. A sequence of samples $\{x_t\}$ are drawn independently from a known distribution $p(x)$, the RDA and a standard kernel density estimate $\hat{p}(x|\omega_t)$ are compared at determining whether a new point x_t has $p(x_t) < \alpha$. For the RDA this is done by updating the recurrences and testing $p_{t+1}(x) \geq \ell$; for the standard kernel density estimate the anomaly condition is $\hat{p}(x|\omega_{t-1}) < \alpha$. For the sake of simplicity, $p(x)$ is chosen as the standard normal distribution and $\alpha = 0.02$. The parameters are set as follows: $w = 500$ the width of window ω_t; $h = 0.21$ the width of the kernel in K_S and $\hat{p}(x)$, the standard normal kernel is used; $u_t = (1-b)$; $z_t = x_t$; $b = 0.9$; $d = 0.999$; $g = 0.01$; $\beta = \alpha = 0.02$; $\ell = 0.1$. These parameters achieve the required stability in negative feedback. Note that b and d lie either side of the point of equal variance $w - 1/w + 1$. The equality of variance condition will not hold exactly for the density estimate carried by the negative feedback, but $w - 1/w + 1$ still provides a good metric in setting b and d

The RDA and $\hat{p}$ are initialised with 500 data points, and then 40000 test points are evaluated. The two methods are compared in terms of receiver operator characteristics (ROC), which compares the result of a classifier against the true result. The terminology and conditions on the RDA and kernel density estimator (KDE) are as follows,

	$p(x_t)$	KDE	RDA
True Positive (TP):	$p(x_t) \leq \alpha$	$\hat{x}_t(v\|\omega_{t-1}) \leq \alpha$	$\exists x\ p_{t+1}(x) \geq \ell$
True Negative (TN):	$p(x_t) > \alpha$	$\hat{p}(x_t\|\omega_{t-1}) > \alpha$	$\forall x\ p_{t+1}(x) < \ell$
False Positive (FP):	$p(x_t) > \alpha$	$\hat{p}(x_t\|\omega_{t-1}) \leq \alpha$	$\exists x\ p_{t+1}(x) \geq \ell$
False Negative (FN):	$p(x_t) \leq \alpha$	$\hat{p}(x_t\|\omega_{t-1}) > \alpha$	$\forall x\ p_{t+1}(x) < \ell$

There a number of common meta-statistics defined on TP, TN, FP and FN. Three considered here are,

$$\text{True Positive Rate: } \frac{\text{TP}}{\text{TP}+\text{FN}}, \quad \text{False Positive Rate: } \frac{\text{FP}}{\text{FP}+\text{TN}},$$

$$\text{Accuracy: } \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}.$$

	KDE	RDA
TP	451	**492**
TN	**39359**	39295
FP	**80**	144
FN	110	**69**
True Positive Rate	0.803922	**0.877005**
False Positive Rate	**0.002028**	0.003651
Accuracy	**0.995250**	0.994675

Table 4.1: *Type 1* anomaly results. Boldface indicates the superior score. 561 anomalous x_t were drawn from $p(x)$, which is appropriate given the area under the normal curve outside the decision boundaries $561/40000 = 0.014 \approx 1 + \text{erf}\left(-\sqrt{-\ln(\alpha\sqrt{2\pi})}\right)$

The true assignment of each x_t is known by inverting the normal distribution, which gives decision boundaries at $\pm\sqrt{-2\ln(\alpha\sqrt{2\pi})}$. Then x_t is anomalous if $|x_t| \geq \sqrt{-2\ln(\alpha\sqrt{2\pi})}$. For $\alpha = 0.02$ this gives $|x_t| \geq 2.447$. The results are given in Table 4.1. The RDA and KDE perform well and comparably. The RDA is slightly more sensitive (more likely to determine any x_t anomalous) than the KDE in that the RDA has greater TP, FP lesser FN. The advantage of this simplest test is that the distributions of TP, TN, FP, FN can be calculated and presented in a clear manner. During the test each x_t was marked dependent on whether it contributed to TP, TN, FP, FN. of either algorithm. The distributions are given in Figure 4.17. No standard parametric or non-parametric significance tests [Siegel, 1988; Vargha and Delaney, 2000] find any differences between the distributions from either algorithm for a given ROC statistic. Clearly, the mass of the FP, FN, and TN distributions lie over the decision boundary. Observe the long tail of the RDA FP distribution toward zero: with current parameters the RDA has greater susceptibility to noise in $\{x_t\}$ as receptor activations occur in areas of dense x_t.

4.6.2 The RDA and *type 2* anomaly detection

This section discusses the more interesting anomaly detection problem, the *type 2* anomaly, and is driven by the example of a 1-dimensional time-varying distribution $\psi_t(x)$. The distribution is described by a standard normal distribution whose mean shifts at rate ν,

$$\psi_t(x,\nu) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(x-\theta_t)^2}{2}\right), \qquad \theta_t = \begin{cases} -3 & 0 \leq t < t_s \\ -3 + \nu(t - t_s) & t_s \leq t \leq 6/\nu + t_s \\ 3 & 6/\nu + t_s < t \end{cases}, \tag{4.133}$$

θ_t is the mean of $\psi_t(x)$ with initial value $\theta_0 = -3$. From time t_s, θ_t increases at rate ν until $\theta_t = 3$ at time $6\nu + t_s$. Samples are drawn from $\psi_t(x,\nu)$ to generate stream $\{z_t\}$ that is the input to the RDA. With a rearrangement notice that the distribution of input at any one

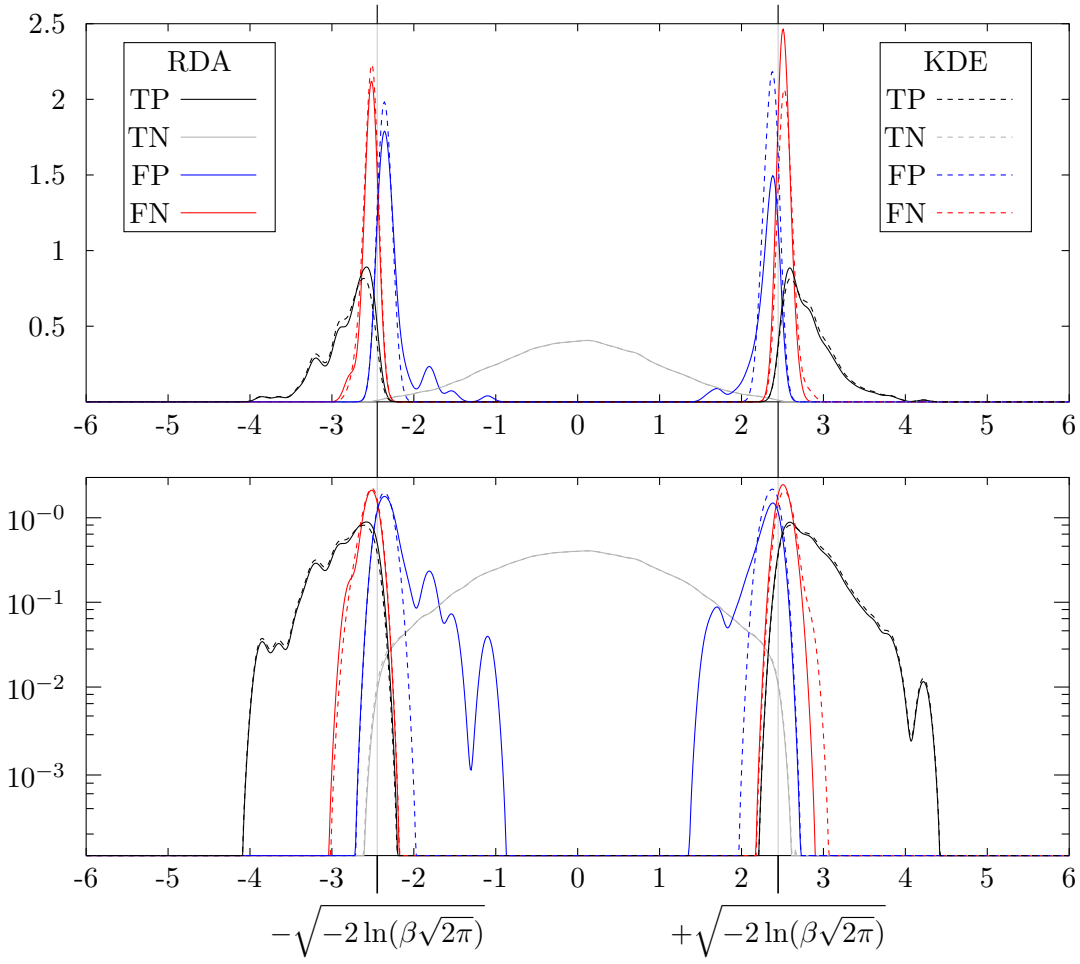

Figure 4.17: The distributions of TP, TN, FP, FN for RDA and KDE on a linear scale (*top*) and a log-scale (*bottom*) . Distributions are themselves kernel density estimates calculated from the portion of the sample that contributed to each ROC class. Note the mass of the distribution lies on the decision boundaries. The distributions for TP, TN and FN for each algorithm are almost identical. Only the RDA FP distribution shows a difference in the tails, which stretch toward zero.

location x has the form of the normal distribution in time, for $t_s \leq t < 6/\nu + t_s$,

$$\psi_t(x,\nu) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\nu^2(t-(t_s+\frac{3+x}{\nu}))^2}{2}\right), \tag{4.134}$$

so $\nu\psi_t(x,\nu)$ is a normal distribution with mean $t_s + \frac{3+x}{\nu}$ and variance $1/\nu^2$. Any x which is passed by the leading edge of ψ_t will experience the maximum change in underlying distribution. Approximating with continuous t the time of the maximum positive rate of change in ψ_t can be found by solving,

$$\frac{\partial^2 \psi_t}{\partial t^2} = 0 \qquad \text{which gives} \qquad t = t_s + \frac{3+x}{\nu} - \nu, \tag{4.135}$$

$$\text{and so} \qquad \max\left\{\frac{\partial \psi_t}{\partial t}\right\} = \frac{\nu}{\sqrt{2\pi e}}. \tag{4.136}$$

Consequently all receptors with location $-2 < x < 2$ will experience the maximum rate of change of the leading edge of ψ_t. If the receptors in this range are robust to this passing of the leading edge, then no anomaly will be detected. For example, if $\nu < g\sqrt{2\pi e}$ then the expectation, with respect to the sample, is that no anomaly will occur. The above is closely related to the MSD defined in (4.64) in Section 4.3.2. However, it is clear from the convolution (4.123) and summation (4.126) conditions that type two anomaly detection in the RDA is related to an integrated MSD. Rather than define an integrated MSD it is clearer to discuss the ψ_t with respect to the convolution and summation conditions of the RDA.

This example can be modelled by considering a single receptor with input $\{\psi_t\}$ (rather than $u_t = u$ or $u_t = 0$ as a function of $\mathbf{z}_t$) and it should convey the expectation of anomaly. It is necessary to render the sequence $\{n_t\}$ independent of $\{\psi_t\}$ to allow calculation of the convolution and summation anomaly conditions. Given x and ν defining ψ_t assume that $p_0, n_0 = 0$ then the time at which n_t begins to grow is the smallest t_n such that $\varphi_{t_n} * \psi_{t_n} \geq \beta$. The convolution and summation conditions are then calculated from time t_n using the $p_t \geq \beta$ solution for n_t (4.16). If either condition indicates an anomaly at time t_ℓ, letting $\tau = t_\ell - 1 - t_n$ be the time of the anomaly with respect t_n, then the anomaly is only valid if,

$$\varphi_\tau * (u_\tau - n_\tau) - p_{t_n} b^\tau \geq \beta. \tag{4.137}$$

The RDA is demonstrated with two ψ_t scenarios, a fast $\nu_f = 0.01$ and a slow $\nu_s = 0.002$. For each, demonstrations are given of the RDA tracking the shift (such that no anomalies occur) and then RDA detecting the shift as an anomaly. The parameters are chosen by combining the results of the chapter. A discussion of the meaning of the parameters is given in the subsequent conclusions and discussion section. For the moment, b and d should be chosen near 1 with $d > b$ and g near zero, this will help smooth the noise in the sample and provide good quality density estimates. However, b and d also define the window for the internal density estimate in the $(1+b)/(1-b)$ in the manner described in Section 4.4.1. Thus, b and d must also be set at a rate relevant to the underlying distribution. As such, the fast distribution requires fast rates for tracking and the slow distribution requires very

slow decay rates to detect the slow shift as an anomaly. There is some freedom in the choice of β and ℓ, here they are set such that type 1 anomaly detection does not occur (i.e. a single u_t is insufficient to result in $p_t(x) \geq \beta$). Arbitrarily, β is set at 1, this defines a scale for p_t which is greater than the underlying distribution, however n_t still contains a good density estimate and this scheme provides good results. Finally, note that the slower distribution has a greater area for one x when evaluated in time; this combined with a slow decay rate increases the size of the convolutions. As a consequence, this defines the scale by which it is appropriate to set ℓ.

The parameter setting proceeds by first choosing $-2 < x < 2$ which is passed by the leading edge of ψ_t, $x = 0$ is chosen. The conditions are used with the constraints of the parameter setting described above. The results are given in Figure 4.18 for the fast shift and Figure 4.19 for the slow shift. The corresponding parameters are given in Table 4.2. The parameters for fast anomaly detection and slow tracking have been purposefully chosen identical for comparison (apart from an increase in ℓ for the slow tracking). Note that as b increases the error of the summation condition 4.128 increases, which results in a difference in predicted time of anomaly in Figure 4.128. As stated above, the use of ψ_t directly in the anomaly conditions only describes the conditions for anomaly in the expected sense. The true behaviour of the receptor at $x = 0$ may be investigated by sampling from the time-course of ψ_t. A receptor with parameters given in Table 4.2 is presented with input $u_t \sim \text{Bernoulli}(\psi_t(0, \nu))$ with $\nu = \nu_f, \nu_s$. As noted in Section 4.4.1 this will have greater variance than the input passed through the kernel. One thousand instances of the ψ_t are simulated and are also presented in Figures 4.18 and 4.19. The results are given in Table 4.3, and are as desired. The 1/10 occurrence of slow tracking anomalies can easily be reduced by an increase in g or ℓ. However, leaving the parameters as they are allows comparison between fast anomaly detection and slow tracking, moreover it later provides an opportunity to demonstrate the variance reduction of the kernel.

The results thus far have demonstrated the appropriate parameter setting for the extremes of the ψ_t distribution. Now the behaviour is examined with the full ψ_t distribution. A 1-dimensional lattice of receptors is used with receptor separation $\delta = 0.01$ and extremes $k_{min} = -6$ and $k_{max} = 6$. The input to the lattice at time t has location z_t drawn from $\psi_t(x, \nu)$, the magnitudes of all u_t are set equal with $u_t = 1$. A standard normal kernel is used in function K_S, the kernel widths are given in Table 4.2. Figure 4.20 gives the results for fast $\psi_t(x, \nu_f)$ and Figure 4.21 for slow $\psi_t(x, \nu_s)$. The desired tracking and anomaly detection behaviour is observed. Moreover, the statistics of the tracking and anomaly detection are confirmed with 1000 repetitions of the ψ_t shift. The results are given in Table 4.3 and demonstrate that the parameters chosen for a single receptor translate well to many receptors. The stimulation kernel K_S plays an important role in reducing the variance of $\{u_t\}$ which greatly reduces the occurrence of tracking anomalies for the slow distribution.

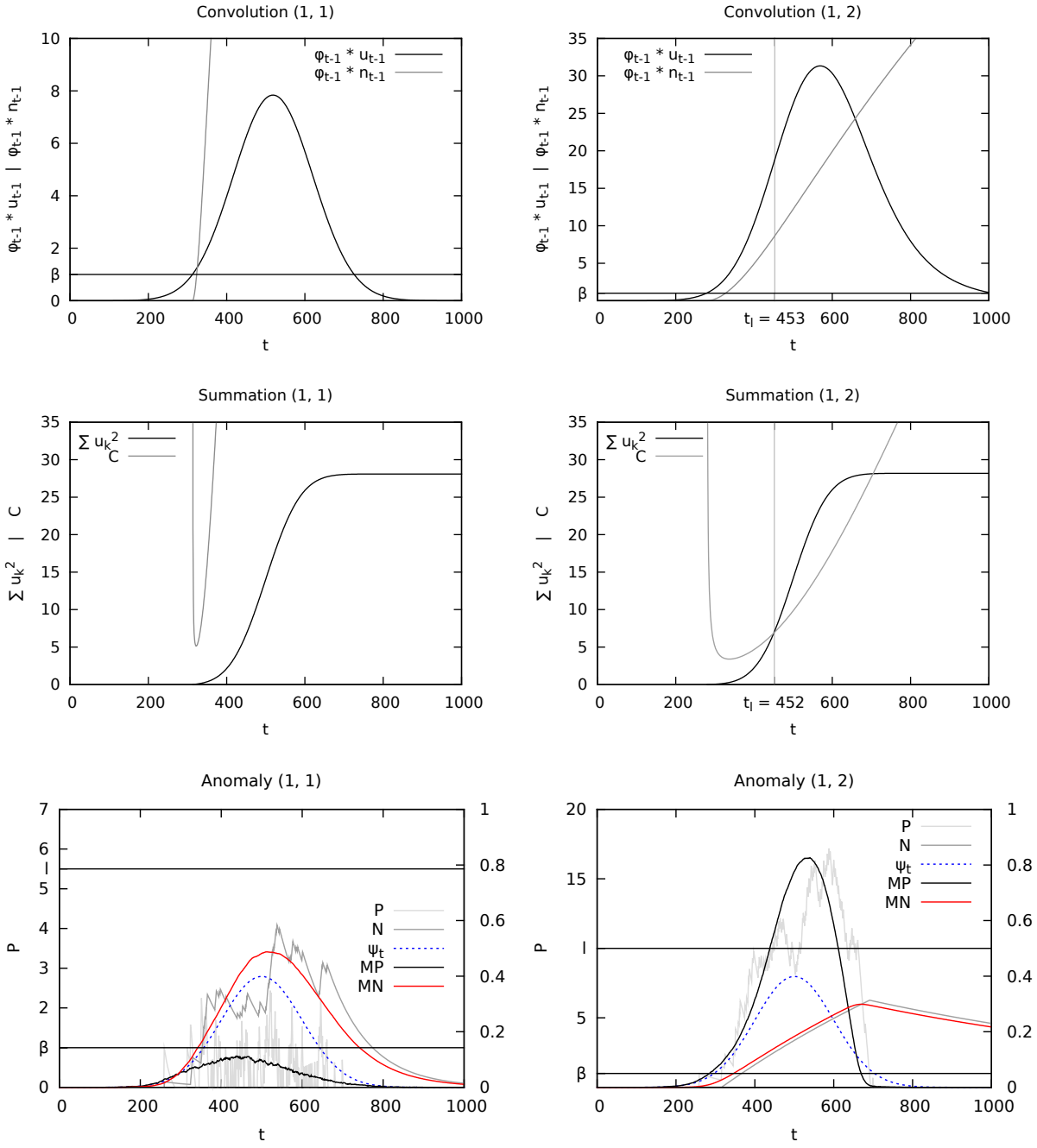

Figure 4.18: Parameter setting for fast $\psi_t(0, v_f)$ tracking (*left*) and anomaly detection (*right*). *Top:* Convolution condition, $t = t_\ell$ marks first satisfaction of the condition, the difference between lines is $\ell - p_b^{t-t_n}$. *Middle:* Summation condition, C marks the R.H.S. of (4.126), $t = t_\ell$ marks first satisfaction of the condition, C and $\sum u_k^2$ intersect. Note 1 step disagreement with convolution condition. Observe in *top* and *middle* tracking occurs when function of u_t grows slower than the function n_t and vice versa for anomaly detection. *Bottom:* Sampling 1000 timecourses of ψ_t. P and N give p_t and n_t for a single timecourse. MP and MN give the mean across the 1000 repeats. P and MP are read on the left axis and N, MN and ψ_t are read on the right axis. For tracking the MP $< \ell$ and for anomaly detection MP $> \ell$.

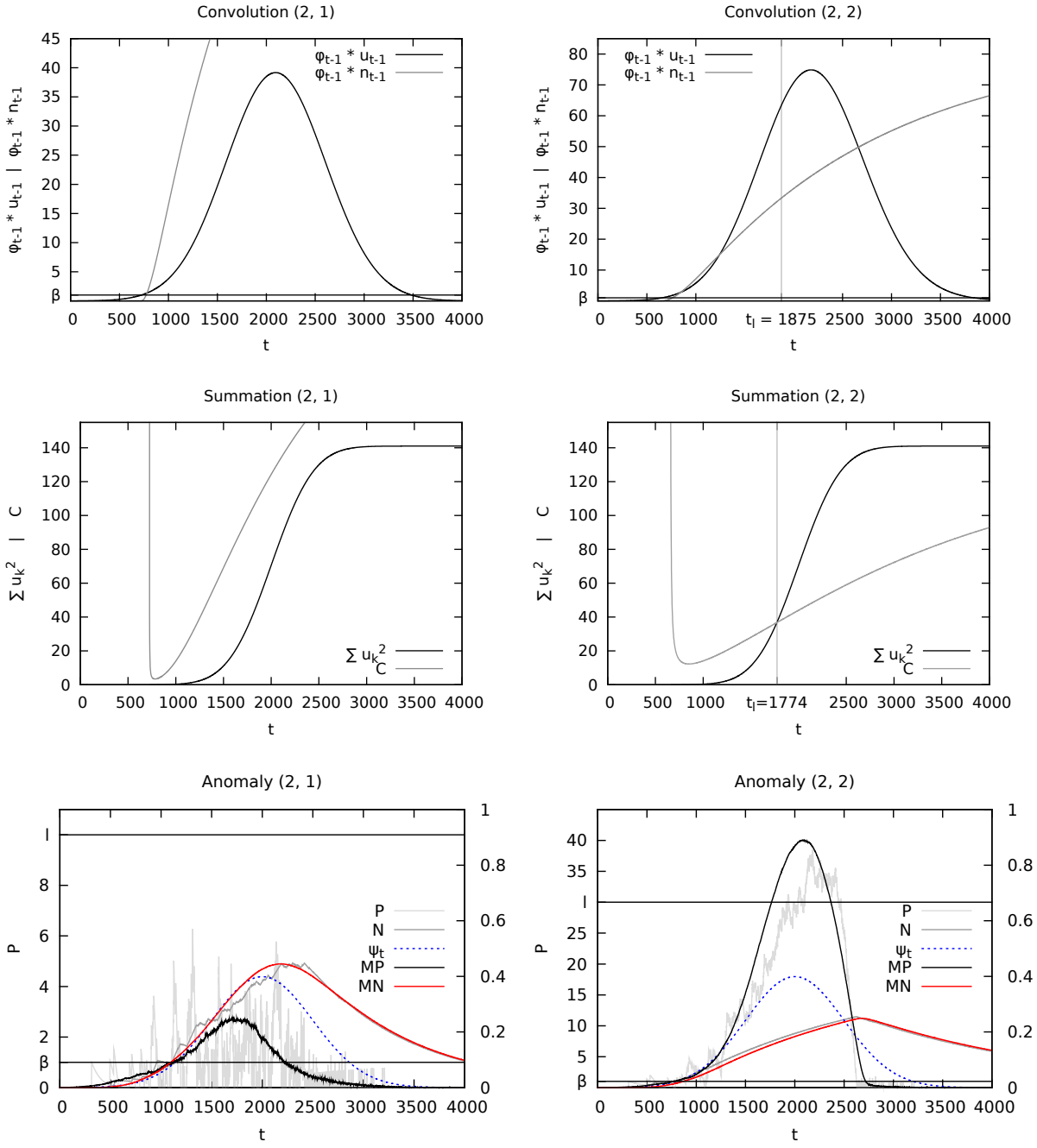

Figure 4.19: Parameter setting for slow $\psi_t(0, v_s)$ tracking (*left*) and anomaly detection (*right*). *Top:* Convolution condition. *Middle:* Summation condition. *Bottom:* random sampling from ψ_t. See caption of Figure 4.20 for detailed description. Note with increased b, the error in the summation condition increases, resulting in a 100 time step difference in predicted anomaly.

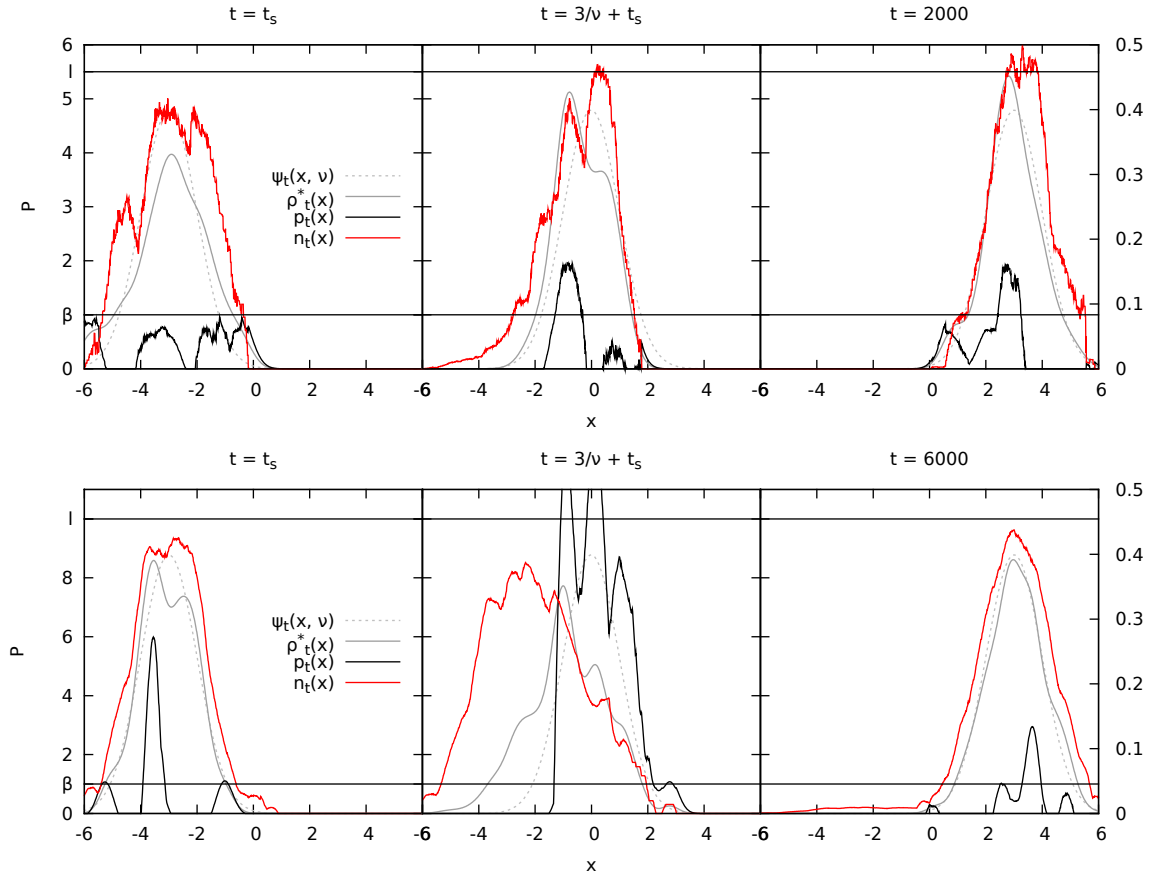

Figure 4.20: Type 2 anomaly detection for fast $\psi_t(x, \nu_f)$, for tracking (*top*) and anomaly detection (*bottom*). The three panels display snap shots of $\psi_t(x, \nu_f)$ (dashed grey, *right* axis); $\rho_t^\star(x)$ (solid grey, *right* axis) the decay density estimate estimate with the same b as used in RDA shown for comparison; $p_t(x)$ (solid black, *left* axis); $n_t(x)$ (solid red, *right* axis). The snapshots are time the start of the shift t_s; mid-shift $3/\nu + t_s$; and at the end of the shift after an appropriate settle time. When tracking, the negative feedback moves with the shift in ψ_t ensuring no anomaly occurs. For anomaly detection, the negative feedback loses correspondence with ψ_t and the receptor position builds and exceeds ℓ.

	Fast: $\nu_f = 0.01$		Slow: $\nu_s = 0.002$	
Parameter	Tracking	Anomaly	Tracking	Anomaly
b	0.95	0.99	0.99	0.995
d	0.99	0.999	0.999	0.9995
g	0.02	0.001	0.001	0.00021
ℓ	5.5	10	15	30
β	1	1	1	1
w_b	39	199	199	399
w_d	199	1999	1999	3999
h	0.4	0.3	0.3	0.25
u_t	1	1	1	1

Table 4.2: Parameters for RDA tracking and anomaly detection of fast and slow ψ_t. Additional parameters, w_b and w_d, give the window size of the sliding window kernel density estimator with equivalent variance to the decay density estimate with parameters b and d respectively.

	Single Receptor		Lattice of Receptors	
	Fast	Slow	Fast	Slow
Tracking	1	105	2	23
Anomaly	997	1000	1000	1000

Table 4.3: Tracking and anomaly detection results, table gives number of anomalies detected for the single receptor (Figures 4.18 and 4.19) and a lattice of receptors (Figures 4.20 and 4.21) for fast $\psi_t(x, \nu_f)$ and slow $\psi_t(x, \nu_s)$. All parameters are given in Table 4.2 and are identical between single receptor and lattice. All results out of 1000 runs. Note the improvement due to the kernel in reducing erroneous tracking anomalies in the slow distribution.

.

4.7 Discussion and Conclusions

This chapter began by identifying general properties of the T cell signalling processes, that looked to be of relevance to bio-inspired algorithm development. To confirm this, an abstracted computational model of the T cell signalling processes was defined. This model was found to have certain properties in common with classical statistical anomaly detection techniques. With comparison to the classical concepts, the abstracted model has been developed into an anomaly detection system. In more detailed summary:

Section 4.1 defines an abstracted model from the details of the T cell signalling model. The abstracted model focuses on the details of the hypothesis at the end of the last chapter, and as a consequence discards much detail but retains the concept of the negative feedback base state. The abstracted model defines a *receptor* which has a continuous position in analogy to TCR kinetic proofreading state and a continuous negative feedback in analogy to local pSHP1. The dynamics of a receptor are given in terms of discrete recurrent equa-

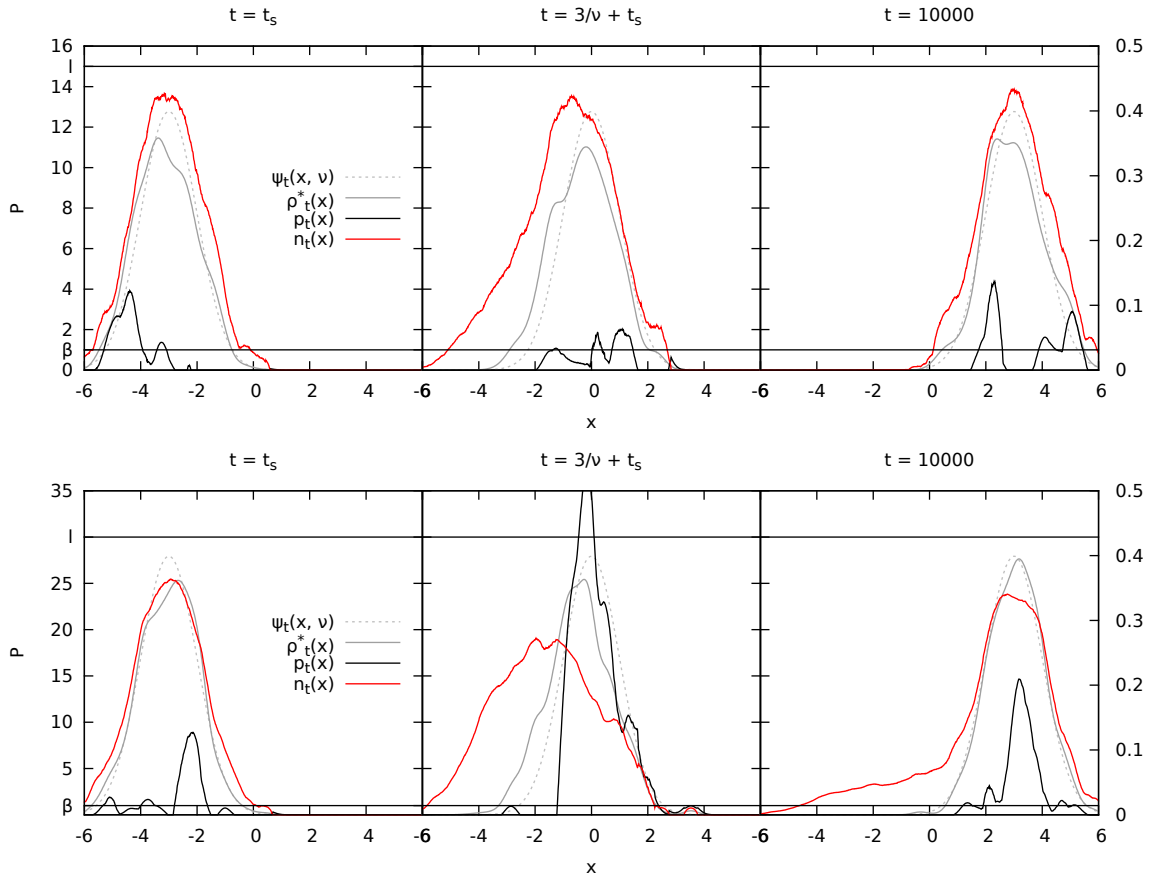

Figure 4.21: Type 2 anomaly detection for slow $\psi_t(x, \nu_f)$, for tracking (*top*) and anomaly detection (*bottom*). See caption of Figure 4.20 for details. Additionally, note the smoothing the receptor position and negative feedback in comparison to Figure 4.20. This is due to improvement of the density estimates by the reduction in variance by increasing the proximity of b and d to 1. Further, in the bottom anomaly detection case, even at $t = 10^4$ the negative feedback displays clear residue of the shift.

tions. This is a form of particular utility given the intended computational goal. The details of combining multiple receptors presents challenges. First, the receptor lacks the molecular recognition part of the TCR. Second, that the T cell signalling pathways perform a comparison task that is remarkably difficult for biology, but relatively simple for computers. The solution to first problem is the introduction of receptors in a lattice. Then interactions between multiple receptors are defined by kernel function K_S and K_N which describe the spreading of input and negative feedback respectively. Figure 4.3 demonstrates the abstracted model is able to replicate the hypothesis presented in the final section of Chapter 3. This is the intent of the abstracted model: to remove detail such that the desired properties of the biological system remain. The connection between the abstracted model and the biological model through assumption set A are discussed concluding chapter of this thesis.

Initial theoretical properties regarding the behaviour of a single receptor are then presented in Section 4.2. The results include two basic solutions with respect to the negative feedback barrier β, and details of the fixed points of the recurrences. Of most interest to

the chapter is input range $\mathcal{U}_2$, for an input $u \in \mathcal{U}_2$ the receptor position and negative feedback are shown to orbit an approximate equilibrium point in the (n, p)-plane. The feature of importance is that at this equilibrium point the p becomes independent of u and the magnitude of u is translated to n.

Section 4.3 commences with a definition of two types of anomaly of relevance to the chapter. Type 1 anomaly detection is concerned with the problem of evaluating the similarity between a test point $\mathbf{v}$ and a sample $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{m-1}$. Type 2 anomaly detection inherently concerns a stream of samples with an underlying distribution thought to change in time. The problem of type 2 detection is to track these changes and flag an anomaly if the underlying distribution changes too quickly. The section then notes that a solution to the receptor position equation has the form of a kernel density estimate. A background to kernel density estimation and the use of the kernel density estimate in type 1 and type 2 detection is then given. Clearly, this connection to density estimation is a pivotal step in the conversion of the abstracted computational model to the anomaly detecting RDA. This particular connection was not a prior intent and its initial discovery was accidental and arose when considering the summation of kernel functions.

The details of the density estimate present in the dynamics of a receptor, defined as the decay density estimate, are then developed in Section 4.4. In particular, it is shown the decay density estimate performs an exponential smoothing of the input. The result is an estimator with identical bias to the standard kernel density estimator (i.e. only the bias due to the kernel) and the ratio of variances of each estimator is bound by a constant.

Section 4.4 also unites the decay density estimate with negative feedback. In doing so, a key step is taken in the construction of an anomaly detection system: the removal of negative feedback diffusion. This step constitutes additions to set assumption set A and, interestingly, destroys the mechanism with which the T cell performed "anomaly detection". In general, one would not expect to remove the mechanism with which the biological system performs the behaviour that inspired the algorithm development. A direct abstraction of this mechanism was of limited computational value (Section 4.1.2) because of a missing molecular-recognition portion of the TCR and the signalling system that performs a comparison which can be performed far more directly using conventional computation. However, the *receptor* has interesting dynamical properties (Section 4.2) and the removal of the negative feedback diffusion is the ingredient needed to bring its properties to the forefront. This topic is given further discussion in the concluding Chapter 6. After removing the negative feedback diffusion, Section 4.4 also develops a connection between the negative feedback barrier β and the anomalous distribution threshold α.

Section 4.5 returns to the single receptor to develop the conditions for breaking ℓ. The one-step breaking conditions describe the type 1 anomaly detection conditions. The step increase from u to $u_\star$ and varying u_t conditions describe the type 2 anomaly detection. The consideration of varying input gives the clearest picture of the behaviour of a receptor under increasing input. The receptor performs a convolution of the input sequence $\{u_t\}$ and the growing negative feedback $\{n_t\}$ with the sequence $\{b^t\}$. If the convolution involving the input grows significantly faster than the convolution involving the negative feedback as

measured by their separation ℓ, then an anomaly is detected. The summation condition is then derived by the weakening of the convolution condition. It is of use it allows algorithm parameters to be considered independently from input.

In Section 4.6 the abstracted model of biology with negative feedback diffusion removed is defined as the *Receptor Density Algorithm* (RDA). It is important to note that the behaviour of the algorithm can be described without any biological metaphor, this is an important stage in the development of any biologically inspired algorithm. Certainly, the terms such as the *receptor* have biological origin, but they are not necessary to the algorithm's description.

Demonstrations of the RDA performing type 1 and type 2 anomaly detections are given in Sections 4.6.2 and 4.6.1 respectively. The type 1 demonstrations reveal the RDA to be highly comparable to thresholding of a standard kernel density estimate. The major differences in behaviour of the two techniques lie in the false positive distributions. With the parameters used in the demonstration, the RDA was more susceptible to noise in the input sample, resulting in false positive distribution tails that stretch away from the decision boundary. Further comments on the implications of this are given below. The demonstration of type 2 anomalies presented a method for setting the parameters of the RDA by consideration of expected extremes of change in the underlying distribution in conjunction the the convolution and summation conditions. If the timescales associated with the parameters of the RDA are equivalent or faster than the timescale of change in the underlying distribution then the RDA will track the change. However, if the RDA responds slower than the change the internal density estimate is lost and an anomaly occurs.

With the contribution of the chapter summarised, a discussion of the type 2 anomaly detection mechanism is now given. One can interpret the behaviour of the RDA at type 2 anomaly detection with a connection back to the kernel density estimate and the Bayes optimal decision boundary (4.58). Consider a potentially time-varying probability distribution with the probability density function $p(\mathbf{x}, t)$. It is possible to evaluate whether this distribution is changing in time by the consideration of two estimates defined by sliding windows: $\hat{p}(\mathbf{x}|\omega_{t,l})$ and $\hat{p}(\mathbf{x}|\omega_{t,s})$. The sliding windows are of different lengths, so $\omega_{t,l}$ is a long window and contains the l most recent $\mathbf{x}_t$ and $\omega_{t,s}$ is a short window and contains the s most recent points. So, $s < l$ and $\omega_{t,s} \subset \omega_{t,l}$. Then, one evaluates every $\mathbf{x}_t$ to identify if it was more likely that $\mathbf{x}_t$ was drawn from the distribution estimated by the short window than the long window. So, assign $\mathbf{x}_t$ to the short window distribution if,

$$\hat{p}(\mathbf{x}_t|\omega_{t-1,s}) > \hat{p}(\mathbf{x}_t|\omega_{t-1,l}). \tag{4.138}$$

This condition conveys the notion that if $p(\mathbf{x}, t)$ is changing the changes will be more dominant in the short window and $\mathbf{x}_t$ will have a greater probability given the short window than the long window. It is a formalisation of a type 2 anomaly. Note that since the density estimates are themselves probability density functions and sum to one, if there is a decrease in probability in $p(\mathbf{u}, t)$ for some point $\mathbf{u}$ then there must be a distinct point $\mathbf{v}$ elsewhere in the distribution which experiences an increase in probability. The issue now is that the short window estimate, by definition, has greater variance than the long window estimate.

To prevent this variance from erroneously indicating a type 2 anomaly, it is best to supply the condition (4.138) with a confidence,

$$\hat{p}(\mathbf{x}_t|\omega_{t-1,s}) > \hat{p}(\mathbf{x}_t|\omega_{t-1,l}) + \ell. \tag{4.139}$$

That is, the type 2 anomaly is only detected when the probabilities are separated by ℓ. Clearly, it is not an accident that ℓ has been chosen to denote the threshold, this is in essence the type 2 detection of the RDA. The short window estimate is the decay density estimate contained in the receptor position, the long window estimate is contained in the negative feedback. This is why $d > b$ as the time scale of the negative feedback should be longer than that of the receptor position. While there are many commonalities between the RDA and the windowed estimates, certainly enough to make the comparison worthwhile, the two methods are by no means equivalent.

Brief comments on the meaning of the parameters are now given. It is worth explicitly stating that the convolution with sequence $\{b^t\}$ in the condition for type 2 detection is, of course, the exponential smoothing analysed in Section 4.4.1. The RDA performs an exponential smoothing of the difference between the input and negative feedback. The smoothing means that an estimate of the expected value of the difference between the input and negative feedback at reduced variance is compared to ℓ. This is the reason why the condition is appropriate.

The negative feedback barrier β deserves some discussion, the base negative feedback state plays a crucial role in chapter 3. The importance of β comes from the consideration of single receptor dynamics. It is the feature that renders the receptor position independent of input $u \in \mathcal{U}_2$ and makes the negative feedback dependent on u. For $u \in \mathcal{U}_2$ the trajectory of p_t and n_t in the (n, p)-plane orbits around the point $(u - \beta(1 - b), \beta)$. Then, this allows ℓ to be set constant across the receptor lattice and provides a clean platform for anomaly detection. Note, that rendering one variable of a system independent of the input at the expense of another variable in the system is common concept in Control Theory [Nise, 2000]. The single receptor can be interpreted as a simple controller which endeavours to maintain the receptor position at level β regardless of input. The control mechanism is then very simplistic, it is an on-off switch dependent on $p \geq \beta$ [8]. Certainly, if the problem is to maintain $p_t = \beta$ then one would discard this on-off control for one of the many more sophisticated techniques in control theory that would prevent the large divergence of p_t from β when the input changed. However, it is exactly because the control mechanism is simplistic that p_t can diverge from β which provides the mechanism from anomaly detection.

An alternative perspective on β arises from the analogy with the anomalous distribution threshold α. Clearly, β can be used in the methods described in Section 4.4 as an anomaly detection threshold. However, a further interpretation of β can be gained by examining the motivation in which the thresholding scheme was presented in [Bishop, 1994]. The objective of [Bishop, 1994] was to construct a neural network classifier, and outlying sample points severely hindered the minimising of the sum of squares error function used to set

[8] This perspective gives an alternative argument for $d > b$, as the negative feedback decays it will not overshoot its equilibrium position any more than the simple switch dictates.

network weights. The solution of [Bishop, 1994] was to use a thresholded density estimate to remove anomalies in the training sample. So, the anomaly detection was used to improve the performance of a different classification mechanism. In analogy here, type 2 anomaly detection is the different and even, the more interesting, anomaly detection problem. If type 2 detection alone is desired, type 1 detection can be used to improve its performance by filtering the outlying samples that may alter the shape of the type 2 density estimation. In the context of the RDA, this is not achieved by detecting type 1 anomalies, but by setting the parameters such that a single sample is not enough to break β. Thus outlying samples will never generate negative feedback, and since $\beta < \ell$ neither will they cause an anomaly. In this scenario, β acts like its kinetic proofreading origins, all below threshold inputs are ignored. It is known that the performance kernel density estimation can suffer in the tails of distribution [Silverman, 1986]. Thus, β can side-step the issue of the tails of the distribution by simply ignoring them. It must be noted that this is a benefit in few dimensions, but as the dimensions increase by the argument at the end of Section 4.3 this effect of β becomes detrimental. To use the RDA in high dimensions it is necessary to have a very large number of sample points contributing to the internal estimates. Therefore, b and d must be set exceedingly close to 1, and as a consequence the RDA can only be used to track very slow moving distributions in high dimensions. In this scenario the recursive nature of the RDA becomes very useful as it not necessary to store the large number of samples.

CHAPTER 5

Anomaly Detection in Spectra

This chapter presents an application of the Receptor Density Algorithm. The application shares similarities with *type 2* anomaly detection, but has some fundamental differences. In some difference to the previous chapters the application is now related to chemistry, and is anomaly detection in *spectrometer* data. The problem requires the detection of anomalies in data originating from a device that measures chemical properties of its surroundings. In general, this chapter will assume that these devices are not operated in controllable laboratory conditions, and moreover there may be limited computational resources to process the data.

The chapter is organised as follows: Section 5.1 gives a background to spectrometers and gives a description of the problem; Section 5.2 describes how the RDA may be used to address the problem of anomaly detection in spectra. Section 5.3 and Section 5.4 are case studies, the first describes a Mass Spectrometry problem under relatively controlled conditions; the second describes a more challenging problem involving a robot-mounted Ion Mobility Spectrometer.

5.1 Spectrometry

An *analytical chemistry* device measures chemical properties of a chemical sample. The word sample here has a quite different meaning to the sample data points of the previous Chapter 4. Here, sample refers to a quantity of molecules (potentially in solid, liquid or gas form). Examples of a measurable chemical property include mass, charge, and light scattering. A *spectrometer* is an analytical chemistry device which measures a *spectrum* of a particular chemical property. For every point in the spectrum the spectrometer gives a measurement of the quantity of the sample with the chemical property corresponding to that point in the spectrum. The study of the chemical properties of samples by a spectrometer is know as *spectrometry*. Three relevant examples are: Mass Spectrometry [Downward, 2004] in which the spectrometer gives the distribution of molecular or atomic masses in the sample. Ion Mobility Spectrometry [Eiceman and Karpas, 2004], in which the sample is ionised and the distribution of ion mobilities calculated (see Section 5.4). Finally, Raman Spectroscopy [Lewis, 2001], in which the composition of the sample is analysed by observing

the Raman scattering of light incident to the sample. A concern of all these spectrometry fields is the identification of chemical substances based on the signatures present in the spectra.

The majority of spectrometry is performed in a controlled laboratory environment, there is a degree of control over the sample and possible sources of error. The resulting spectrometry data can have both high quality and resolution, and the processing of this data can be detailed and performed off-line. However, there is a clear application for spectrometry techniques outside the laboratory. For example, the detection and identification of chemical/biological agents, explosives and drugs [Eiceman and Karpas, 2004]. At a less dramatic level, environmental monitoring [Eiceman and Karpas, 2004] and medical applications [Westhoff et al., 2009]. These situations are often void of the luxuries of the laboratory: sources of error are not controllable; computational processing is limited; and often immediate identification is required. A clear case in point is the detection of Improvised Explosive Devices (IEDs): identification must be immediate; the device must be portable; and there are likely to be many other innocuous chemical substances present. Regarding portability, many spectrometers make use of physical size in producing high resolution data. For example, the separation of ions by drift tube in Section 5.4. Longer drift tubes result in higher resolution data. The applications related to harmful chemical agents and explosives have attracted much attention from the military community. A number of devices have been developed for military purpose including, held hand devices [Eiceman and Karpas, 2004; SmithsDetection, 2010b]; robot mounted devices [Gardner et al., 2008]; and even a Unmanned Air Vehicle (UAV)-mounted device [SmithsDetection, 2010a]. The details of the majority of these devices are restricted and out of the public domain.

The focus of this chapter will the construction of an anomaly detection system for the data generated by a portable spectrometry device. The problem of anomaly detection in spectra is now discussed.

5.1.1 Anomaly Detection in Spectra

First, it is convenient to provide a definition of the data produced by a mobile spectrometer,

Definition 5.1. *A spectrum $S_t = (s_{i,t})_{i=0}^{n-1}$ is a non-negative n-dimensional vector describing the chemical state of the input chemical sample at time t and $s_{i,t} \geq 0$ is the spectral value in the ith-position at time discrete time point $t \geq 0$.*

It is assumed that the device can continually sample its environment (via an air-inlet valve, or similar) to produce the sequence of spectra. Further, since the spectrum is represented at discrete positions with a finite width, then data available from a spectrometer is a binned and so smoothed version of the true spectrum [1]. This chapter makes the following assumptions on data available from a mobile spectrometer[2]:

[1] Despite the name *spectrum* the nature of the underlying chemical property is not required to be continuous.

[2] These assumptions are compiled from [Eiceman and Karpas, 2004] and personal communication from

- The value of spectrum location i at time t correlates with the quantity of molecular substance with property corresponding to location i in the environment at a time close to t. Thus, the spectrum is peaked and the locations and heights of the peak define a molecular *signature* for the current state of the environment.

- A particular signature corresponds to the presence of many different chemical substances. In general, the signature of a mixture of substances is not guaranteed to be a linear combination of the signatures of the individual substances. This point is dependent on the chemistry of a particular device.

- The inlet to the spectrometer is short-range, or can be directed or restricted to limit the number of chemical substances present in the sample at any time. As such, for a given environment a spectrum is relatively constant.

- The spectrometer has a characteristic behaviour regarding the growth and decay of signatures in the spectra. A common behaviour is when a substance is first present in the sample it is immediately present in the spectra, then as the substance leaves the sample a decaying residue is left in the spectrum. The result is a sharp initial increase in spectra positions related to the signature, followed by a comparatively slower decaying tail.

- In comparison to its laboratory counterpart, the mobile spectrometer is assumed to offer lower resolution spectra with greater noise. However, the rate at which new spectra are available is not assumed to be hindered.

- The spectrum may be subject to drift, such that an entire signature may appear at locally shifted position in the spectrum.

Anomaly detection in spectra is the detection of the presence of new chemical substances in the input sample by the presence of new features (peaks) in the spectra. In general, this chapter aims for the following desirable properties of an anomaly detection system applied to spectra: First, the system should be able to isolate a noise-reduced signature of the new component of the spectra. This signature should be used to identify the new substance, and the system should have the ability to learn signatures of new substances. Finally, the system should be able to perform these tasks at a rate faster than the occurrence of anomalies.

Unfortunately, there is very little information available regarding current approaches to this anomaly detection problem. This is due to two factors: the proprietary (military) nature of many mobile spectrometry devices and that the publicly available work often concerns results obtained in a laboratory [Eiceman and Karpas, 2004]. The work that is available [Gardner et al., 2008], and the manuals of proprietary devices [SmithsDetection, 2010b], generally have a far more restricted approach to the anomaly detection. The focus is tailoring the system to detect one or two signatures with high accuracy. Some more

Mark Neal of the University of Aberystwyth and Stuart Cairns & Peter Hickey of the Defence and Science Technology Laboratory (DSTL).

general anomaly detection in spectra techniques such as [Davis and Kroutil, 1990; Goubran and Lawrence, 1991; Bell et al., 1999; Chen and Harrington, 2003] are not applicable in this scenario. This is for reasons such as fragility to noise and the need for off-line processing.

Given the above requirements and the physical correspondence between the RDA's lattice of receptor and a spectrum S_t, the RDA is readily applicable to this problem. The details of which are developed in the next section.

5.2 The RDA and Signature Matching

To aid the development of the RDA for anomaly detection in spectra, a stochastic model of a spectrometer is given.

Definition 5.2. *A model spectrometer $\mathcal{M}$ generates a sequence of artificial spectra with n locations. The spectra generated at time t is written $\mathcal{M}_t$ and has the form of a Spectrum in Definition 5.1. The ith position of the model spectrum at time t is written $m_{i,t} \geq 0$. Associated with a model $\mathcal{M}$ is a rate function: $r_{\mathcal{M},t}$ which defines a Poisson rate of arrival for location $r_{m_i,t} > 0$ of chemical substance with property i at time t. Thus, model location i at time t is distributed $m_{i,t} \sim Poisson(r_{m_i,t})$.*

The above definition describes an appropriate qualitative picture for many spectrometers, and produces data in good quantitative correspondence with the ion mobility spectrometry data collected in Section 5.4. For simplicity, all instantiations of model spectrometers will be a summation of k products of Gaussians and a time profile. Ion mobility spectrometry peaks have the shape of a Gaussian [Goubran and Lawrence, 1991]. In Chapter 4 the normal distribution featured heavily as either a kernel or the underlying distribution for sample points. In this chapter the same function is used to describe the shape of a peaks in a model spectrometer rate function. To distinguish this difference in use they are referred to as Gaussians. Thus, $G(x, y)$ is a normalised Gaussian with mean x and standard deviation or width y. Then, the all rate functions have the form,

$$r_{\mathcal{M},t} = \sum_{j=0}^{k-1} G(s_k, \theta_k)\tau_k(t), \tag{5.1}$$

with s_k the spectrum location at which each Gaussian is centred, θ_k is its width, and $\tau_k(t)$ defines its time profile which describes the peak's growth and decay. For example, consider that the appearance and disappearance of a chemical substance in the sample results in sharp increase in a signature in the spectrum followed by a long decay. If the signature for the chemical substance contained two peaks, then Figure 5.1 shows an example setting of peak locations and $\tau_k(t)$ that could replicate this scenario. An full instantiation of this model will be given in a later example in Figure 5.3.

Due to the physical correspondence between a receptor lattice and the spectrum, the RDA can be mapped to take input directly from a spectrum S_t. If the receptor position at

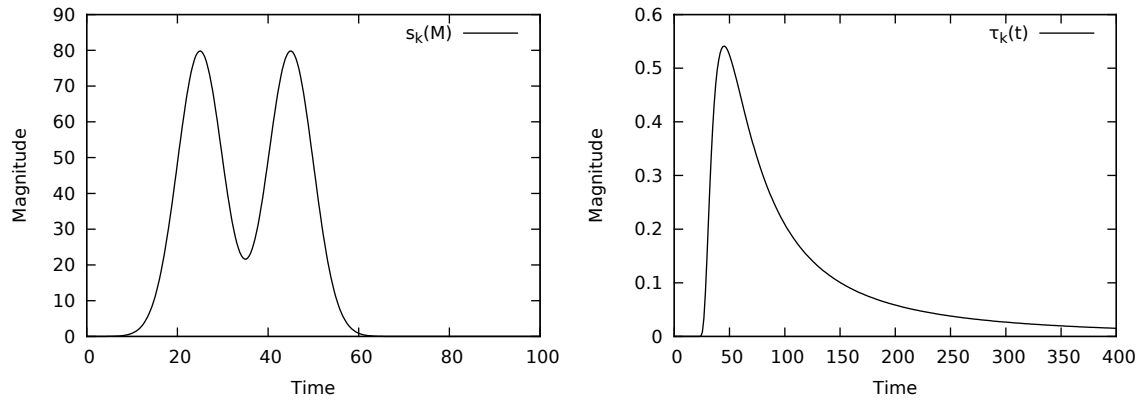

Figure 5.1: Example two-peaked model spectrometer $\mathcal{M}$ with $n = 100$. Each peak has identical time profile so $r_{\mathcal{M},t} = \tau_k(G(25,5) + G(45,5)$. The time profile $\tau(t)$ is defined by a scaled inverse-gamma distribution starting at $t = 20$ with shape and scale parameters set to 1. So, $\tau(t) = 0$ for $t <= 20$ and $\tau(t) = 2500\exp(50/(t-20))/(t-20)^2$ for $t > 20$. The use of the inverse-gamma is simply because it gives the appropriate qualitative picture. Figure 5.3 shows an instantiation of this $\mathcal{M}$.

spectrum location i at time t is written $p_{i,t}$ and the negative feedback similarly $n_{i,t}$, then,

$$p_{i,t+1} = bp_{i,t} + K_i(S_t) - n_{i,t} \tag{5.2}$$

$$n_{i,t+1} = dn_{i,t} + gH(p_{i,t} - \beta), \tag{5.3}$$

with parameters defined with identical constraints as Chapter 4. Note that the stimulation kernel function, defined (4.5), has now become a $K_i(S_t)$. This is a kernel function applied to a spectrum as follows,

$$K_i(S_t) = \frac{1}{nh}\sum_{j=0}^{n-1} f(s_{j,t})K\left(\frac{i-j}{h}\right), \tag{5.4}$$

for some suitable kernel function $K(\cdot)$ with properties (4.6) and $f(\cdot)$ is a preprocessing function. Note, K_i is very similarly defined to the negative feedback diffusion function K_N (4.7). Examples of preprocessing functions $f(\cdot)$ include a multiplication by a scalar factor, or a logarithm. If the spectra have a non-zero background signature the RDA must be capable of tracking this, so the result of applying K_i to this background signature must result in an input in region 2 of Figure 4.4: $(\max\{K\}f(s_{i,t})) \in \mathcal{U}_2$ for all i.

A discussion of appropriate parameter setting for the RDA applied to spectra is given in Section 5.2.4. Before that, a discussion of the required anomaly detection process is given.

5.2.1 The RDA for Signature Generation

In Chapter 4 a *type 2* anomaly is detected when the underlying distribution changes faster than the growth of negative feedback. However, in this chapter more rapid changes are the norm, for example the time profile of the model spectrometer presented in Figure 5.1. With such immediate changes the intuition supplied by the step-increase in u_t in Section 4.5.1

apply. The result is that the divergence of receptor position above β can be used to generate a signature for the new features in the spectrum.

Definition 5.3. *The time of an anomaly is $t_a = t_e - t_s$, with t_s the first time step when $p_{i,t_s} \geq \ell$ and t_e is the first time point after t_s when no receptor is above ℓ, that is $p_{i,t_e} < \ell, \forall i$. Further, $t_e > t_s$ and $t_a > 0$.*

Definition 5.4. *The signature of an anomaly is average of the distance of each receptor above β taken over the time of the anomaly. For an anomaly A, the signature is written $\sigma(A)$:*

$$\sigma(A) = (\sigma_{A_0}, \sigma_{A_1}, \ldots, \sigma_{A_{w-1}}) \tag{5.5}$$

$$\sigma_{A_i} = \frac{1}{t_e - t_s} \sum_{t=t_e}^{t_s-1} (p_t - \beta) H(p_t - \beta), \tag{5.6}$$

with $H(\cdot)$ the Heaviside step function.

The following will use two examples to demonstrate the issues related to signature generation. The first will concern a basic demonstration of signature generation. The second will involve the issues sequential anomalies.

5.2.2 Example 1 – Basic signature generation

Consider a model spectrometer $\mathcal{M}$ with the following rate function,

$$r_{\mathcal{M},t} = \alpha(G(60, 8.9) + G(120, 20) + H(t - 100)G(160, 8.9)), \tag{5.7}$$

with $\alpha = 18000$. The spectrum generated by $\mathcal{M}$ instantly changes from being two-peaked shaped to three-peaked at time step 100. The details of the example are given in Figure 5.2, note that they are not intended to be realistic, they are intended to demonstrate the concepts of signature detection in the RDA. There are a number of important details conveyed by Figure 5.2: First, for $t < 100$ the negative feedback familiarly matches the input, and matters are improved from the Chapter 4 as there are effectively n input points per time step rather than 1. Second, the change in $r_{\mathcal{M},t}$ results in an anomaly A with $t_s = 101$ and $t_e = 130$, and signature proportional to the new peak $G(160, 8.9)$. Third, due to the large step increase there is effectively linear growth of p_{160}, t to its peak. As the growth of negative feedback is not linear the signature is elongated, and this example has been chosen to demonstrate this point. If a time profile like that of Figure 5.1 was used, then this effect would be less pronounced. Further, note that $g = 0.3$ is larger than examples in the Chapter 4. This has three closely related influences. First, it increases the upper bound of input region $\mathcal{U}_2$ and so greater magnitude input may be held at β. Second, it increases the linear growth phase of the negative feedback which reduces the distortion of signature. Third, for input near the bottom end of $\mathcal{U}_2$ it increases the "bottom-out" effect in the equilibrium cycle (Figure 4.7) where p_t is pushed to zero. Consequently, these inputs spend the majority of their $\mathcal{U}_2$ equilibrium cycle below β and so contribute minimal noise

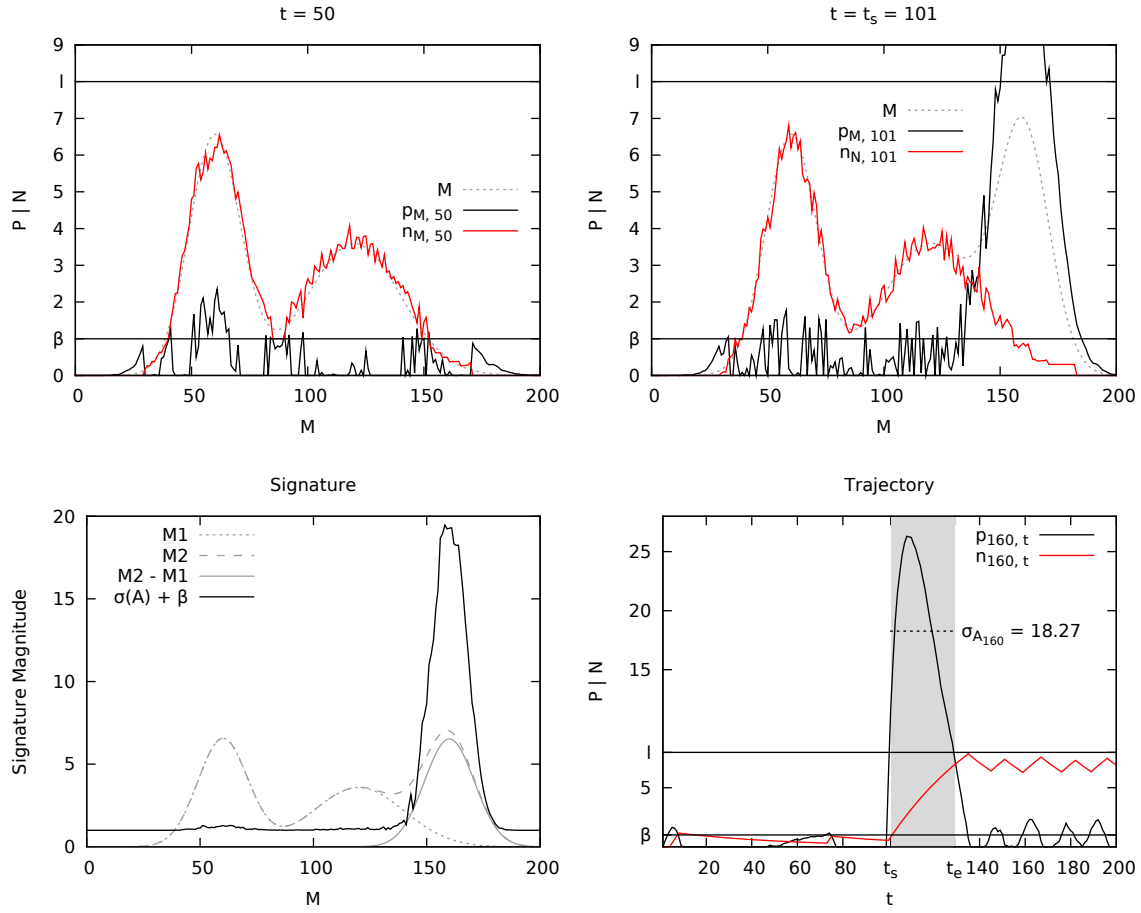

Figure 5.2: Example of signature generation. Model spectrometer $\mathcal{M}$ has rate function defined in (5.7). RDA parameters are set $b = 0.84$; $d = 0.98$; $g = 0.3$; $\beta = 1$; $\ell = 8$; $f(\mathcal{M}) = 0.01\mathcal{M}$, a standard normal kernel is used with $h = 0.5$. *Top right:* the RDA before $t = 100$; *Top left:* the RDA at $t = t_s = 101$; *Bottom Left:* The resulting signature from anomaly A caused by change $\sigma(A) + \beta$ is shown for clarity. $\mathcal{M}_1$ denotes before $t = 100$ and $\mathcal{M}_2$ denotes after. $\sigma(A)$ is proportional to the difference $\mathcal{M}_2 - \mathcal{M}_1$ above β. *Bottom right:* The trajectory of the receptor at the mode of the new peak $i = 160$.

to the signature. These benefits come at the cost of larger jumps in n_t, however as will be seen in the case studies in Sections 5.3 and 5.4 the background spectrum will be effectively zero. With no background spectra for the negative feedback to estimate, the large jumps in negative feedback are less costly.

5.2.3 Example 2 – Anomaly Interleaving

Section 5.1.1 discussed that the spectra will decay after a substance leaves the sample. Figure 5.1 gave an example of such a decay. This example addresses the case when a new anomaly occurs whilst an old anomaly is decaying. Such an occurrence is denoted *anomaly*

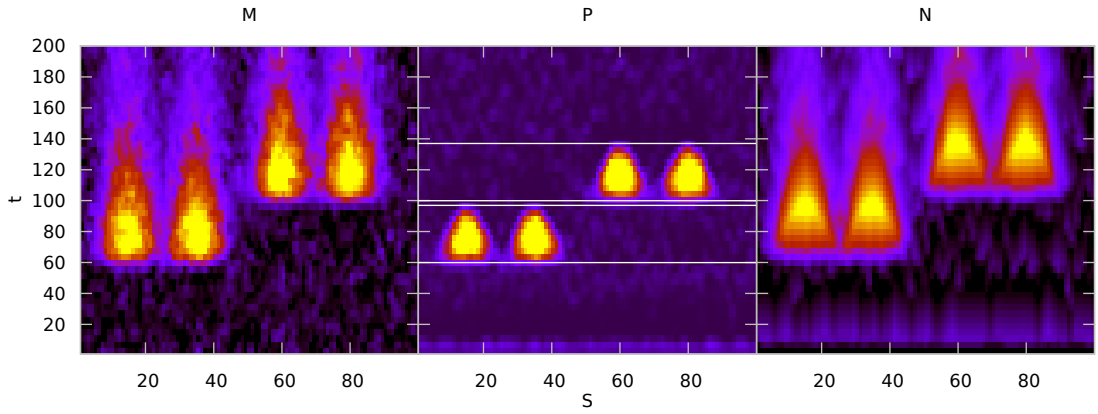

Figure 5.3: Anomaly interleaving. Model spectrometer $\mathcal{M}$ with rate function (5.8), RDA parameters are $b = 0.85$; $d = 0.98$; $g = 0.3$; $\beta = 1$; $\ell = 5$; $f(\mathcal{M}) = 0.21\mathcal{M}$; a standard normal kernel is used with $h = 0.5$. *Left:* The model $\mathcal{M}$; *Centre:* Receptor position, anomalies A_1 and A_2 are detected. Horizontal lines denote $t_{s_1} = 60$; $t_{e_1} = 97$; $t_{s_2} = 100$; $t_{e_2} = 137$. Note that no residue of the first anomaly is in the receptor position at the start of the second anomaly. Resulting signatures are given in Figure 5.4. *Right:* The negative feedback. Note it is a smoothed version of the input, and the discontinuity of the β step is clearly observable.

interleaving. Consider the following rate function for $\mathcal{M}$ with $n = 100$ locations,

$$r_{\mathcal{M},t} = \tau(t-50)(G(15,5)+G(35,5)) + \tau(t-90)(G(60,5)+G(80,5)),$$
$$\tau(t-x) = \begin{cases} 0 & \text{for } t \leq x \\ \alpha \frac{\exp(-50/(t-x))}{(t-x)^2} & \text{for } t > x \end{cases}, \tag{5.8}$$

with $\alpha = 2.5 \times 10^6$. The rate function describes two sets of two peaks that have a time profile as Figure 5.1. The first set of two peaks appears at $t = 50$, the second at $t = 90$. Figure 5.3 displays an instantiation of random model $\mathcal{M}$ with RDA receptor position and negative feedback. Note, that the sets of peaks are detected as separate anomalies. Figure 5.4 gives the two resulting signatures. Clearly, it is possible to generate signatures for interleaved anomalies only if they occupy disjoint regions of the spectrum. The proximity in time of interleaved anomalies will dictate whether they are detectable as different anomalies. This is dependent on the size of anomalies and RDA parameters. This is discussed in the next section.

5.2.4 Parameter setting for RDA signatures

No specific theory related to parameter setting for signatures is developed in this chapter. The discussion will be informal, but rely on some of the results calculated in Chapter 4. In general, the advice is as the last chapter: b, d close to 1 with $d > b$ and g near zero. The characteristics of a particular spectrometry device will force some parameter setting. For example, the resolution of the spectrum and expected horizontal noise (i.e. shifts of peak positions to left or right) will set the details of the kernel. If the spectrum is low resolution, so it is a binned view on reality, then it may be preferable to use no kernel.

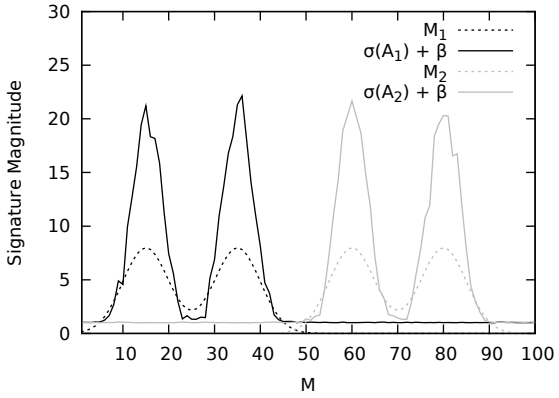

Figure 5.4: Signatures for interleaved anomalies in Figure 5.3. $\mathcal{M}_1$ and $\mathcal{M}_2$ denote the shape of the left and right sets of peaks in (5.8) respectively. Note that $\sigma(A_2)$ contains no residue of $\sigma(A_1)$.

Decay rate d should be set in reference to the expected characteristic decay of the device. In particular, the negative feedback should decay at a rate approximately equal to but slower than the decay of signatures in spectra. This ensures that whilst the anomaly decays in the spectra the anomaly is not re-detected and the receptor positions are free from any residue of the anomaly which allows interleaving with other anomalies.

As noted above, g should be set larger than in Chapter 4. Anomaly interleaving provides another argument for larger g, an increased negative feedback growth rate decreases the length, t_a, of an anomaly. Again, there is a trade-off here as t_a defines the number of spectra that contribute to a signature in (5.5).

If the device has a characteristic baseline-noise, then β should be set just above this level.

Finally, a good scheme with which to set parameters for signature generation is with reference to examples of extremes of detectable anomalies. The concerns are less for the largest anomalies, but the parameters should be set by their desired interleaving times. The small anomaly issues are more significant, setting the parameters incorrectly may mean that small anomalies are missed. The summation and convolution conditions can be used as in Chapter 4 to calculate appropriate parameters for detection of small anomalies. In general, a good scheme is to lower ℓ until the anomaly is detected and that t_a is of sufficient length to give a good signature.

5.2.5 Signature Libraries

Recall the discussion in Section 5.1.1, it is desired that an anomaly detection system is able to identify chemical substances via signatures and potentially learn the signatures of new substances. This can be addressed a training sequence of labelled spectra.

Consider a sequence of spectra $\{S_t\}$ for $t \geq 0$, and associated with this sequence are m anomaly labels, where the kth anomaly label is written $l_k = (l_{s_k}, l_{e_k}, l_{n_k})$. With $l_{s_k} \in \mathbb{Z}$ denoting the start time of the anomaly which is the first time that the anomaly is present in the spectra. Correspondingly, $l_{e_k} \in \mathbb{Z}$ denotes the end time of an anomaly, and so $l_{s_k} < l_{e_k}$. Finally, $l_{n_k} \in \mathbb{Z}$ is the anomaly type label and indicates which chemical substance produced the anomaly. As this is a training sample assume that there is control over the anomalies

such that they do not overlap, $l_{e_k} < l_{s_{k+1}}$.

The RDA may be applied to $\{S_t\}$ to generate a signature for each anomaly. Assume that the spacing between anomalies is such that the receptor position and negative feedback decay to the background spectrum between anomalies. Also assume that the RDA detects m anomalies $A_0, A_1, \ldots, A_m$, and the start and end times of anomaly A_k are written A_{s_k} and A_{e_k} respectively. The final requirement is that each of the m anomalies in the RDA overlaps with the corresponding anomaly label, so $l_{s_k} \leq A_{s_k} \leq l_{e_k}$.

The RDA may now be used to generate a set of signatures for a given chemical substance labelled c. Define $L_c = \{A_k : l_{n_k} = c\}$ as the set of all anomalies for a chemical substance with label c.

Definition 5.5. *The signature for a set of anomalies L_c is written $\sigma(L_c)$ and given by,*

$$\sigma(L_c) = \sum_{A \in L_c} f_L(\sigma(A)), \tag{5.9}$$

with $f_L(\cdot)$ a pre-processing function, for example a linear or logarithmic scaling.

Then, define the signature library for all chemical substances as $\mathcal{L} = \{\sigma(L_c) : \text{ for all } c\}$. To review, a sequence of labelled spectra can be converted into a signature library $\mathcal{L}$ with the RDA. Each library entry corresponds to the sum of the signatures of all anomalies of a particular a chemical substance. The library may be used to classify anomalies in unlabelled spectra. It is necessary to define a method for calculating a *match* between two anomalies, here the correlation (as calculated by the dot product) is chosen.

Definition 5.6. *The match between two anomalies A and B is written $\mu(A, B)$ and is the correlation between their signatures:*

$$\mu(A, B) = \frac{1}{\|\sigma(A)\|\|\sigma(B)\|} \sum_{i=0}^{w-1} \sigma_{A_i} \sigma_{B_i}, \tag{5.10}$$

with $\|\sigma(A)\| = \left[\sum_{i=0}^{w-1} \sigma_{A_i}^2\right]^{\frac{1}{2}}$ the magnitude of an anomaly's signature, and as such $\mu(A, A) = 1$.

The choice of correlation is found to be slightly more robust to horizontal shifts in the spectra than a more conventional distance metric such as Euclidean or Manhattan distance (or any other p-norm). The classification of a new anomaly A is performed by assigning A to the class of the chemical substance whose library signature has the maximum match with A. So, A is assigned to the c that produces $\max_c\{\mu(A, L_c)\}$. In the event of a tie, for example $\mu(A, L_{c_1}) = \mu(A, L_{c_2})$, then A is assigned to the library signature with the greater magnitude scaled by the number instances in each class. So, if $\|\sigma(L_{c_1})\|/|L_{c_1}| > \|\sigma(L_{c_2})\|/|L_{c_2}|$ then A is assigned to class c_1. In general, signatures with large magnitudes imply the chemical substance was in abundance in the sample and gave a good quality signature. Thus, ties are broken by assigning to the class with which there is a greater confidence in the quality of the signatures.

If this does not break the tie, then it is likely A contains signatures for both chemical substances labelled c_1 and c_2. In this event there a number of ways one could proceed, for example attempting to decompose A, or assigning A to a new c_1c_2 class. However, the details of this are left to further work. No ties of any kind, including $\mu(A, L_{c_1}) = \mu(A, L_{c_2})$, occur in the examples in this chapter.

The procedure above for calculating a signature library and the match between anomalies is relatively simplistic. Although, it performs well in the subsequent case studies. Similar to the RDA the methods for generating signatures libraries and matching mechanisms may be best chosen in reference to the characteristic behaviour of a spectrometer. Even careful consideration of which example signatures are used in the library can greatly influence performance [Bell et al., 1999]. On a related matter, and despite being relatively obvious, it is worth explicitly stating that the signature library must be regenerated for every change in RDA parameters.

Finally, an additional detail discussed in Section 5.1.1 is the ability to learn signatures for new chemical substances. The full details of the following are, again, left to further work. It is possible to use the above methods to perform on-line learning. A confidence threshold μ_T can be applied to the matches with the signature library, so if $\max_c\{\mu(A, L_c)\} < \mu_T$ then this method cannot resolve the identity of the anomaly with confidence. In this scenario the signature can potentially be assigned to a new class and given a new entry in the library. However, it is not guaranteed that this signature is representative of its class. Thus, a more appropriate solution is to place the new signature in a pool of unclassified signatures and progressively apply clustering techniques (there are many described in [Duda et al., 2001]) as new below threshold signatures arrive.

With the above definitions of signatures and signature libraries the chapter proceeds to the case studies.

5.3 Case Study 1: Anomaly Detection in Mass Spectrometry Data

This case study involves data collected from an immobile mass spectrometer. The data was and is used in the *International Conference on Artificial Immune Systems* (ICARIS) 2009/2010 DSTL sponsored *Anomaly Detection in Mass Spectra* competition. As a consequence it is available on-line [ICARIS, 2010] and may be used as a public benchmark. The data is collected from Ionicon PTR-TOF 8000 Mass Spectrometer [Ionicon, 2010]. The acronym PTR-TOF stands for Proton Transfer Reaction Time Of Flight. In essence, the technology operates by ionising molecules in the input sample, that is the proton-transfer leaves the molecules positively charged. Then, the charged ions are propelled through an electric field, their *time-of-flight* through this field is use to calculate their mass. This technology is closely related to Ion Mobility Spectrometry (see Section 5.4), but is less susceptible to distortions in the spectra due to the presence multiple different chemical substances [Mayhew et al., 2009].

The important details of the ICARIS data are outlined by the following remarks. The spectra are collected in mass/charge ratios in the range $1m/z$ to $270m/z$ at a resolution of $1m/z$, thus the spectrum width $n = 270$. The PTR-TOF 8000 [Ionicon, 2010], is capable of a resolution approximately 15 times greater than this, and so a binning operation has occurred. This fact removes the need and the validity of using a kernel in the RDA. The spectra approximately occur once every 0.34 seconds, or roughly 3Hz. The spectrometer has an air-inlet capillary, and periodically one of 6 different substances are moved past the inlet at varying rates. The 6 substances are *Shower Gel*, *Shampoo*, *Shaving Gel*, *Coffee Beans*, *Brewed Coffee*, *Olive Oil*, *Smoked Ham*. The data is augmented with one of three strength ratings *1 – weak*, *2 – medium*, *3 – strong*, along with the start time (l_{s_k}), end time (l_{e_k}) and peak time (equivalent to the time of peak in $\tau(t)$ in Figure 5.1). All data are collected in a laboratory, but based on the accompanying notes there are a number of sources of error, such as changing backgrounds, contaminated samples and unlabelled anomalies.

5.3.1 MS – RDA Application and Results

The RDA performs well at this anomaly detection task, although even with the 15 times downsampling the data is still of suitably high resolution and quality that the majority anomalies are relatively simple to detect. The major difficulty is the size 1, weak anomalies, whose magnitude is similar to the background spectrum. Figure 5.5 shows an example of the background spectrum, a strong anomaly and a weak anomaly. There are two orders of magnitude difference in the peak heights of the background spectrum and the signature of a weak anomaly. In RDA terminology, input region $\mathcal{U}_2$ must span at two orders of magnitude, and so $g/(1-d) > 10^2$. It is a challenge to satisfy this condition and retain good anomaly detection performance. One solution would be to use pre-processing function f (5.4) to apply a logarithm to the input data. The disadvantage of this procedure is that the noise is unduly scaled. The best performing solution is simply to mask out the initial portion of the spectrum. The background peaks only occur approximately at locations $< 66m/z$ and are relatively constant. Thus, the preprocessing function f simply sets spectrum locations $1 - 66$ to zero.

Parameter setting proceeds without difficulty using the scheme described in Section 5.2.4 and are given in Table 5.1. A signature library is generated via the methods given in Section 5.2.5, due to the two orders of magnitude difference between weak and strong anomalies f_L in (5.9) is set to a logarithm base 10. The application of a logarithm after signature generation does not suffer from the same noise scaling problems. A further issue is that the weak anomaly signatures tend to exhibit a subset of the peaks of the strong anomaly signatures. Classification can be much improved by generating a library for each anomaly size of each class. That is, a library of weak Olive Oil signatures, a library of medium Olive Oil signatures, etc. Then, an anomaly is said to be classified correctly if it matches the correct class regardless of size.

There are four files in the ICARIS data set and each contains variety of anomalies of all strengths and a number of interleaved anomalies. The RDA is tested on the four files in

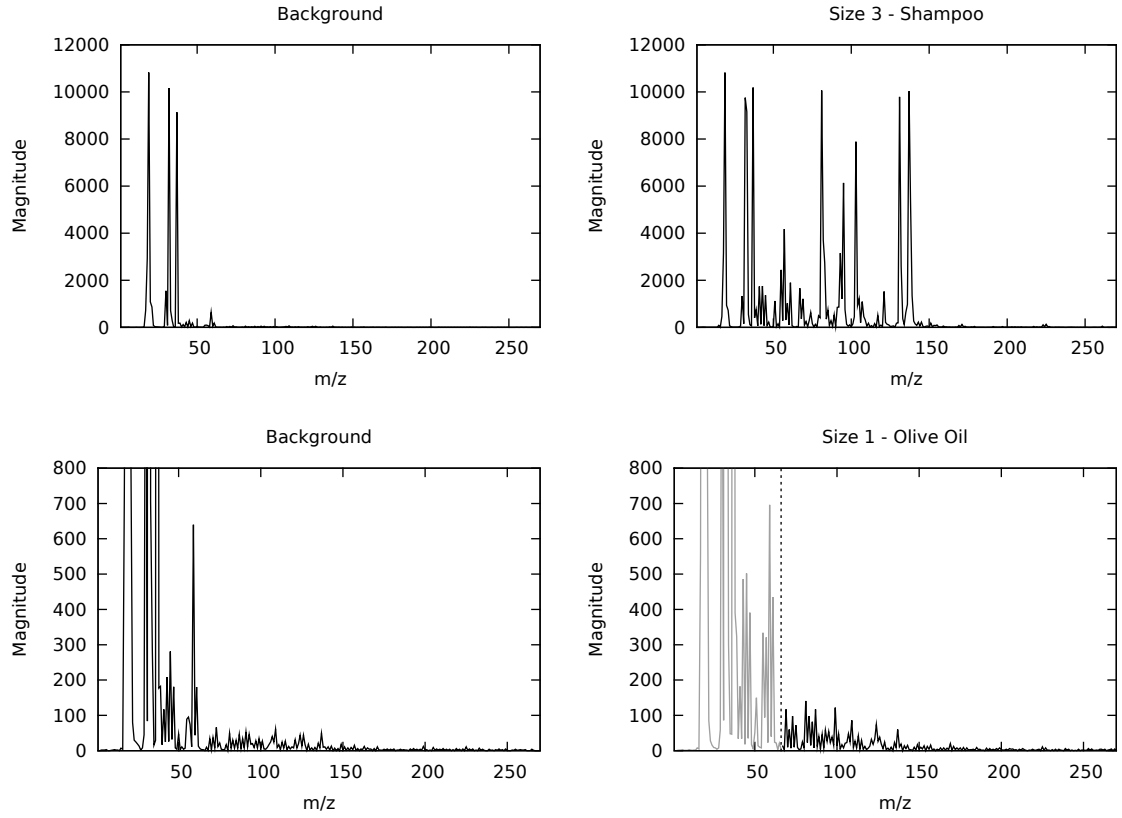

Figure 5.5: Example spectra in the ICARIS data set. *Top Right*: Background. *Top left*: Strong size 3 Shampoo. *Bottom Right*: Background, note change of scale. *Bottom Left*: Weak size 1 Olive Oil. Note change of scale and mask over $i = [1, 66]$.

Parameter	Value	Parameter	Value	Parameter	Value
b	0.82	β	0.75	ℓ	4.5
d	0.98	g	0.3	$f(S)$	$0.045S$

Table 5.1: RDA Parameters for ICARIS Data. Parameters are chosen by convolution and summation conditions given in Chapter 4 on the smallest anomalies as described in Section 5.2.4.

	Detection				Classification	
File	TP	FP	FN	Total Labelled	Correct	Incorrect
1	31	8	2	33	29	1
2	12	0	1	13	12	0
3	19	0	1	20	18	1
4	54	0	0	54	53	1

Table 5.2: Results on the ICARIS data set [ICARIS, 2010]. Shown are detection and classification results, see text for descriptions. The column "Total Labelled" gives the total anomalies in each file. True negatives are not included as they are not meaningful in this example.

terms of detection and classification. The results are presented in Table 5.2 and are given in terms of Receiver Operator Characteristics (see Section 4.6.1). The performance measures are calculated between an anomaly A generated by the RDA and a labelled true anomaly l_k. A true positive (TP) occurs when $l_{s_k} \leq t_s \leq l_{e_k}$ for some labelled anomaly with index k; a false positive occurs (FP) if an anomaly is detected and there is no matching labelled anomaly; a false negative (FN) occurs for every labelled anomaly that is not matched with a RDA generated anomaly. The classification scores are calculated for every detection of a true positive.

From Table 5.2 the RDA performs very well. The library signatures are given in Figures 5.6 and 5.7, note the clear similarity between the two types of coffee and the three toiletries. The strong anomalies are very easy to detect and classify and there are 14, 7, 9, 22 anomalies in each of the four files respectively. The mis-detections, and mis-classification are generally the weakest anomalies that have been interleaved after a larger anomaly. There are two examples of immediately interleaved anomalies, i.e. two substances presented a small number of time steps apart. The RDA detects these as one anomaly, which gives rise to one mis-detection, and the resultant signature contains the signatures of the two anomalies and so does not match with any anomaly. The other mis-classifications are generally mistaken coffees or mistaken toiletries. The majority of false positives look to be unlabelled anomalies, for example Figure 5.8. Finally, the Smoked Ham anomalies are some of the smallest in the data set, Figure 5.9 gives an example of a very small interleaved Smoked Ham anomaly. Despite size, the Smoked Ham has the advantage of a distinctive signature, see Figure 5.7, so if the ham is detected it is classifies correctly.

5.4 Case Study 2: Robot-Mounted Ion Mobility Spectrometry

This case study concerns a demonstration of the RDA applied to data on a robot-mounted Ion Mobility Spectrometer. The objective requires a Chemical Agent Monitor (CAM) Ion Mobility Spectrometer [SmithsDetection, 2010b] mounted on a Pioneer Robot [MobileRobots, 2010] (see Figure 5.10) to detect sources of Deep Heat in a noisy environment.

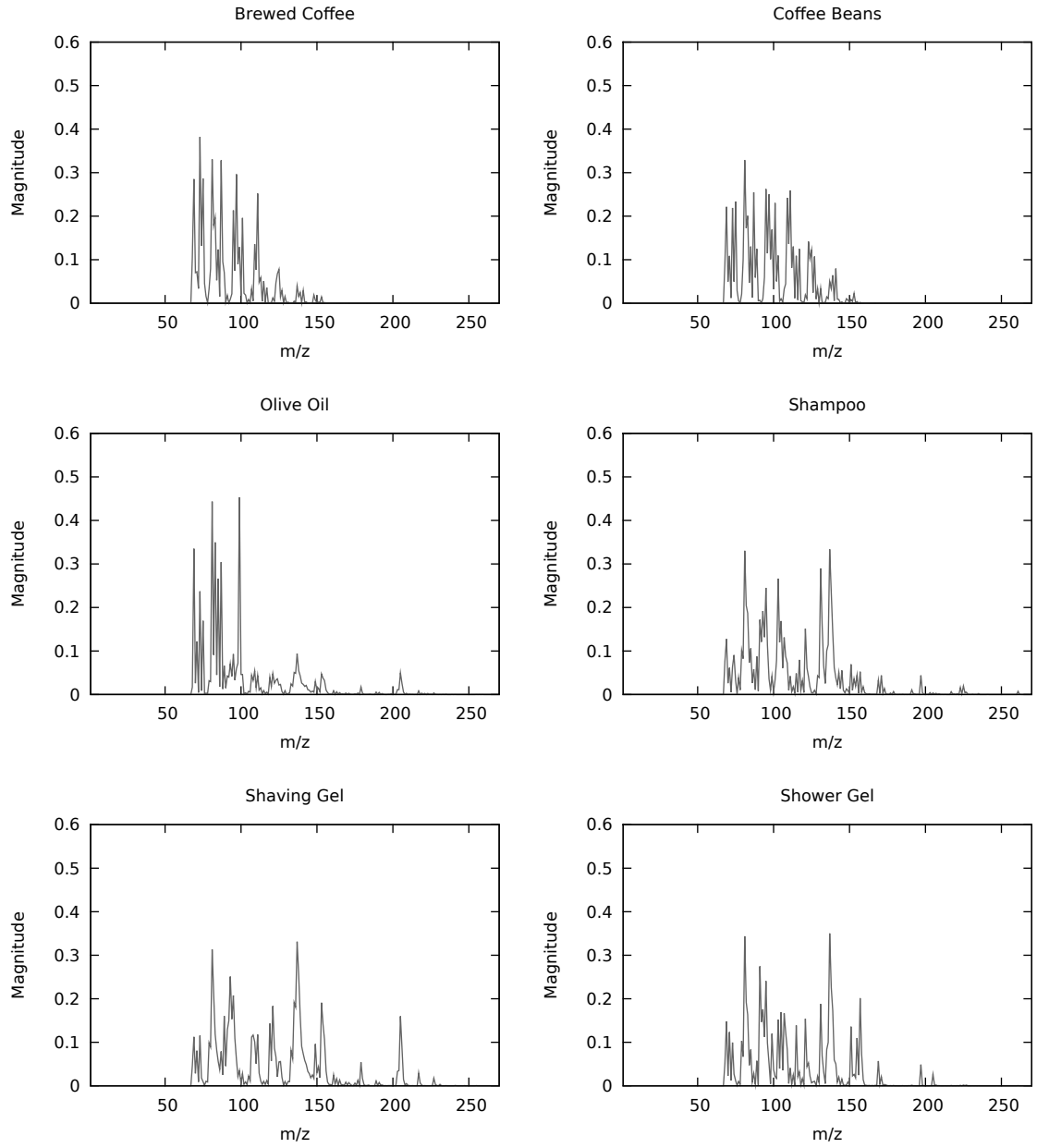

Figure 5.6: Library of Signatures for the ICARIS data set. All signatures are normalised. Note the similarity between Brewed Coffee & Coffee beans and between the toiletries. The signature for Smoked Ham is given in Figure 5.7

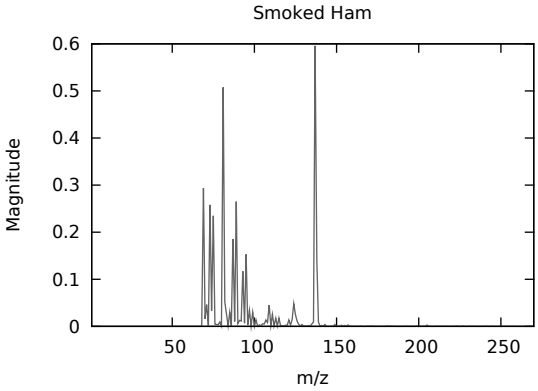

Figure 5.7: Normalised Library Signature for Smoked Ham. Other library signatures are given in Figure 5.6

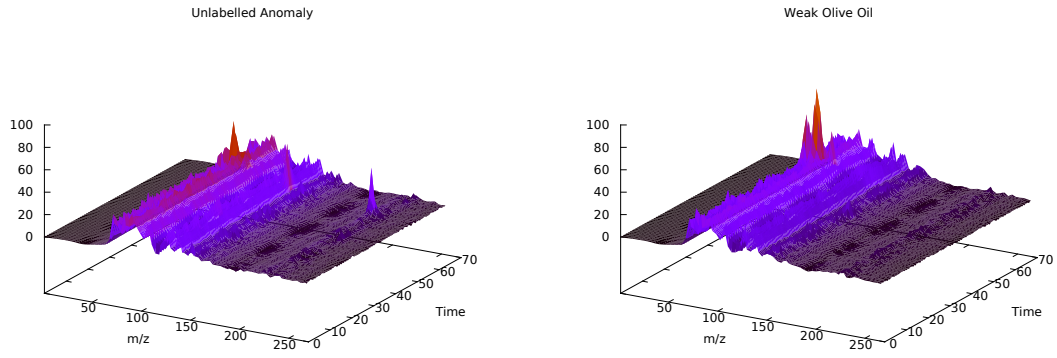

Figure 5.8: Example of a unlabelled anomaly and a similar size small anomaly for comparison. The structure of the unlabelled anomaly would suggest that it is not just noise. The weak Olive Oil anomaly is that shown in Figure 5.5.

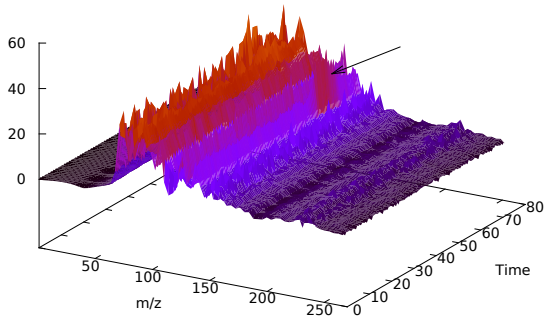

Figure 5.9: Very weak Smoked Ham anomaly. The arrow indicates the location of the distinctive peak in its signature, shown in Figure 5.7. The anomaly is interleaved the activity in the spectra before the arrow (in time) is the decaying tail of a much larger Brewed Coffee anomaly.

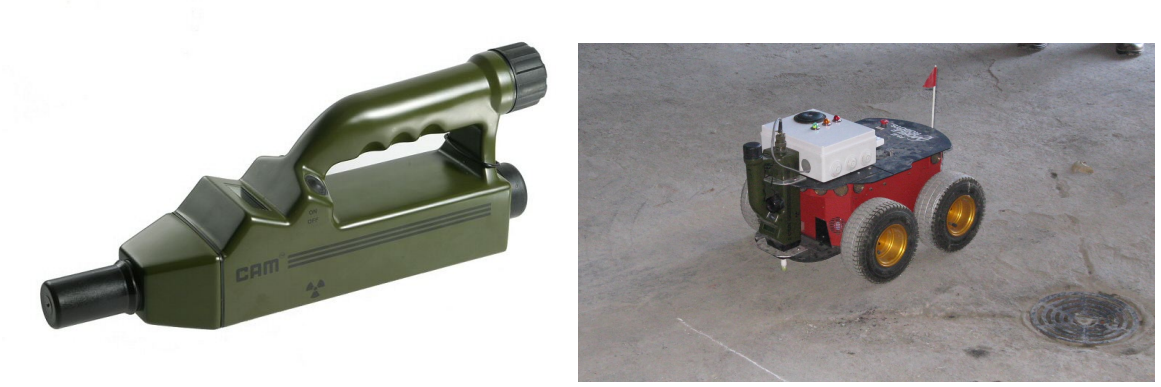

Figure 5.10: *Left:* The Chemical Agent Monitor (CAM) developed by Smith's Detection (formerly Graseby Dynamics) [SmithsDetection, 2010b]. *Size:* 350 x 80 x 145mm. *Weight:* 1.9kg (including standard battery). *Power Source*: Single 6V battery, rechargeable battery, or 4 x D-cell batteries with training adaptor. *Right:* The CAM mounted on a Pioneer Robot [Mobile-Robots, 2010].

All data processing must be performed on the robot. This section only gives comment to the data analysis and does not discuss any details of the robot platform implementation. Assuming the signature library is small, the computational complexity for anomaly detection and classification with the RDA is $O(n)$ in time and space with n the spectrum width. The pioneer robot contains a simple PC which is more than capable of performing the required computation.

5.4.1 Ion Mobility Spectrometry and the CAM

Ion Mobility Spectrometry (IMS) [Eiceman and Karpas, 2004] is a similar technology to the PTR-TOF mass spectrometer described in the last section. IMS operates by first generating *reactant ions* which then react with substances in an input sample, forming *product ions*. The product-ions are shuttered into a drift tube and subject to an electric field, their *drift-time* is the time take to travel the drift tube. Thus, IMS devices produce drift-time spectra and the drift-time of an ion is related to its *mobility*, which in turn is related to mass, charge and cross-sectional area [Eiceman and Karpas, 2004]. The advantage of IMS is that the process can be performed at atmospheric temperature and pressure, and typical drift tubes 4–20cm in length and drift velocities are typically 1–10 m/s. Thus, the devices can be small, and have a rapid response time. However, the ionisation chemistry is complicated [Eiceman and Karpas, 2004]. The presence of certain chemicals in the sample, such as solvents, can skew the spectra [Mayhew et al., 2009].

Under good conditions many ions travel down the drift tube in an *ion bundle*, for which the drift time t_d is given by,

$$t_d = \frac{d_t}{K_M E} \tag{5.11}$$

with d_t the length of the drift tube and K_M the mobility coefficient and E the strength of the electric field. The important issue here is that K_M is dependent on atmospheric temperature and pressure. A normalised mobility coefficient is often used: $K_o = K_M(273/T)(P/760)$

with T the ambient temperature in Kelvin and P the ambient pressure in Torr [Eiceman and Karpas, 2004]. Thus, the normalisation is with respect to the ambient temperature and pressure. One can potentially expect spectra to skew dependent on environmental conditions and many IMS devices have mechanisms to correct for the skew.

The CAM is a hand-held Ion Mobility Spectrometer, developed by Smith's Detection (formerly Graseby Dynamics) [SmithsDetection, 2010b], Figure 5.10 gives the dimensions. The CAM is highly successful device with over 70000 units deployed worldwide, and it is configured to detect "Nerve, blood, blister, & choking agents and selected TICs" (Toxic Industrial Chemicals) [SmithsDetection, 2010b]. The use of Deep Heat in this case study is a consequence of the CAM. Deep Heat contains Methyl salicylate which has chemical properties in common with Mustard Gas.

The difficulty of the CAM is that it is a proprietary device and out of the public domain. The specifics of its IMS are unclear, for example whether there is any correction for temperature and pressure. The data that is available from the CAM[3] is shown in Figure 5.11. The data is significantly lower is resolution with greater noise than the mass-spectrometry data in the last section. The spectra available through an analog-to-digital converter at a width of $n = 120$. The spectra are found to be oscillating around zero, and the existence of negative spectral values demonstrates that the spectra are not drift-times (which are strictly positive). It is likely that the spectra are differentials of the drift-times. This is for two reasons. First, taking differentials of spectra is a common analysis technique in the literature [Goubran and Lawrence, 1991], and second, summing out the differences smooths out the spectra and renders them positive in a manner that is appropriate for IMS [Eiceman and Karpas, 2004]. The purpose of taking the differential is to effectively double the resolution, by adding one minima for every maxima. Due to the details of the analog-to-digital converter there is an uncontrollable drift in the spectra from the CAM. This results in the need to calculate a suitable zero-point with which to sum out the differences of every spectrum. Since the purpose of the differential is to increase the resolution, it seems reasonable not to reverse this step. Taking the absolute value of the differential spectra results in a non-negative signal suitable for use with the RDA, see Figure 5.11.

A further detail of IMS chemistry is clearly visible in Figure 5.11, it is the *reactant ion-peak* (RIP), which is the large peak near the bottom end of the spectra. The RIP is the result of all reactant-ions that do not react to form product-ions with the sample. The number of product-ions formed is dependent the chemical species and the concentration of each substance present in the sample. The RIP can change in height, shape and even disappear dependent on the configuration of the sample. As a consequence changes in the RIP can be used as an indicator of changes in the chemical sample. However, the RIP presents difficulties for the RDA, the disappearance and reappearance of the RIP can cause an anomaly after a substance disappears from the sample. To alleviate this difficulty the spectra are masked in a similar manner to Section 5.3 and spectra locations 0-30 are set to zero.

[3]The author would like to thank Mark Neal for finding the appropriate pin inside the CAM that contained spectra information

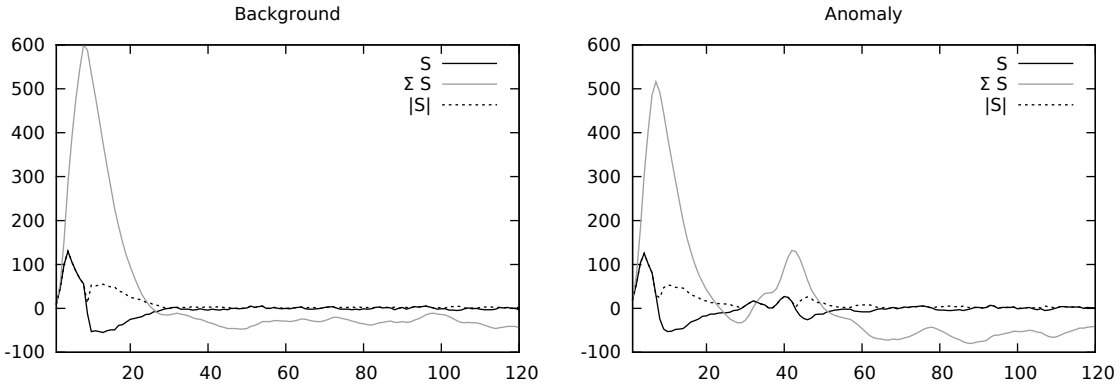

Figure 5.11: The spectra available from the CAM. *Left:* A background signature. *Right:* A background signature. S gives the spectrum available from the CAM. ΣS is the cumulative sum, so $\Sigma S_i = \sum_{j=0}^{i} S_j$. $|S|$ gives the absolute value used in the RDA. Note the reactant-ion-peak.

Parameter	Value	Parameter	Value	Parameter	Value
b	0.88	β	1	ℓ	5
d	0.98	g	0.22	$f(S)$	$0.22S$

Table 5.3: RDA Parameters for CAM Data. Parameters are chosen by convolution and summation conditions given in Chapter 4 based on a number of Deep Heat Anomalies.

5.4.2 IMS – RDA Application and Results

The case study demonstrates the detection of Deep Heat in a noisy environment. The CAM has an air-inlet (white tip in Figure 5.10) that is only able to detect chemical substances in close proximity (effectively, under the tip). As the robot moves through its environment, the task is to signal when any chemical substance is detected and to flag (literally, see Figure 5.10) when the detected substance correlates with a deep heat signature.

A range of Deep heat anomalies are collected and used to set RDA parameters, see Table 5.3. The Deep Heat anomalies are then used to construct a representative signature, see Figure 5.12. Despite the low resolution and noise of the CAM, the RDA performs well in this case study. The system detects Deep Heat with accuracy, Figure 5.13 displays the spectra and receptor position and negative feedback for a deep heat anomaly.

5.5 Conclusions

This chapter has discussed an alternative application of the Receptor Density Algorithm to the *type 1* and *type 2* anomaly detection problems in Chapter 4. The application concerns anomaly detection in data from spectrometers.

Section 5.1 gives definitions related to spectrometry and states the interest in mobile spectrometers which have reduced performance in comparison to their laboratory counterparts. Section 5.2 demonstrates how the RDA may be applied to this anomaly detection

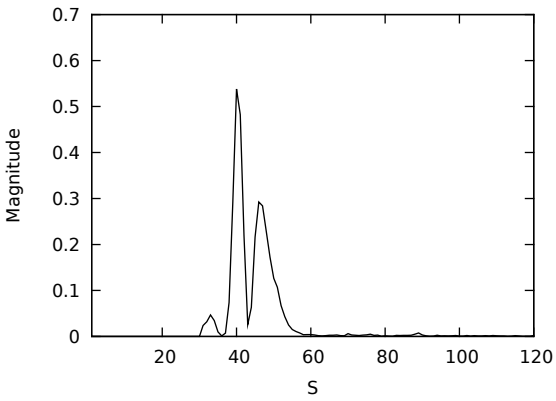

Figure 5.12: Normalised signature generated from 14 Deep Heat anomalies, note the correspondence with Figure 5.13. Compare the resolution and detail to those obtained in the mass spectrometry case study, Figure 5.6.

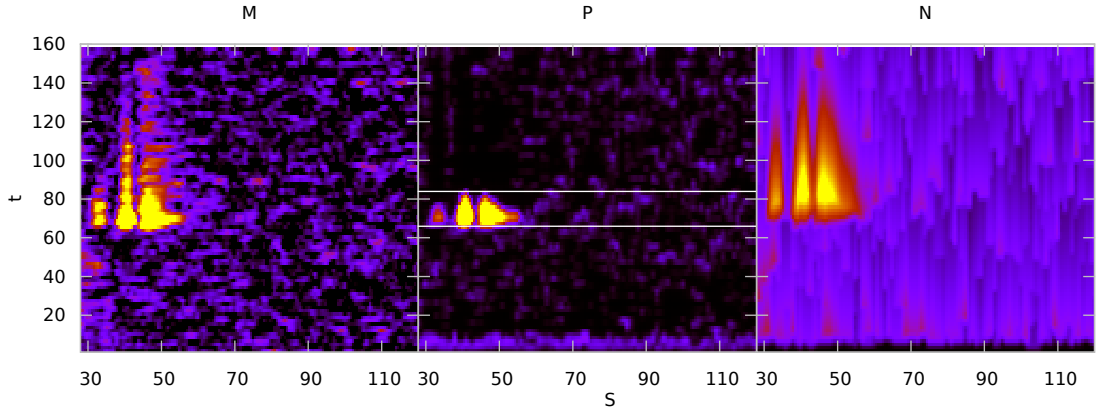

Figure 5.13: A Deep Heat anomaly. *Left:* the spectra. *Centre:* receptor position. *Right:* negative feedback. The anomaly has with $t_s = 66$ and $t_e = 84$. Note the noise removal in receptor position, and the discontinuous growth and decay of negative feedback.

problem. Demonstrations of RDA signature generation are given, including examples on the more difficult interleaved anomaly problems. Finally, Section 5.2 demonstrates how signature matching and signature library generation may be performed.

Sections 5.3 and 5.4 describe case studies. The first in Section 5.3, is the application of the RDA to mass spectrometry data. Despite downsampling of the data and anomalies showing two orders of magnitude difference in size, the RDA performs well on detection and classification. The second case study in Section 5.4 applies the RDA to a robot-mounted Ion Mobility Spectrometer. This case study does not involve an experimental evaluation of RDA performance and so is a proof-of-principle of the RDA applied in a real-world scenario.

The case studies clearly demonstrate the efficacy of the RDA at this anomaly detection task. However, many of the idea are preliminary and require significant theoretical and experimental development. In particular, a theoretical analysis that relates the spectra to the RDA signatures generated is necessary, this should give insight into parameter setting for high quality signature generation. Finally, there has been no comparison between the RDA and other approaches to detect anomalies on the same dataset. From what has been presented it is unclear whether the success of the RDA is due to the difficulty (or sim-

plicity) of the problem rather than features of the RDA. However, the RDA has received some comparison at this problem, it was the winning entry in the ICARIS 2009 DSTL *Anomaly Detection in Mass Spectra* competition. It seems that detecting and classifying the small anomalies in Case Study 1 (Section 5.3) is indeed a difficult problem at which the RDA performed better than the other competition entrants. However, a detailed experimental comparison between the RDA and alternative techniques is required to establish the applicability of the RDA to this problem.

CHAPTER 6

Conclusions

This final chapter concludes the thesis, summarises the contribution and gives evaluation and discussion of the bio-inspired algorithm design process. The chapter is organised as follows: Section 6.1 presents a summary and evaluation of the work in each chapter is presented. Section 6.2 reflects and evaluates the bio-inspired algorithm development performed. Suggestions of further work are given in Section 6.3. Finally, the hypothesis of this thesis given in Chapter 1 is reviewed in Section 6.4.

6.1 Summary and Contribution

The structure of this work of this thesis is shown in Figure 6.1. The figure depicts the flow of work from biological model, through an abstracted model to the algorithm. The F_i mark possible directions for further work. A summary of each chapter is now given and the contribution of each is then explicitly stated.

Chapter 1: A background and motivation for this thesis is presented along with reviews of bio-inspired methodologies; the current state of Artificial Immune Systems; and strategies for modelling biology. The explicit set of assumptions A that connects a biological model and the abstracted model is introduced.

Contribution: Relatively minor – making A explicit.

Chapter 2: The focus is on models of biology that are described by chemical kinetics, and particularly those defined by chemical reaction equations. The stochastic formulation of chemical kinetics is introduced and definitions of the Chemical Master Equation (CME) and the Gillespie Algorithm are given. The CME defines a continuous time Markov Chain (CTMC) and methods to numerically solve a large sparse CTMC rate matrix are presented. Finally, in reference to a cell-surface receptor, a method for scaling the volume of the cell by a factor ϵ is discussed. The method allows tractable numerical calculation of transient and stationary distributions, with the caveat that the results may not scale to the full system.

Contribution: The use of ϵ to scale the volume reaction volume Ω.

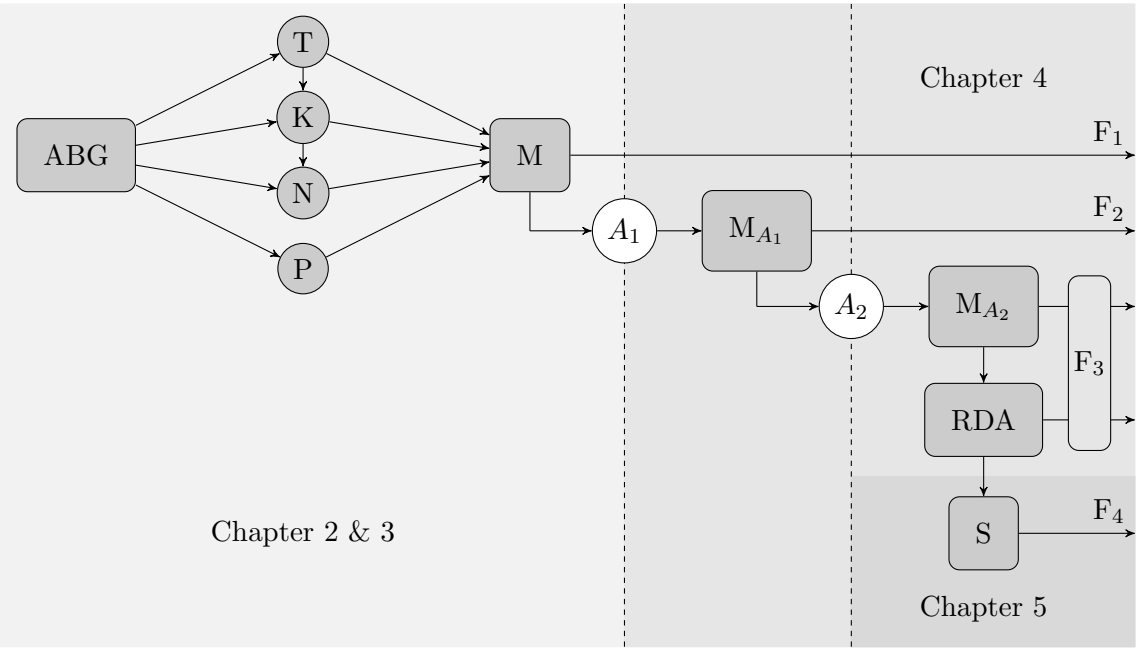

Figure 6.1: Outline of the work of this thesis. *Key:* ABG – The original model of T cell signalling [Altan-Bonnet and Germain, 2005]; T, K, N, P – Individually analysed components of the stochastic ABG model; M – The recomposed model; A_1 – First abstractions: the generalised receptor and the receptor lattice; M_{A_1} – the abstracted model after applying A_1; A_2 – The removal of negative feedback diffusion; M_{A_2} – the result of applying A_2 to M_{A_1}; RDA – The Receptor Density Algorithm; S – Application of RDA to anomaly detection in spectra. F_1, F_2, F_3, F_4 indicate further work, see Section 6.3.

Chapter 3: The biological modelling/analysis work of thesis, no discussion of algorithms is given. Initially a review is given of the salient concepts from immunology, and in particular, T cell signalling. The bulk of the chapter is concerned with the analysis of a model of T cell early signalling events originally presented in [Altan-Bonnet and Germain, 2005]. The chapter refers to this as the ABG model. The model is decomposed into reactions concerning the following: the binding of TCR, pMHC and CD8; kinetic proofreading; kinetic proofreading with negative feedback; protection of the TCR by ppERK and the MAPK cascade. These are scaled to the volume of a single TCR via ϵ, with the exception of the MAPK cascade which scaled to $\epsilon = 10^{-2}$. The most notable features discovered are the low probability of state TC; the influence of the ZAP-70 locking mechanism; and the influence of the base negative feedback state. The decomposed processes are re-composed into the full model and ϵ increased. Results at the small-ϵ level guide the simulation at the large-ϵ level, and the results are in qualitative agreement. A discussion of the chosen values of ϵ is then given. Finally, a hypothesis for T cell activation is presented:

> *Good agonists that are on average able to complete kinetic proofreading before weaker endogenous ligands are able to reach the base negative feedback state will reliably activate a T cell.*

Contribution: All results obtained at the small-ϵ scale. In particular, the definition of $\mathcal{B}_n$ binding systems; the low probability of state TC; the influence of the locking mechanism; the influence of the negative feedback base state (this was implicit in the ABG model and to the author's knowledge never explicitly discussed anywhere in T cell signalling literature); the hypothesis for T cell activation.

Chapter 4: The transition from the model in Chapter 3 to an algorithm via an abstracted model is presented. The hypothesised mechanism for T cell activation forms the abstracted model in terms of a *receptor*. Discussions of the difficulty of combining many receptors lead to an argument for a lattice of receptors. The elements in assumption set A are noted, and these will be discussed in more detail in Section 6.2. Properties of the behaviour of a single receptor are given and it is noted that a lattice of receptors has a connection to kernel density estimation. Topics related to kernel density estimation are reviewed and definitions of *type 1* and *type 2* anomalies are given. It is noted that removing the negative feedback diffusion transfers the density estimate from the receptor position to the negative feedback. This connects the abstracted model to *type 1* anomaly detection. The chapter continues with conditions for a single receptor breaking ℓ under variable input. The *Receptor Density Algorithm* is then defined. The algorithm is demonstrated on *type 1* and *type 2* anomaly detection problems. A comparison is made between the RDA and a kernel density estimate with a threshold; performance is found to be very comparable. As noted in the concluding section of Chapter 4 the behaviour of the algorithm may be described without reference to biology. This is an important step for any bio-inspired algorithm and will aid future dissemination.

Contribution: The definitions of the generalised receptor and all associated properties. The definitions of *type 2* anomalies. The definition of the *decay density estimate*. The properties calculated in relation to $w_t^\star$ and $\rho_t^\star$. Finally, the definitions of anomaly conditions of the Receptor Density Algorithm.

Chapter 5: The application of the Receptor Density Algorithm to anomaly detection in spectrometer data is presented. Discussions related to mobile spectrometers are given along with definitions of spectra. The RDA is used to smooth the noise in spectral data and provide a background subtracted signature of new features in the stream of spectra. Demonstrations are given on two case studies. The first is in part proof of concept on static immobile mass spectrometer data. The second case study concerns a robot-mounted Ion Mobility Spectrometer. Both cases studies successfully demonstrate the applicability of the RDA to this problem.

Contribution: The application of the RDA to anomaly detection in spectrometer data. The details of the signature generation were developed with discussions with James Hilder. The robot-platform implementation of the algorithm were developed by Mark Neal and James Hilder. All other analysis and results presented in the chapter are the author's own.

6.2 Evaluation the Bio-Inspired Algorithm Development

This section evaluates the bio-inspired algorithm development of the thesis. First a discussion of the biological modelling is given in Section 6.2.1. This is followed by a discussion of the assumptions A that link the biological and the abstracted model in Section 6.2.2. Finally, in Section 6.2.3 some comments are given on bio-inspired algorithm development for applications.

6.2.1 Modelling for Biology and Model Choice

It is stated in Chapter 1 and Chapter 3 that the Tuneable Activation Threshold hypothesis directed the choice of biological model. The following describes how this led to the choice of the model presented in [Altan-Bonnet and Germain, 2005]. There are many general models of tuneable activation thresholds [Grossman and Paul, 1992; van den Berg and Rand, 2004; Scherer et al., 2004; Carneiro et al., 2005], and these do not deal with tunability on the level of chemical reactions. There are also models [Chan et al., 2001; Artyomov et al., 2007; Lipniacki et al., 2008] which deal with the same negative and so-called positive feedbacks as [Altan-Bonnet and Germain, 2005] and Chapter 3, but contain less detail regarding the specifics of reactants; they are less realistic. All models contain bias in abstractions and simplifications. However, the model in [Altan-Bonnet and Germain, 2005] attempts to faithfully reconstruct the exact chemical reactions of the T cell signalling system. It can be regarded as the least biased, but also requires the most work to develop understanding. Choosing this model allows the biology to dictate the algorithmic development more than any of the alternative models. In this sense it is the next best option to developing one's own model. Of course, there is the risk that the analysis of a detailed model results in abstraction to an existing model in the field. What is lost in novelty is gained in model insight. Given the choice of model it is essential to model for the purposes of biology, as [Altan-Bonnet and Germain, 2005] demonstrate that the model can replicate the biological behaviour but it remains unclear how the signalling processes map the observed behaviour. Had this thesis moved straight to an abstracted model from [Altan-Bonnet and Germain, 2005], the details of the base negative feedback state would likely have been missed.

6.2.2 The Assumptions A

This section discusses the assumptions made to make the transition from the biological model to the abstract model. The outline of work given in Figure 6.1 depicts two sets of assumptions A_1 and A_2, these connect abstract models M_{A_1} and M_{A_2}. The model M_{A_1} is the abstract model before the removal of negative feedback diffusion, and M_{A_2} is the abstracted model after, so,

$$A_1 \ : \ \begin{cases} \text{Assumptions related to the definition of a receptor} \\ \text{The combination of many receptors in a lattice} \end{cases},$$

$$A_2 : \text{The removal of negative feedback diffusion, } h \to 0 \text{ in } K_N.$$

In A_1 the assumptions regarding the definition of a generalised receptor (for example, the continuous receptor position and negative feedback) will not be discussed further in this thesis. They are left to further work. The model M_{A_1} must remain an abstracted model and not an abstract biology model (i.e. it must remain on the right of Figure 1.1) until these assumptions are given biological discussion. The lattice of receptors is, however, given discussion. Chapter 4 introduced the lattice to avoid replicating the behaviour of the molecular recognition of the TCR. As this molecular-recognition is not explicitly modelled in Chapter 3 (it is abstracted to a dissociation rate), it seems inappropriate to include any details of molecular-recognition in the abstracted model. One could use similar reasoning to argue against the inclusion of the lattice in the abstracted model. However, the inclusion of the lattice is relatively computationally simple and, moreover, a high diffusion rate lattice can give equivalent results to a non-spatial collection of receptors. The point of interest is that the removal of the negative feedback is key in the development of an anomaly detection system, this has interesting implications for biology:

- A major difference between a lattice of receptors and the biological model is that an input u_t can only stimulate close receptors due to function K_S. Whereas, in the ABG model a pMHC ligand can dissociate and re-associate to any TCR in the population. As a consequence the influence of serial triggering is reduced in the ABG model, and so the simulation should predict a greater difficulty of activation than the hypothesis of T cell activation would suggest. Certainly, [Altan-Bonnet and Germain, 2005] derived TCR pMHC association parameters based on a surface diffusion model, however it seems necessary to return to the original model and confirm the details of the non-spatial assumptions.

- If the hypothesis for T cell activation is correct, then it would be necessary for every TCR to experience the average level of pSHP1. This allows fair comparisons between all TCRs. A high rate of negative feedback diffusion is required for this behaviour. If in reality pSHP1 does not diffuse, as in M_{A_2}, then the hypothesis of T cell activation would be correct. However, agonists would locally construct maximally dense regions of pSHP1. This would suggest that pSHP1 does indeed diffuse away, or the major interpretations of T cell signalling in the literature are incorrect. The former seems the more likely.

6.2.3 Comments on Application

Chapter 1 made arguments for an *approximate application* due to the potential bias of the application. This thesis has corroborated that fact, in the first instance the mechanism for discrimination was not clear until after the T cell model analysis. Second, even once the initial abstracted was defined it was still not clear what algorithm would be resultant. In fact, the concept of *type 2* anomaly detection arose purely from the consideration of receptors performing a *type 1* problem. The abstracted model defined its own application. Many applications have constraints on computational complexity, for example resource limitations in

time/space and the need to be applicable in high dimensions. These computational details did not become clear till late on in RDA development. Now it can be stated that the RDA is ideally suited to high data rate streams of low dimension numerical data. Thus the RDA's applicability to anomaly detection in spectra. It should be noted that this application arose by fortuitous coincidence during the development of the algorithm. Had this application been chosen from the outset, it would not be clear which biology/immunology to model. Even if T cell signalling had been chosen, it seems likely that the resultant algorithm would be very different. There is, of course, nothing to prevent a bio-inspired development directed by an application from being very successful. It is only that the approximate application allows the biology to direct proceedings and so increasing the chances of capitalising of interesting behaviour.

6.3 Further Work

The outline of work in Figure 6.1 depicts an four extensions for further work labelled F_1, F_2, F_3, F_4. Each will be discussed in turn,

$\mathbf{F}_1$ – Modelling methods and the T cell Signalling Model

The use of scale parameter ϵ lacks formal justification, predominantly qualitative arguments have been made for its use. An avenue of further work would be to investigate when, if it all, its predictions are valid.

There are many directions in which one could extend the modelling work. The most interesting of these would be further modelling and ideally experimental work to investigate the influence, and even existence, of the base negative feedback state. An alternative direction of further work would be to return to the non-spatial assumptions to investigate if any are unfounded and whether any of the above discussed implications of the abstracted model lattice apply.

$\mathbf{F}_2$ – Abstracted Model $\mathbf{M}_{A_1}$

This is the model of T cell signalling with negative feedback diffusion. An avenue of further work would be to investigate the full implications of assumption set A_1, in particular, connecting the definitions of the generalised receptor to biology. If the definitions could be justified the abstracted model M_{A_1} could prove a useful tool in returning to biology.

Figure 6.1 does not depict an abstracted model defined by a collection of non-spatial generalised receptors. This scenario was discarded as simple outlier detection, however no formal arguments have been made on this statement. Further work could calculate the exact anomaly conditions for many non-spatial receptors with shared negative feedback. The anomaly conditions could be elucidating to the hypothesis of T cell activation.

In the opposite direction, there are many biological features not included in M_{A_1}, such as the influence of co-receptor CD8. In particular, CD8 controls the re-association and kinetic proofreading speed in the biological mode. This could translate across to the abstracted

model as variable width kernel in function K_S, and variable kernel width methods have been shown to give improvements on kernel density estimation [Silverman, 1986].

$\mathbf{F}_3$ – Abstracted Model $\mathbf{M}_{A_2}$ / Receptor Density Algorithm

The abstracted model M_{A_2} and the Receptor Density Algorithm are essentially the same entity. Further work here involves additional analysis of the behaviour of the single receptor and additional consideration of the anomaly conditions of a lattice of receptors.

For example, single receptor cycles in input region 2 have not been bounded. Without close bounding to the point (n_e, p_e) the assumptions for the ℓ breaking in Section 4.5 may be unjustified. Further, the influence of the discrete jumps in the p and n recurrences with regards the uniqueness of region 2 cycles has not been investigated. More importantly, there has been no explicit investigation of a stochastic input u_t to the receptor with negative feedback. This will be required to fully understand the exact anomaly detection conditions.

There are many topics in Chapter 4 that have well established fields. Some comparisons to density estimation methods are given, but other areas remain largely without comparison. In particular, returning to exponential smoothing techniques may provide a method of reducing the bias in shifting distributions. Certainly, control theory is likely to have a perspective on the behaviour of the receptor equations. The concepts of anomaly detection are clear: β renders the receptor position independent of input and the negative feedback dependent on input; the convolution smooths the noise; the growth of negative feedback defines the maximum rate of change of underlying distribution. With these concepts, one can return to the literature on conventional techniques to investigate if better performing conventional mechanisms can be swapped in for features originating in biological detail.

$\mathbf{F}_4$ – Anomaly Detection in Spectra

The work of the Chapter 5 is partially proof of concept, there are many possible extensions. A deeper analysis of setting RDA parameters to maximise the quality of signatures is necessary. Further, the signature matching is simplistic, though good enough for purpose. An investigation of alternative signature matching techniques is necessary. Concerning implementation, certainly alternative spectrometer devices are of interest, and understanding which spectrometers produce data suitable for the RDA is an avenue of investigation. Further, implementations of the Receptor Density Algorithm in hardware could prove very practical in mobile-spectrometer data analysis. There seems no reason that the algorithm could not be implemented with a constant time complexity update step.

6.4 Concluding Statements and Returning to the Hypothesis

As a concluding comment, the relatively small detail of the maximum in expected kinetic proofreading state before the rise of expected pSHP1 in Figure 3.9, was attributed to the relatively small detail of the non-zero kinetic proofreading required to generate phosphory-

late SHP1 and so generate negative feedback. From Figure 3.9 onwards this concept into the base negative feedback state has been tremendously influential on this thesis.

Chapter 1 defined a hypothesis to guide this thesis. Reproducing here,

> *Through investigation and analysis of a realistic and precise model of T cell signalling it is possible to abstract properties and via an abstracted model design an algorithm which replicates the discrimination abilities of the T cell.*

Given the summaries of contribution of this chapter, it is the author's opinion that this thesis has been a success; the challenge of the hypothesis has been met.

APPENDIX A

Supplementary Modelling Material

A.1 Components and Rates

All components and their labels are given in Table A.1, all rates are given in Table A.2.

Component	Label	Comment
TCR	T_z	T cell Receptor complex with z ZAP-70 molecules bound.
pMHC	M	Antigenic peptide bound to major-histocompatability complex MHC
CD8	C	Co-receptor CD8
Lck	L	Leukocyte-specific protein tyrosine kinase (Lck)
SHP1	S	SH2 domain containing tyrosine phosphatase (SHP-1). Phosphorylated pSHP-1 represented as S^1
ERK	E	Extracellular signal-regulated kinase (ERK). Single and double phosphorylated versions represented as E^1 and E^2 respectively.
ZAP-70	Z	ζ-chain associated protein kinase 70 (ZAP-70)
Adapter	A	Adapter protein that initiates the MAPK cascade. Phosphorylated form represented as A^1.
Raf1	R	Part of MAPK cascade. Phosphorylated form represented as R^1.
Mek	K	Part of the MAPK cascade. Single and double phosphorylated forms represented as K^1 and K^2 respectively
MAPKPase	P_E	MAPK Phosphatase, acts dephosphorylating ERK
MAPKKPase	P_K	MAPK Phosphatase, acts dephosphorylating Mek
MAPKKKPase	P_R	MAPK Phosphatase, acts dephosphorylating Raf1
Adapter Pase	P_A	Adapter Phosphatase, acts desphosphorylating Adapter
SHP Pase	P_S	pSHP-1 Phosphatase, acts desphosphorylating S^1
TCR-pMHC	B_z^q	TCR-pMHC complex with q non-protected phosphorylations and z ZAP-70 molecules bound.
Protection	$\overline{C_z^q}$, $\overline{T_z}$	Internal TCR complex after protection by ppERK, with q non-protected phosphorylations of the ζ-chain

Table A.1: Components in the model and their labels.

Rate	Value	Comments
m_a	1×10^4	Association rate of pMHC to TCR.
m_d	–	Dissociation rate of pMHC from TCR.
c_t	1000	Association rate of CD8 to TCR.
c_a	3×10^5	Association rate of CD8 to pMHC, TCR-pMHC complex.
c_d	100	Dissociation rate of CD8 from TCR, pMHC TCR-pMHC complexes.
c_s	0.1	Stabilising effect of CD8 to TCR-pMHC complex, multiplies dissociation rate m_d.
l_a	1×10^4	Association rate of Lck to internal TCR complex.
l_d	50	Dissociation rate of Lck from internal TCR complex.
l_c	10.4	Rate of Lck phosphorylation of the internal TCR complex.
l_{ca}	50.0	Association rate of CD8 associated Lck to ITAMs, ZAP-70 etc.
l_{cd}	40.0	Dissociation rate of CD8 associated Lck from ITAMs, ZAP-70 etc.
l_{cc}	10.4	Rate of CD8 associated Lck phosphorylation of the internal TCR complex.
z_a	1.2×10^7	ZAP-70 association to phosphorylated ITAMs.
z_d	0.11	ZAP-70 dissociation from phosphorylated ITAMs.
s_a	2.9×10^5	Association rate of SHP-1 to TCR complex.
s_d	0.13	Dissociation rate of SHP-1 from TCR complex.
s_c	35.0	Rate of phosphorylation of SHP-1.
p_a	3.2×10^5	Association rate of pSHP-1 to TCR complex.
p_d	0.05	Dissociation of pSHP-1 from complex.
a_a	5.0×10^6	Association rate of adapter to TCR complex.
e_a	1.0×10^7	Association rate of ppERK to TCR complex.
e_d	2.0	Dissociation rate of ppERK from TCR complex.
e_c	3.4	Rate of protection of TCR by ppERK.
k_a	1.2×10^7	Association rate of any component of MAPK cascade.
k_d	0.15	Dissociation rate of any component of MAPK cascade.
k_a	2.6	Rate of (de)phosphorylation of any component of MAPK cascade.
q_a	1.0×10^4	Association rate of pSHP1 phosphatase.
q_d	0.15	Dissociation rate of pSHP1 phosphatase.
q_c	2.6	Rate of (de)phosphorylation of pSHP1 phosphatase.

Table A.2: Macroscopic rates from the ABG model in [Altan-Bonnet and Germain, 2005]. All association rates are given in $\text{mol}^{-1}s^{-1}$ all other rates have units s^{-1} with the exception of c_s which is dimensionless.

Component	Quantity	Comment
Ω	15×10^{-15}	Cytoplasmic volume of a T cell
N_{T}	3×10^4	Quantity of TCR
N_{M}	3×10^4	Quantity of pMHC
N_{C}	3×10^4	Quantity of CD8
N_{L}	3×10^4	Quantity of Lck
N_{S}	8×10^5	Quantity of SHP1
N_{E}	10^5	Quantity of ERK
N_{Z}	1.2×10^6	Quantity of ZAP-70
N_{A}	1.5×10^5	Quantity of Adapter
N_{R}	10^5	Quantity of Raf1
N_{K}	4×10^5	Quantity of Mek
$N_{\mathcal{P}}$	2000	$\mathcal{P} = \{\mathrm{E}, \mathrm{K}, \mathrm{R}, \mathrm{A}, \mathrm{S}\}$ Quantity of all dephosphorylating phosphatases

Table A.3: Quantity of all components found in the cytoplasmic volume Ω, taken from [Altan-Bonnet and Germain, 2005]

A.2 Negative Feedback

The action of pSHP1 on the TCR internal complex with 3 ZAP-70 molecules is a one step process, see Reaction 8 (as reproduced from [Altan-Bonnet and Germain, 2005]). The action of pSHP1 on the TCR internal chains containing 0, 1, or 2 ZAP-70 molecules is a many step process. To clarify, if N and P are non-phosphorylated and phosphorylated complexes respectively, L is Lck and S is pSHP1, in the standard case:

$$\begin{aligned} \mathrm{N} + \mathrm{L} + \mathrm{S} &\underset{l_d}{\overset{l_a}{\rightleftharpoons}} \mathrm{NL} + \mathrm{S}, & \mathrm{NL} + \mathrm{S} &\xrightarrow{l_c} \mathrm{P} + \mathrm{L} + \mathrm{S}, \\ \mathrm{P} + \mathrm{L} + \mathrm{S} &\underset{p_d}{\overset{p_a}{\rightleftharpoons}} \mathrm{PS} + \mathrm{L}, & \mathrm{PS} + \mathrm{L} &\underset{l_d}{\overset{l_a}{\rightleftharpoons}} \mathrm{PSL}, \\ \mathrm{PSL} &\xrightarrow{l_c} \mathrm{NS} + \mathrm{L}, & \mathrm{NS} + \mathrm{L} &\xrightarrow{p_d} \mathrm{N} + \mathrm{L} + \mathrm{S}. \end{aligned} \tag{A.1}$$

In the B_3 case:

$$\begin{aligned} \mathrm{N} + \mathrm{L} + \mathrm{S} &\underset{l_d}{\overset{l_a}{\rightleftharpoons}} \mathrm{NL} + \mathrm{S}, & \mathrm{NL} + \mathrm{S} &\xrightarrow{l_c} \mathrm{P} + \mathrm{L} + \mathrm{S}, \\ \mathrm{P} + \mathrm{L} + \mathrm{S} &\xrightarrow{p_a} \mathrm{N} + \mathrm{S} + \mathrm{L}, & \mathrm{N} + \mathrm{L} + \mathrm{S} &\underset{p_d}{\overset{p_a}{\rightleftharpoons}} \mathrm{NS} + \mathrm{L}. \end{aligned} \tag{A.2}$$

Rate matrix $\mathbf{R}_S$ describes the standard case:

$$\mathbf{R}_S = \begin{pmatrix} -l_a - p_a|\mathrm{S}| & l_a & 0 & 0 & 0 & p_a|\mathrm{S}| \\ l_d & -l_d - l_c & l_c & 0 & 0 & 0 \\ 0 & 0 & -p_a|\mathrm{S}| & p_a|\mathrm{S}| & 0 & 0 \\ 0 & 0 & p_d & -p_d - l_a & l_a & 0 \\ 0 & 0 & 0 & l_d & -l_d - l_c & l_c \\ p_d & 0 & 0 & 0 & 0 & -p_d \end{pmatrix}. \tag{A.3}$$

The stationary distribution given by $\boldsymbol{\pi}_{\mathbf{R}_S}\mathbf{R}_S = 0$:

$$\boldsymbol{\pi}_{\mathbf{R}_S} = \frac{\boldsymbol{\phi}}{\sum \phi_i}, \qquad \boldsymbol{\phi} = \begin{pmatrix} p_d p_a |\mathrm{S}|(l_c + l_d) \\ l_a p_d p_a |\mathrm{S}| \\ p_d^2(l_c + l_d) + p_d l_a l_c \\ p_d p_a |\mathrm{S}|(l_c + l_d) \\ l_a p_d p_a |\mathrm{S}| \\ (p_a|\mathrm{S}|)^2(l_c + l_d) + p_a|\mathrm{S}|l_a l_c \end{pmatrix}. \tag{A.4}$$

The rate matrix $\mathbf{R}_3$ describes the B_3 case:

$$\mathbf{R}_3 = \begin{pmatrix} -l_a - p_a|\mathrm{S}| & l_a & 0 & p_a|\mathrm{S}| \\ l_d & -l_d - l_c & l_c & 0 \\ p_a|\mathrm{S}| & 0 & -p_a|\mathrm{S}| & 0 \\ p_d & 0 & 0 & -p_d \end{pmatrix}. \tag{A.5}$$

The stationary distribution in this case is:

$$\boldsymbol{\pi}_{\mathbf{R}_3} = \frac{\boldsymbol{\theta}}{\sum \theta_i}, \qquad \boldsymbol{\theta} = \begin{pmatrix} p_d p_a |\mathrm{S}|(l_c + l_d) \\ l_a p_d p_a |\mathrm{S}| \\ l_a l_c p_d \\ (p_a|\mathrm{S}|)^2(l_c + l_d) \end{pmatrix}. \tag{A.6}$$

The stationary distributions exhibit differences dependent on $|\mathrm{S}|$. For example, the difference in probability of the phosphorylated state ξ is:

$$\xi = \frac{l_a l_c p_a |\mathrm{S}|}{p_d^2(l_c + l_d) + l_a l_c(p_d + l_d) + p_a|\mathrm{S}|(p_d(l_a + l_c + l_d) + l_a l_c)} \tag{A.7}$$

$$\approx \frac{10^{-10}|\mathrm{S}|}{40 + |\mathrm{S}|}. \tag{A.8}$$

The approximation is taken for CD8-associated Lck phosphorylation rates and holds for $N_T^{-1} \leq \epsilon \leq 1$. The stationary differences between the two are negligible, numerical solutions of the transient behaviour reveal larger differences however the behaviour is qualitatively identical.

A detailed look at the pSHP1 reaction equations reveals that doubly-phosphorylated ITAMs are exempt from the SHP1 negative signal. This is a consequence of the precondition for Lck binding: $q < 2$ when $z = 0, 1, 2$ and $q < 3$ when $z = 3$. A comparison is made between the negative feedback as found in the ABG model in [Altan-Bonnet and Germain, 2005] and a uniform application of negative feedback with multi-step dephosphorylation, and no state exempt from the action of pSHP1, this is termed *uniform negative feedback*. Calculating the transient distributions reveals the two methods are qualitatively similar, but uniform negative feedback (despite applying to a larger number of kinetic proofreading states) gives a higher stationary probability of expected kinetic proofreading states. Further, for ABG negative feedback the ordering of most probable m_d is preserved in the hump but

is broken as the transient moves towards stationarity. The multi-step form with uniform action on all kinetic proofreading states is deemed preferable and so all results in this thesis use uniform negative feedback.

A.3 Histogram Plots

This section describes a technique to display the simulation results of many independent runs of Gillespie Algorithm simulations, particularly, the technique is used to display the kinetic proofreading progress of many simulations. It is necessary to perform many independent runs of the stochastic Gillespie Algorithm to gain an estimate of the underlying statistics. It is possible to lose information related to rare events and bistabilities through averaging results across many simulations. Here a *histogram plot* is devised to reduce this kind of information loss. The number of TCRs in each relevant state at time t are averaged across simulation runs and plotted. Specifically, let $\mathcal{S}$ by the set of all unbound and bound TCRs in all possible phosphorylation and ZAP-70 configurations:

$$\mathcal{S} = \{\mathrm{T}_0, \mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3, \mathrm{B}_0^0, \mathrm{B}_0^1, \mathrm{B}_0^2, \mathrm{B}_1^0, \mathrm{B}_1^1, \mathrm{B}_1^2, \mathrm{B}_2^0, \mathrm{B}_2^1, \mathrm{B}_2^2, \mathrm{B}_3^0, \mathrm{B}_3^1, \mathrm{B}_3^2, \mathrm{B}_3^3\}. \tag{A.9}$$

Then, the set $\mathcal{X} = \{\mathcal{S}, \mathcal{S}\mathrm{S}^1, \overline{\mathcal{S}}\}$ is the set of all relevant TCR states, i.e. all non-pSHP1-bound and non-protected states, all pSHP1-bound states and all protected states. If K independent simulations have been performed then define the following:

$$\begin{aligned} o_i(s,t) = {} & \text{Number of TCRs in state } s \in \mathcal{X} \text{ in simulation } i = 0, \ldots, K-1 \\ & \text{at time } t, \end{aligned} \tag{A.10}$$

then a state occupancy score $\zeta(s,t)$ for state s at time t can by simply defined:

$$\zeta(s,t) = \frac{1}{K} \sum_{i=0}^{K-1} o_i(s,t). \tag{A.11}$$

The histogram plot gives $\zeta(s,t)$ for all $s \in \mathcal{X}$ and all relevant times t. Figure A.1 gives an increased detail example of the $\mathcal{S}$ section of Figure 3.16. Figures 3.19 is the other histogram plot in this thesis.

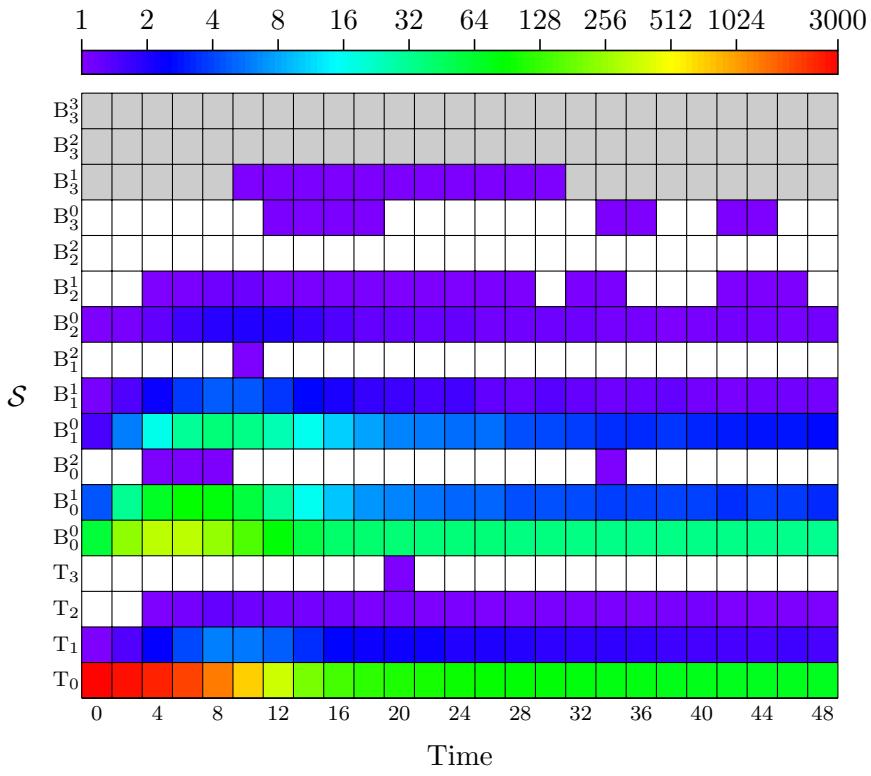

Figure A.1: Example histogram plot of the $\mathcal{S}$ section of Figure 3.16 for $t \in [0, 50]$ which plots data from 50 independent simulations of 3000 TCRs. The $\zeta(s, t)$ values are denoted by the colour bar and the kinetic proofreading signalling region (states B_3^q for $q > 0$) is shaded grey.

Bibliography

A. Adamatzky and J. Jones. Road planning with slime mould: If physarum built motorways it would route m6/m74 through newcastle. *International Journal of Bifurcation and Chaos (accepted)*, 2009.

R. A. Adams. *Calculus: a complete course*. Addison-Wesley, 1991.

Grégoire Altan-Bonnet and Ronald N Germain. Modeling T Cell Antigen Discrimination Based on Feedback Control of Digital ERK Responses. *PLoS Biol*, 3(11):e356, Nov 2005.

Paul S. Andrews. *An Investigation of a Methodology for the Development of Artificial Immune Systems: A Case-Study in Immune Receptor Degeneracy*. PhD thesis, Department of Computer Science, University of York, 2008.

Steven S. Andrews, Tuan Dinh, and Adam P. Arkin. Stochastic models of biological processes. In Robert Meyers, editor, *Encyclopedia of Complexity and System Science*. Springer, NY, 2009.

Maxim N. Artyomov, Jayajit Das, Mehran Kardar, and Arup K. Chakraborty. Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. *Proc Natl Acad Sci U S A*, 104(48):18958–18963, Nov 2007.

Mark A. Bedau. Artificial life: Organization, adaption and complexity from the bottom up. *Trends in Cognitive Sciences*, 7(11):505–512, 2003.

S. Bell, E. Nazarov, Y.F. Wang, and G.A. Eiceman. Classification of ion mobility spectra by functionial groups using neural networks. *Analytica Chimica Acta*, 394:121–133., 1999.

Claude Bernard. *Experimental Medicine*. Macmillan, 1927.

H. Bersini. Reinforcement learning and recruitment mechanism for adaptive distributed control. *Annual Review in Automatic Programming*, 17:467–473, 1992.

C. M. Bishop. *Neural Networks for Pattern Recognition.scott*. Oxford University Press., 1995.

C.M. Bishop. Novelty detection and neural network validation. *Visual, Image and Signal Processing, IEE Proceedings*, 141(4), 1994.

Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems.* Oxford University Press. USA., 1999.

S. K. Bromley, W. R. Burack, K. G. Johnson, K. Somersalo, T. N. Sims, C. Sumen, M. M. Davis, A. S. Shaw, P. M. Allen, and M. L. Dustin. The immunological synapse. *Annu Rev Immunol*, 19:375–396, 2001. doi: 10.1146/annurev.immunol.19.1.375. URL `http://dx.doi.org/10.1146/annurev.immunol.19.1.375`.

M. M. Bruce. Estimation of variance by a recursive equation. Technical report, NASA Technical Note, 1969.

Seth Bullock and Eric Silverman. Levins and the legitimacy of artifical worlds. In *Third Workshop on Epistemological Perspectives on Simulation. (Unpublished).*, 2008. In:.

F. M. Burnet. *The clonal selection theory of acquired immunity.* Vanderbilt University Press. Nashville, 1959.

Scott Camazine, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau, editors. *Self-Organization in Biological Systems.* Princeon University Press, 2001.

Walter B. Cannon. *The Wisdom of The Body.* Norton, New York, 1932.

Yang Cao, Hong Li, and Linda Petzold. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J Chem Phys*, 121(9):4059–4067, Sep 2004. doi: 10.1063/1.1778376. URL `http://dx.doi.org/10.1063/1.1778376`.

Jorge Carneiro and John Stewart. Rethinking "shape space": Evidence from simulated docking suggests that steric shape complementarity is not limiting for antibody – antigen recognition and idiotypic interactions. *Journal of Theoretical Biology*, 169:391–402, 1994.

Jorge Carneiro, Tiago Paix ao, Dejan Milutinovic, Jo ao Sousa, Kalet Leon, Rui Gardner, and Jose Faro. Immunological self-tolerance: Lessons from mathematical modelling. *Journal of Computational and Applied Mathematics*, 184:77–100, 2005.

C. Chan, A. J. George, and J. Stark. Cooperative enhancement of specificity in a lattice of T cell receptors. *Proc Natl Acad Sci U S A*, 98(10):5758–5763, May 2001.

G. Chen and P. B. Harrington. Real-time two-dimensional wavelet compression and its application to real-time modeling of ion mobility data. *Analytica Chimica Acta*, 490: 59–69, 2003.

Irun R. Cohen. *Tending Adam's Garden: evolving the cognitive immune self.* Academic Press., 2000.

Irun R. Cohen. Real and artifical immune systems: computing the state of the body. *Nature Reviews Immunology*, 7:569–574, 2007.

Joel E. Cohen. Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. *PloS Biology*, 2(12):2017–2023, 2004.

D. M. Davis and R. T. Kroutil. Application of digital filters to process data for ion mobilty spectrometry. *Analytica Chimica Acta*, 232:261–266, 1990.

L.N. de Castro and J. Timmis. *Artificial Immune Systems: A New Computational Approach.* Springer-Verlag, London. UK, 2002.

Kevin Downward. *Mass Spectrometry: A Foundation Course.* The Royal Society of Chemistry, 2004.

O.R. Duda, P. Hart, and D.G. Stork. *Pattern Classification.* Wiley & Sons., 2001.

Gerald M. Edelman and Joesph A. Gally. Degeneracy and complexity in biological systems. *Proc Natl Acad Sci U S A*, 98(24):13763–13768, 2001.

G. A. Eiceman and Z. Karpas. *Ion Mobility Spectrometry.* CRC Press, 2nd edition edition, 2004.

Saber N. Elaydi. *An Introduction to Difference Equations.* Springer-Verlag New York, 1995.

J. D. Farmer and N. H. Packard. The immune system, adaptation, and machine learning. *Physica D*, 22:187–204, 1986.

Ofer Feinerman, Ronald N Germain, and Grégoire Altan-Bonnet. Quantitative challenges in understanding ligand discrimination by alphabeta T cells. *Mol Immunol*, 45(3):619–631, Feb 2008a.

Ofer Feinerman, Joël Veiga, Jeffrey R. Dorfman, Ronald N. Germain, and Grégoire Altan-Bonnet. Variability and Robustness in T Cell Activation from Regulated Heterogeneity in Protein Levels. *Science*, 321(5892):1081–1084, Aug 2008b.

C. W. Gardner, R. Wentworth, P. J. Treado, P. Batavia, and G. Gilbert. Remote chemical biological and explosive agent detection using a robot-based raman detector. *Proceedings of SPIE. Unmanned Systems Technology X*, 6962, 2008.

Everett S. Gardner. Exponential smoothing: The state of the art – part ii. *International Journal of Forecasting*, 22(4):637–666, 2006.

R. N. Germain and I. Stefanová. THE DYNAMICS OF T CELL RECEPTOR SIGNALING: Complex Orchestration and the Key Roles of Tempo and Cooperation. *Annu Rev Immunol*, 17:467–522, 1999.

Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104:1876–1889, 2000.

Daniel T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, December 1977.

Daniel T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188: 404–425, 1992.

Daniel T. Gillespie. The chemical langevin equation. *Journal of Chemical Physics*, 113:297, 2000.

R. A. Goubran and A.H. Lawrence. Experimental signal analysics in ion mobility spectrometry. *International Journal of Mass Spectrometry and Ion Processes*, 104:163–178, 1991.

Julie Greensmith, Uwe Aickelin, and Gianni Tedesco. Information fusion for anomaly detection with the dendritic cell algorithm. *Information Fusion*, 11(1):21–34, 2010.

Z. Grossman and W. E. Paul. Adaptive cellular interactions in the immune system: the tunable activation threshold and the significance of subthreshold responses. *Proc Natl Acad Sci U S A*, 89(21):10365–10369, Nov 1992.

Z. Grossman and W. E. Paul. Autoreactivity, dynamic tuning and selectivity. *Curr Opin Immunol*, 13(6):687–698, Dec 2001.

Peter Hall and Prakash Patil. On the efficiency of on-line density estimators. *IEEE Transactions on Information Theory*, 40(5):1504–1512, 1994.

Emma Hart and Jon Timmis. Application areas of ais: The past, the present and the future. *Journal of Applied Soft Computing*, 8(1):191–201, 2008.

J.L. Hodges and E.L. Lehmann. The efficiency of some nonparametric competitors of the t-test. *Annals of Mathematical Statistics*, 27(2):324–335, 1956.

J. J. Hopfield. Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proc Natl Acad Sci U S A*, 71(10):4135–4139, Oct 1974.

C. Y. Huang and J. E. Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci U S A*, 93(19):10078–10083, Sep 1996.

ICARIS. Icaris 2010: Industry competition: Dstl: Anomaly detection in mass spectra. `http://www.artificial-immune-systems.org/icaris/2010/competition.html`, 2010. URL Valid on 29/06/2009.

Donald E. Ingber. From cellular mechanotransduction to biologically inspired engineering. *Annals of Biomedical Engineering*, 38(3):1148–1161, 2010.

Ionicon. Ptr-tof 8000 – product webpage. `http://www.ptrms.com/products/ptrtofms/ptrtof8000.html`, 2010. URL valid 29/06/2010.

N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, third edition edition, 2007.

M Kendall and J. K. Ord. *Time Series*. Edward Arnold, 1990.

Hiroaki Kitano. Biological robustness. *Nature Reviews in Genetics*, 5:826–837, 2004.

Donald E. Knuth. *The Art of Computer Programming*, volume 2. Addison-Wesley, 3rd edition, 1979.

G. A. Korn and T.M. Korn. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems and Formulas for Reference and Review.* Dover Publications., 2nd edition, 2000.

S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Paul Levi and Serge Kernbach, editors. *Symbiotic Multi-Robot Organisms*. Springer, 2010.

Richard Levins. The stratergy of model building in population biology. *American Scientist*, 54(4):421–431, 1966.

Richard Levins. A response to orzack and sober: Formal analysis and the fluidity of science. *The Quarterly Review of Biology*, 68:547–555, 1993.

I. R. Lewis. *Handbook of Raman Spectroscopy (Practical Spectroscopy).* CRC Press, 2001.

Hong Li and Linda Petzold. Logarithmic direct method for discrete stochastic simulation of chemically reacting systems. Technical report, Department of Computer Science University of California at Santa Barbara, 2006.

Tomasz Lipniacki, Beata Hat, James R Faeder, and William S Hlavacek. Stochastic effects and bistability in T cell receptor signaling. *J Theor Biol*, 254(1):110–122, Sep 2008.

P. Matzinger. The danger model: A renewed sense of self. *Science*, 296:301–305, 2002.

C. A. Mayhew, P. Sulzer, F. Petersson, S. Haidacher, A. Jordan, L. Märk, P. Watts, and T.D. Märk. Applications of proton transfer reaction time-of-flight mass spectrometry for sensitive and rapid real-time detection of solid high explosives. *International Journal of Mass Spectrometry*, 289(1):58–63, 2009.

James M. McCollum, Gregory D. Peterson, Chris D. Cox, Michael L. Simpson, and Nagiza F. Samatova. The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior. *Computational Biology and Chemistry*, 30:39–49, 2006.

Chris McEwan and Emma Hart. Representation in the (artificial) immune system. *Journal of Mathematical Modelling and Algorithms*, 8:125–149, 2009. ISSN 1570–1166.

T. W. McKeithan. Kinetic proofreading in T-cell receptor signal transduction. *Proc Natl Acad Sci U S A*, 92(11):5042–5046, May 1995.

Melanie Mitchell. *An Introduction to Genetic Algorithms.* MIT Press, 1998.

Mobile-Robots. The pioneer 3-at product page. `http://www.mobilerobots.com/ResearchRobots/ResearchRobots/P3AT.aspx`, 2010. URL Valid 29/06/2009.

Cleve Moler and Charles Van Loan. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review*, 45(1):3–49, 2003.

Kenneth M. Murphy, Paul Travers, and Mark Walport. *Janeway's Immunobiology.* Garland Science, 2008.

John Von Neumann. *Theory of Self-Reproducing Automata.* University of Illinois Press, 1966.

John Newborough and Susan Stepney. A generic framework for population-based algorithms, implemented on multiple fpgas. In *Proceedings of the Fourth International Conference on Artificial Immune Systems. (ICARIS 2005)*, 2005.

Norman S. Nise. *Control Systems Engineering.* John Wiley & Sons, 3rd edition. edition, 2000.

J. R. Norris. *Markov Chains.* Cambridge Series in Statistical and Probabilistic Mathematics (No. 2). Cambridge University Press, 1998.

Jay Odenbaugh. Complex systems, trade-offs and theoretical population biology: Richard levin's "strategy of model building in population biology" revisted. *Philosophy of Science*, 70(5):1496–1507, 2002.

Jay Odenbaugh. The strategy of "the strategy of model building in population biology". *Biology and Philosophy*, 21:607–621, 2005.

Steven H. Orzack and Elliot Sober. A critical assessment of levins's the strategy of model building in population biology (1966). *The Quarterly Review of Biology*, 68(4):533–546, 1993.

Nick D. L. Owens, Jon Timmis, Andy Greensted, and Andy Tyrrell. Modelling the Tunability of Early T Cell Signalling Events. *Proceedings of the 7th International Conference on Artificial Immune Systems (ICARIS). LNCS 5132.*, pages 12–23, 2008.

Nick D. L. Owens, Andy Greensted, Jon Timmis, and Andy Tyrrell. T-cell receptor signalling inspired kernel density estimation and anomaly detection. In *ICARIS 2009*, 2009.

Nick D L Owens, Jon Timmis, Andrew Greensted, and Andy Tyrrell. Elucidation of t cell signalling models. *J Theor Biol*, 262(3):452–470, Feb 2010. doi: 10.1016/j.jtbi.2009.10.017. URL `http://dx.doi.org/10.1016/j.jtbi.2009.10.017`.

Emanuel Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

Alan S. Perelson and George F. Oster. Theoretical studies of clonal selection: Minimal antibody repetoire size and reliability of self-non-self discrimination. *Journal of Theoretical Biology*, 81(4):645–670, 1979.

Alan S. Perelson and Gérard Weishbuch. Immunology for physicists. *Reviews of Modern Physics*, 69(4):1219–1267, 1997.

Bernard Philippe, Youcef Saad, and William J. Stewart. Numerical Methods in Markov Chain Modelling. *Operations Research*, 40:1156–1179, 1996.

Andrew Phillips and Luca Cardelli. Efficient, correct simulation of biological processes in the stochastic pi-calculus. In Muffy Calder and Stephen Gilmore, editors, *Computational Methods in Systems Biology*, volume 4695 of *Lecture Notes in Computer Science*, pages 184–199. Springer Berlin / Heidelberg, 2007.

Mark Read, Jon Timmis, Paul S. Andrews, and Vipin Kumar. A domain model of experimental autoimmune encephalomyelitis. In *Complex Systems Simulation and Modelling Workshop (CoSMoS 2009), York UK.*, 2009.

Peter T. Saunders. The organism as a dynamical system. In F. Varela and W. Stein, editors, *Thinking about Biology (SFI Studies in the Science of Complexity, Lecture Notes)*, volume Volume III. Addison-Wesley, 1993.

Almut Scherer, Andre Noest, and Rob J. de Boer. Activation-threshold tuning in an affinity model for the t-cell repertoire. *Proceedings of the Royal Society B*, 271(1539):609–616, 2004.

D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization.* John Wiley & Sons, Inc., 1992.

Roger B. Sidje. Expokit: A Software Package for Computing Matrix Exponentials. *ACM Trans. Math. Softw.*, 24(1):130–156, 1998.

Sidney Siegel. *Nonparametric Statistics for Behavioural Sciences.* McGraw-Hill, 2nd edition edition, 1988.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, 1986.

SmithsDetection. Uav technology enables rapid aerial chemical detection and tracking. Press Release. `http://www.smithsdetection.com/eng/1025_4834.php`, 2010a. URL `http://www.smithsdetection.com/eng/1025_4834.php`.

SmithsDetection. Chemical agent monitor (cam). produce webpage. `http://www.smithsdetection.com/eng/CAM.php`, 2010b. URL valid 11/05/2010.

Irena Stefanová, Bernhard Hemmer, Marco Vergelli, Roland Martin, William E Biddison, and Ronald N Germain. TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways. *Nat Immunol*, 4(3):248–254, Mar 2003.

Susan Stepney. Emboidment. In *In Silico Immunology*. Springer, 2007.

Susan Stepney, Robert E. Smith, Jonathan Timmis, Andy M. Tyrrell, Mark J. Neal, and Andrew N. W. Hone. Conceptual frameworks for artificial immune systems. *International Journal of Unconventional Computing*, 1(3):315–338, July 2005.

Thomas Stibor. *On the Appropriateness of Negative Selection for Anomaly Detection and Network Intrusion Detection*. PhD thesis, Techniche Universität Darmstadt, 2006.

Thomas Stibor. An empiricial study of self/non-self discrimination in binary data with a kernel estimator. In *Prooceedings of the 7th International Conference on Artificial Immune Systems*, 2008.

Thomas Stibor, Robert Oates, Graham Kendall, and Jonathan M. Garibaldi. Geometrical insights into the dendritic cell algorithm. In *Prooceedings of the Genetic and Evolutionary Computation Conference (GECCO-2009)*, pages 1275–1282, 2009.

Audrius B. Stundzia and Charles J. Lumsden. Stochastic simulation of coupled reaction–diffusion processes. *Journal of Computational Physics*, 2007:196–207, 1996.

Bernd Sturmfels. Can biology lead to new theorems. Clay Mathematics Institute. Annual Report., 2005.

J. Timmis, P. Andrews, N. D. L. Owens, and E. Clark. An interdisciplinary perspective on artificial immune systems. *Evolutionary Intelligence*, 1(1):5–26, 2008a.

Jon Timmis. Artificial immune systems – today and tomorrow. *Natural Computation*, 6: 1–18, 2007.

Jon Timmis, Emma Hart, Andy Hone, Mark Neal, Susan Stepney, and Andy Tyrrell. Immuno-engineering. In *2nd IFIP International Conference on Biologically Inspired Collaborative Computing, 20th IFIP World Computer Congress.*, 2008b.

Jon Timmis, Andy Hone, Thomas Stibor, and Edward Clark. Theoretical advances in artifical immune systems. *Theoretical Computer Science*, 403:11–32, 2008c.

R. S. Tsay. *Analysis of Financial Time Series*. A Wiley-Interscience Publication Wiley & Sons., 2002.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer New York, 2009.

A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transaction of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952.

A. M. Turing. Intelligent machinery. In D. C. Ince, editor, *Collected Works of A. M. Turing: Mechanical Intelligence*. Elsevier, 1992.

S. Valitutti, S. Müller, M. Cella, E. Padovan, and A. Lanzavecchia. Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature*, 375(6527):148–151, May 1995.

Hugo A. van den Berg and David A Rand. Dynamics of t cell activation threshold tuning. *Journal of Theoretical Biology*, 228:397–416, 2004.

Hugo A. van den Berg, Linda Wooldridge, Bruno Laugel, and Andrew K Sewell. Coreceptor CD8-driven modulation of T cell antigen receptor specificity. *J Theor Biol*, 249(2):395–408, Nov 2007.

Andra Vargha and Harold D. Delaney. A critique and improvement of the "cl" common language effect size statistics of mcgraw and wong. *Journal of Educational and Behavioural Sciences*, 25(2):101–132, 2000.

Dario A A Vignali, Lauren W Collison, and Creg J Workman. How regulatory t cells work. *Nat Rev Immunol*, 8(7):523–532, Jul 2008. doi: 10.1038/nri2343. URL `http://dx.doi.org/10.1038/nri2343`.

C. H. Waddington. *The strategy of Genes: A discussion of some aspects of Theoretical Biology*. George Allen & Unwin Ltd., 1957.

D.H.D. West. Updating mean and variance estimates: an improved method. *Communications of the ACM*, 22(9):532–535, 1979.

M. Westhoff, P. Litterst, L. Freitag, W. Urfer, S. Bader, and J-I Baumbach. Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study. *Thorax.*, 64:744–748, 2009.

C. T. Wolverton and T. J. Wagner. Asymptotically optimal discriminant functions for pattern classification. *IEEE Transactions on Information Theory*, 15(2):258–265, 1969.

J. R. Wyer, B. E. Willcox, G. F. Gao, U. C. Gerth, S. J. Davis, J. I. Bell, P. A. van der Merwe, and B. K. Jakobsen. T cell receptor and coreceptor CD8 alphaalpha bind peptide-MHC independently and with distinct kinetics. *Immunity*, 10(2):219–225, Feb 1999.

Dennis C Wylie, Jayajit Das, and Arup K Chakraborty. Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. *Proc Natl Acad Sci U S A*, 104(13):5533–5538, Mar 2007.