Modeling the Role of the Basal Ganglia in Motor Control and Motor Programming

by

Zhi-Hong Mao

S.M. (2000), Massachusetts Institute of Technology

Submitted to the Harvard-MIT Division of Health Sciences and Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

©Massachusetts Institute of Technology 2005. All rights reserved.
Modeling the Role of the Basal Ganglia in Motor Control and Motor Programming

by

Zhi-Hong Mao

Submitted to the Harvard-MIT Division of Health Sciences and Technology on August 22, 2005, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

The basal ganglia (BG) are a group of highly interconnected nuclei buried deep in the brain. They are involved in an important range of brain functions, including both lower-level movement control and higher-level cognitive decision making. Dysfunction of the BG has been linked to the human neurological disorders such as Parkinson’s disease, Huntington’s disease, and schizophrenia.

This thesis proposes a unified functional model of the BG, called multi-input multi-output adaptive switching (MIMOAS) model that attempts to account for the role of the BG in both higher-level rote behavior and lower-level motor control. In the model, BG circuitry effectively implements a large set of parallel noncompetitive logical OR and NOR circuits that can be driven by specific patterns of cortical activity. These afford selective gating of target thalamocortical neurons. This process can be viewed as a general mapping between binary context and response vectors. The mapping is proved to be learnable via a reinforcement mechanism that is consistent with actions commonly proposed for nigro-striatal dopaminergic pathways in the striatum and homeostasis in synaptic physiology. It appears that the cortico-striatal connections provide a biologically plausible realization of winner-take-all dynamics that is different from many engineering alternatives implementing the same function.

With the winner-take-all units as functioning as a hidden layer, using corticostriatal weights as the only tunable parameters, the adaptive BG network can develop the capacity to perform universal binary mappings. In this way, the model can simulate important features of procedural learning in human experiments. At the same time, it can be shown that derangement of the winner-take-all dynamics could underlie the tremor and rigidity seen in Parkinson’s disease.

Thesis Supervisor: Steve G. Massaquoi
Title: Assistant Professor of Health Sciences and Technology and Electrical Engineering and Computer Science

Thesis Supervisor: Eric M. Feron
Title: Associate Professor of Aeronautics and Astronautics
Acknowledgments

First I would like to express my deepest gratitude to my two advisors and Ph.D. thesis supervisors, Prof. Eric Feron and Prof. Steve Massaquoi. Eric is such a caring mentor to me. It is his strongest support, both intellectual and emotional, that have helped me through these unforgettable seven years at MIT (Eric was also my Masters thesis supervisor). I benefited greatly from his wild and insightful imagination and constant enthusiasm in research. He is my perfect role model in academia as well as in life. Steve is my first and, by all means, best teacher of neuroengineering. He introduced me to this fantastic world of brain research, and this thesis grew out of his original idea on the basal ganglionic function. It is his knowledge, wisdom, and inspiration that have guided me to the frontier of scientific inquiry. His very pleasant personality has certainly made such a journey extremely enjoyable. I will always be indebted to Eric and Steve for everything they have so kindly offered me.

I would sincerely thank Prof. Munther Dahleh, my Ph.D. thesis advisor, and Prof. Dava Newman, my academic advisor at HST. I would thank Munzer for his most generous and timely offer of help and trust. As a mentor to me, Munzer is simply perfect. As a student of his, I have been deeply impressed by his profound insight in both control engineering and systems biology—I have learned a lot from him. I would like to thank Dava for her constant support these years at HST. She is a wonderful advisor to me, and her heartwarming words of encouragement have saved my confidence in medical course study as well as in job application.

I appreciate from heart the valued mentoring (and even parenting) from Prof. Erol Gelenbe of Imperial College London and Prof. Yan-Da Li of Tsinghua University. I am really honored to have been their student, and I also feel extremely lucky to have kept a long time of pleasant contact and collaboration with them.

I wish to thank Prof. Emilio Frazzoli of UCLA, Dr. John Hu of Energied Technologies, Prof. Yi Ma of UIUC, and Dr. Liang-Jie Zhang of IBM Watson Research Center. They have been both friends and mentors to me, and have been such an endless source of support that I can always rely on. I would also like to express sincere gratitude to Dr. Zhongwei Li and Dr. Chi-Sang Poon for their warm and tremendous help through my early days at MIT.

I want to thank Ola Ayaso (really appreciate your warm encouragement), Dr. Karl Bilimoria, Animesh Chakravartly, David Dugail, Sungho Jo (cannot forget your delightful company and collaboration and our regular meetings and dinners together), Dr. Xuan Kong, Xiaojun Liang, Eugene Lim, Tri Ngo (Tri did excellent programming and experimental work, and helped me a lot in the study of procedural learning), Kazutaka Takahashi (we have been very good friends since we were offecmates at 35-223, and I am sure we will continue our collaboration in the future), Ji Hyun Yang (Ji Hyun did excellent work and helped me a lot in the study of EEG and driver drowsiness detection), and Jijun Yin (my old buddy from Tsinghua University) for being closely working with me and being so supportive to me these years.

It is always so wonderful to be surrounded by a group of intelligent friends. I wish to take this opportunity to thank them for making my life at MIT color-
ful and enjoyable. They are (besides the names that I already mentioned above): Dr. Jan De Mot, Dr. Vladislav Gavrillets, Dr. Sommer Gentry, Shan-Yuan Ho, Masha Ishutkina, Prof. Fadi Karameh, Georgios Kotsalis, Dr. Bernard Mettler, Dr. Jae-Hyuk Oh, Tom Schouwenaars, Dr. Tiangang Shang, Eric Smith, Dr. Kyungyeol Song, Funxin Sun, Mario Valentí, Dr. Yong Wang, and Dr. Daxi Xiong. I would also like to thank my friends in my offices, Eric’s group, Steve’s group, Munzer’s group, HST, Course 16, Mount Auburn Hospital, Westgate, and in particular the Chinese student society.

I want to thank the administrative staff of HST and LIDS, especially, Domingo Altarejos, Lisa Gaumond, Eileen Ng Ghavidel, Doris Inslee, Fifa Monserrate, and Ron Smith. They were very helpful and supportive, and made my study at MIT and the Harvard Medical School easier.

I would like to acknowledge the financial supports for my graduate study. The financial supports by NASA Ames Research Center (grant numbers NCC2-1044 and 011779-001), Air Force Office of Scientific Research (UCLA MURI, grant number F49620-01-1-0361), and Air Force Research Laboratory (UIUC MURI, grant number F49620-02-1-0325) are greatly appreciated. I am also very grateful for the support from the David and Beatrice Yamron Fellowship and the Medical Engineering and Medical Physics Fellowship, both awarded by HST, and the Zakhartchenko Fellowship awarded by MIT.

Finally I thank my dear family. I owe myself to my wife Ying Huang, my baby Dou-Dou Mao, my parents Ruiyang Mao and Weijun Zhang, and my little sister Wendy Mao, whose love is always with me. I want to say it again: “Without you, my personal accomplishments (if any) would be meaningless.” This thesis is dedicated to them.
Contents

1 Introduction ...................................................... 13
  1.1 Anatomy of Basal Ganglia ................................... 15
  1.1.1 Direct and Indirect Pathways .............................. 18
  1.1.2 Multiple Parallel Loops ................................... 19
  1.2 Basal Ganglia Function and Models ......................... 20
  1.3 Thesis Contribution and Outline ............................ 23

2 Multi-input Multi-output Adaptive Switching Model of Basal Ganglia ...................................... 27
  2.1 Basal Ganglia Input and Output .............................. 29
    2.1.1 Selection and Winner-take-all Competition in Striatum ...... 29
    2.1.2 Reverberatory Thalamocortical Interaction and Cerebellar Co-processing ...................................... 31
    2.1.3 Permissive Role of Basal Ganglia Output .................. 34
  2.2 Logic-like Operation of Basal Ganglia ....................... 35
    2.2.1 Indirect Networks and Role of STN ....................... 35
    2.2.2 Internal Switching ........................................ 37
    2.2.3 Parallel Modular Architecture ............................ 40
    2.2.4 Effectively Binary Context Input and Net Logic-like Operation ........................................ 41
    2.2.5 Volume Convergence in Basal Ganglia .................... 43
  2.3 Adaptation Mechanism ........................................ 44

3 Pattern Classification and Competitive Learning of Basal Ganglia ................................... 45
  3.1 Simplified Equations for Basal Ganglia Operations .......... 45
  3.2 Basal Ganglia as Universal Pattern Classifier ............... 49
  3.3 Competitive Learning of Corticostriatal Networks .......... 56
    3.3.1 Three-Factor Rule of Competitive Learning .............. 56
    3.3.2 Learning Rule Combining Three-factor Rule and Synaptic Normalization ...................................... 59
    3.3.3 Convergence of Competitive Learning .................... 62

4 Dynamics of Winner-take-all Competition in Striatal Networks ................................... 65
  4.1 Equations for Neuronal Networks with Lateral Inhibition .... 67
  4.2 Existence, Uniqueness, and Stability of Equilibrium ........ 69
    4.2.1 Existence of Equilibrium ............................... 69
4.2.2 Isolated Equilibria and Uniqueness of Equilibrium ... 70
4.2.3 Stability of Equilibrium ........................................ 74
4.2.4 Examples of Equilibria ......................................... 77
4.3 Winner-take-all Competition ...................................... 80
4.3.1 Order Preserving Equilibrium ................................. 81
4.3.2 Increased Differences Between Neuronal Activities ....... 83
4.4 Dynamics of Winner-take-all in a Limiting Situation ....... 87
4.4.1 Concepts of Filippov Solutions ............................. 88
4.4.2 Winner-take-all in a Limiting Situation ................. 89
4.5 Discussion .......................................................... 99
4.5.1 Hysteresis ..................................................... 99
4.5.2 Selection Reliability and Efficiency in Striatal Networks . 101

5 Role of Basal Ganglia in Procedural Learning 103
5.1 Frontal Registers Targeting Parietal Circuits ................ 104
5.2 Simulations ...................................................... 108
5.2.1 Procedural Learning of a Sequence of Movements ....... 108
5.2.2 Derangement in Procedural Learning ..................... 112
5.2.3 Regeneration of a Movement Sequence ................. 112
5.2.4 Learning General Vector Sequences .................. 114
5.3 Discussion .......................................................... 117
5.3.1 Sensitivity of Procedural Learning Efficiency to Sequence Complexity ...................... 117

6 Role of Basal Ganglia in Ballistic and Cruise Movement Generation 123
6.1 Sensorimotor Cortical Servo Control of Single Joint Motion . . . . 123
6.2 Simulations ...................................................... 125
6.2.1 Automated Cruise Movements .......................... 125
6.2.2 Derangement of Movement in Disease ................ 127
6.2.3 Sensitivity Analysis .................................... 127
6.3 Discussion .......................................................... 133
6.3.1 Speed and Postural Control and Dyscontrol ........ 133
6.3.2 Internal Neuronal Signal Characteristics .............. 135

7 More Discussion on MIMOAS Model 137
7.1 Robustness and Binary Formalism ............................. 137
7.2 Modeling Assertions and Suggested Neuroanatomical Connections . 138
7.3 Frontocortical-Basal Ganglionic Interaction .................. 141

8 Conclusions .......................................................... 146

A Properties of Filippov Solutions 149

B Main Abbreviations Used in Text 152
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Basal ganglia anatomy.</td>
<td>16</td>
</tr>
<tr>
<td>1-2</td>
<td>Direct and indirect pathways of the basal ganglia.</td>
<td>18</td>
</tr>
<tr>
<td>2-1</td>
<td>Winner-take-all competition in the corticostriatal network.</td>
<td>30</td>
</tr>
<tr>
<td>2-2</td>
<td>Proposed reverberatory interaction between cortical and thalamic units.</td>
<td>32</td>
</tr>
<tr>
<td>2-3</td>
<td>Interchangeable schematic representations for thalamocortical modules.</td>
<td>33</td>
</tr>
<tr>
<td>2-4</td>
<td>Logic-like processing by principal internal connections of the basal ganglia.</td>
<td>38</td>
</tr>
<tr>
<td>2-5</td>
<td>Input-output mapping of an individual neuron.</td>
<td>40</td>
</tr>
<tr>
<td>3-1</td>
<td>The simplified binary input-output operation of a basal ganglia channel.</td>
<td>47</td>
</tr>
<tr>
<td>3-2</td>
<td>Striatal neuronal connections and neurotransmitters.</td>
<td>58</td>
</tr>
<tr>
<td>4-1</td>
<td>Dynamics of a two-neuron network with only one equilibrium.</td>
<td>78</td>
</tr>
<tr>
<td>4-2</td>
<td>Dynamics of a two-neuron network with three equilibria.</td>
<td>78</td>
</tr>
<tr>
<td>4-3</td>
<td>Dynamics of a two-neuron network with infinite number of equilibria.</td>
<td>79</td>
</tr>
<tr>
<td>4-4</td>
<td>Distribution of the network inputs versus individual elements of an order preserving equilibrium.</td>
<td>84</td>
</tr>
<tr>
<td>4-5</td>
<td>Winner(s)-take-all competition in a network with lateral inhibition.</td>
<td>86</td>
</tr>
<tr>
<td>4-6</td>
<td>Dynamics of a two-neuron network with discontinuous activation function.</td>
<td>98</td>
</tr>
<tr>
<td>4-7</td>
<td>Dynamics of a three-neuron network with discontinuous activation function.</td>
<td>98</td>
</tr>
<tr>
<td>4-8</td>
<td>Hysteresis in two-neuron networks with lateral inhibition.</td>
<td>100</td>
</tr>
<tr>
<td>5-1</td>
<td>Operation of a register in response to a specific cue.</td>
<td>104</td>
</tr>
<tr>
<td>5-2</td>
<td>Operation of several registers with partially overlapping inputs.</td>
<td>105</td>
</tr>
<tr>
<td>5-3</td>
<td>Proposed interaction of the basal ganglia and frontal cortex in the &quot;programming&quot; of single-joint movements under stabilized long-loop proprioceptive control.</td>
<td>107</td>
</tr>
<tr>
<td>5-4</td>
<td>Experimental and simulated procedural learning during SRT (serial reaction time) task using sequences of different lengths.</td>
<td>110</td>
</tr>
<tr>
<td>5-5</td>
<td>Examples of some register activities during SRT task procedural learning.</td>
<td>111</td>
</tr>
<tr>
<td>5-6</td>
<td>Experimental and simulated procedural learning during SRT task in health, Parkinson’s disease, and cerebellar degeneration.</td>
<td>113</td>
</tr>
<tr>
<td>5-7</td>
<td>Learning a multi-step movement.</td>
<td>115</td>
</tr>
</tbody>
</table>
5-8 Activities of cortical units during learning of a vector sequence. ..... 116
5-9 Hypothetical reduction of response time in a SRT test. .............. 119
5-10 State transitions during a SRT test with A-B-C-C-B-A-B-C-A as the repeating sequence. .................................................. 121

6-1 Simulated ballistic movement and cruise movements in health and Parkinson's disease. ............................................................... 126
6-2 Sensitivity test on a multi-step motion with high frequency (band) Gaussian noise added to all basal ganglia units. ..................... 128
6-3 Sensitivity test on a simple ballistic (one-step) motion with high frequency (band) Gaussian noise added to all basal ganglia units. ... 129
6-4 Sensitivity test on a multi-step motion with a pure delay added in neuronal signal conduction in the cortico-basal ganglionic-thalamocortical loop. ................................................................. 130
6-5 Sensitivity test on a multi-step motion with neuronal membrane time constant increased 25% of the original value. ..................... 131
6-6 Sensitivity test on a multi-step motion with inertia of plant increased to 500% of the original value. ........................................... 132
List of Tables

3.1 Main notation used in this chapter. ................................. 48
3.2 Effects of the three factors and dopamine receptor ratio on corticostriatal synaptic plasticity. ................................. 57
Chapter 1

Introduction

The study of human motor control is to understand how the motor system of the brain and spinal cord controls movements and maintains balance and postures. Since motor system anatomically connects with other neural systems intensively and functionally lies between consciousness and the physical environment, study of motor control is likely to give insight into mechanisms serving more general brain functions [113].

Although the importance of motor control can be hardly underestimated—just imagine what life would be if one could not move, people may have the bias that motor control takes very little “intelligence.” This bias might come from the effortlessness with which we carry out complicated motor tasks without thinking about the actual joint motions or muscle contractions. However, motor control actually requires highly sophisticated computational and cognitive mechanism [158]. Modern intelligent robots are still not able to perform many motor tasks that most 5-year-olds can do reasonably well, while the contemporary computer systems can already win grand-masters of chess (IBM’s Deep Blue beat World Champion Garry Kasparov in 1997 [62]). So it is not that motor performance is computationally undemanding. On the contrary, it is so demanding that we are only beginning to understand it [158].

In the study of motor control as well as other brain functions, a major difficulty roots in the fact that neural systems are extremely large-scale and complex systems. How can a system of interacting components, i.e., neurons, of relatively limited individual capabilities exhibit powerful and intelligent behavior? Investigation on this fundamental problem may reveal the underlying mechanisms of some emergent features of neural systems. Although accurate mathematical equations, such as
Hodgkin-Huxley equations, have been proposed to demonstrate the spiking responses of single neuron, it is almost impossible to understand the neural systems by coupling these equations for billions of neurons. To overcome this difficulty, we need system-level computational models that are able to construct compact representations of populations of neurons.

Computational models can also help build bridges between different levels of description and to identify unifying concepts and principles in motor control [33]. Due to the grand challenges in this area, human motor control has attracted attention of researchers from a variety of disciplines, ranging from cellular and molecular neurobiology to human psychology. For such reason, there are diverse viewpoints on what is relevant research. For example, neurobiologists search for clues on what cellular mechanism are behind information processing of neurons during movement. In contrast, psychologists are interested in cognitive and higher-level mechanisms underlying the selection, planning, learning, initiation, and execution of movement. Computational models may provide means of communication among different disciplines; with the help of modeling work and theoretical analysis, different disciplines can complement and support each other more effectively. It is therefore not surprising to see that computational modeling has become an important tool to tackle motor control problems.

This thesis presents a related study on human motor control, mainly following an approach of computational modeling. Specifically, this thesis proposes a unified model to describe the role of the basal ganglia in both lower-level movement control and higher-level motor programming and cognitive decision making.

The basal ganglia (BG) are important brain structures. They are a group of highly interconnected nuclei buried within the cerebrum. The BG were historically considered only as major components of the motor system; however, these nuclei have a much broader role than that: they are linked to an extensive range of cognitive functions such as perception [18], learning [56], memory [104], and attention [19]. In addition, dysfunction of the BG has been related to a cluster of human brain disorders, including Parkinson’s disease (PD), Huntington’s disease, and schizophrenia. For such reason, the BG have produced a long time of strong clinical interest, as well as a great
The need for a new neurobiological theory or model of BG function is enormous. Given the wealth of anatomical and physiological data, the BG provide one of the most exciting prospects for computational modeling of brain function. Although the importance of BG in movement and cognitive activities is well recognized, there has been no consensus on the fundamental roles of the basal ganglia. Several computational models were proposed, focusing either on the lower-level motor function of BG (e.g. [34]) or their possible role in decision making (e.g. [13, 49]) or behavioral learning (e.g. [78]). However, the current models failed to offer a unified view for both the motor and non-motor functions of the basal ganglia. And these models lacked elaborate validations of model predictions. Little was done to reproduce normal movement profiles and signs in disorders involving BG dysfunction. Thus a new theory or model of BG function is needed. Not restricted to the field of basic neurobiology, application fields also need a new theory. Developing new medical treatments is increasingly expensive and time consuming, and choosing a wrong strategy based on an inadequate theory wastes resources. All this has urged us to work with tremendous effort to model the function and dysfunction of the basal ganglia.

In the rest of this chapter, we first introduce the BG circuitry and synaptic organization, and then briefly review the current understanding, including several computational models, of the BG function. Finally, we present an outline of the following chapters and highlight the main contribution of this thesis.

1.1 Anatomy of Basal Ganglia

The BG consist of four major nuclei: the striatum, the globus pallidus (GP), the substantia nigra (SN), and the subthalamic nucleus (STN) (Fig. 1-1).

The striatum is the major recipient of inputs to the basal ganglia. Almost all areas of the cerebral cortex send excitatory, glutaminergic projections to the striatum. The striatum also receives excitatory inputs from the thalamus and the midbrain. The striatum is further divided into the caudate nucleus and putamen. The caudate nucleus receives afferent fibers from the prefrontal cortex, and the BG circuit involving
the caudate nucleus is linked to motor planning (the higher level of motor hierarchy) and cognitive functions; the putamen is mostly connected with the motor areas in the cortex, and the putamen circuit is linked to the middle level of motor hierarchy, which is thought to scale the intensity of execution of motor patterns. The striatum contains both projection neurons and several populations of interneurons. The major type of projection neuron is the medium size spiny neuron. They account for more than 90% of the total populations of striatal neurons [202]. The spiny projection neurons use γ-aminobutyric acid (GABA) as their major neurotransmitter, and project inhibitorily to the globus pallidus and substantia nigra. The interneurons, though much fewer in number than the spiny projection neurons, also play an important role in the signal processing of the striatum.
The globus pallidus (GP) is divided into internal and external segments, i.e., globus pallidus pars interna (GPI) and globus pallidus pars externa (GPE). The GPI, together with the substantia nigra pars reticulata (SNr, which will be discussed below), are the output nuclei of the basal ganglia. The GPI is primarily composed of large neurons that project outside of the basal ganglia. About 70% of GPI neurons send collaterals to both thalamus (via thalamus to the cortex) and brainstem, and the other GPI neurons (20%) project to the centromedian-parafascicular complex of the thalamus or to the lateral habenula \[145\]. The GPI neurons are inhibitory and use GABA as a neurotransmitter \[152\]. The GPE may be viewed as an intrinsic nucleus of the basal ganglia. It receives input from the striatum and STN and sends output to STN, GPI and SNr.

The substantia nigra (SN) consists of the pars reticulata (SNr) and pars compacta (SNC). As mentioned previously, the SNr is another BG output nucleus. The SNr is composed of large neurons that receive similar patterns of input as those of the GPI input. Further, like the GPI, the SNr sends output to the midbrain and thalamus, and the output is GABAergic and inhibitory. The primary difference between the SNr and GPI outputs is that the lateral portion of SNr is connected with the areas of the cortex (via thalamus) and of the brainstem (via superior colliculus) that are involved in eye movement control \[123\]. In this thesis, we are more interested in the BG functions that involve the GPI rather than the SNr, so when talking about the output part of the BG, we only mention the GPI part for most cases without even referring to the SNr, but the readers should keep in mind that these two structures have similar circuitry and operate in a similar way.

The SNC is a densely cellularly region containing dopamine cells, which receives GABAergic and inhibitory input from the striatum. The SNC dopamine neurons also project back to the striatum. The dopamine pathway as well as the reciprocal connection between the striatum and SNC are thought to play a crucial role in the learning carried by the basal ganglia.

The subthalamic nucleus (STN) receives an inhibitory, GABAergic input from the GPE, and an excitatory, glutaminergic input from motor areas of cerebral cortex. The output from the STN is excitatory and glutaminergic and projects to GPI, GPE and
SNr. The STN to GPi projection is highly divergent in such sense that each axon from the STN nucleus ensheathes many GPi neurons [146].

1.1.1 Direct and Indirect Pathways

The output part of BG, GPi (here we ignore the SNr part), is thought to be modulated by two parallel pathways (Fig. 1-2): direct pathway and indirect pathway [2, 34, 123]. The direct pathway runs from the striatum to the GPi; the indirect pathway passes to the GPe and from there to the output nucleus GPi, or from the GPe to the STN and finally to the GPi (see Section 2.2.1 for more discussion on the indirect pathway).

It is widely accepted that the direct and indirect pathways have opposing effects on the BG output nucleus. Activation of the direct pathway suppresses the neurons in the GPi, thus allowing the thalamus and eventually the cortex to be activated. Therefore the direct pathway has a positive effect, which disinhibits the thalamus and increases thalamocortical activities. Specifically for motor control, activation of the direct pathway facilitates movement. In contrast, the indirect pathway has a negative effect: The activation of it increases inhibition of the thalamus, thereby decreasing thalamocortical activities. And from the motor control point of view, the activation
of the indirect pathway inhibits movement.

Striatal spiny projection neurons in the two pathways express different dopamine receptors: The striatal neurons of direct pathway mainly express D1 dopamine receptors that facilitate suprathreshold signal transmission, while those from indirect pathway mainly express D2 dopamine receptors that reduce signal transmission \[35\]. Since these two types of receptors function differently when dopamine is present, the direct and indirect pathways are affected differently by the dopaminergic projection from the SNc. However, the dopaminergic inputs to the two pathways lead to the same effect, i.e., reducing inhibition of the thalamocortical neurons and thus facilitating movements or other actions. In addition to their effect on activities of the two pathways, the dopamine D1/D2 receptors contribute greatly to the formation of synaptic plasticity of striatal neurons. This is crucial for the learning capability of the BG, and will be discussed in detail in Chapter 3.

Some movement disorders result from imbalances in the BG direct and indirect pathways. For example, overactivity in the indirect pathway is a major factor in Parkinsonian signs, and underactivity in the same pathway causes hyperkinetic disorders such as "ballism."

### 1.1.2 Multiple Parallel Loops

The BG circuit can be considered as a big loop, starting with inputs from multiple cortical areas to the BG and then returning, via thalamus, to the cortical areas \[138\]. This big loop can be further divided into multiple parallel loops. All these loops are similar in principle, but each uses different cortical areas and a distinctive portion of the striatum and globus pallidus.

Conventionally, the terms of "lower" or "motor loop" and "higher" or "complex loop" are referred to the cortico-BG-thalamocortical loops that correspond to the putamen and caudate nucleus, which receive inputs from motor/somatosensory cortex and association cortex, respectively \[15\]. The lower loop basically participates in the control of movements, while the higher loop contributes roughly to non-motor, cognitive functions. More recently, the motor loop has been considered to be further partitioned into oculomotor and several skeletomotor circuits and the complex
loop is considered to include dorsolateral prefrontal, lateral and medial orbitofrontal paths [2, 120]. Each of these circuit divisions appears to derive input preferentially from the same areas of cerebral cortex to which they project. Thus, the BG loops or channels appear to operate largely as parallel loops with relatively little intercommunication [93]. The sub-architecture of each loop is very similar (Fig. 1-2) including both direct and indirect pathways so that it is widely felt that the BG provide the same or similar computational processing for different regions of the cerebral cortex.

1.2 Basal Ganglia Function and Models

The BG play a crucial role in normal voluntary movement [123] and body posture regulation [39]. However, unlike most other components of the motor system, they do not have direct connections with the spinal cord. As described previously, the BG receive afferent fibers from the cerebral cortex and send efferent fibers to the brain stem and thalamus and, via the thalamus, back to the frontal cortex (FC). The BG motor functions are mainly mediated by motor areas of the FC. The relationship of BG with movement has been strongly suggested by clinical studies on some motor disorders such as PD and dystonia. These motor disorders exhibit disturbance in lower-level movements and postural control, and are related to some dysfunction of the basal ganglia.

Recent advances in knowledge of the BG have led to a hypothesis that the BG do not generate movements. Rather, they act to inhibit competing motor mechanisms that would possibly interfere with the desired movement [123]. In other words, the BG help “filter” out competing, unwanted motor patterns and keep the desired one(s) for the motor system. The BG make the focused selection of an on-going movement or action based on the current brain states including sensory information. As brain states change, the BG may favor an action other than the current one and then switch from the current action to the newly selected action. In such a way, the BG can trigger a sequence of actions.

Besides their role in movement control, the BG are also involved in non-motor aspects of behavior. These include cognitive tasks such as organizing behavioral
responses, using verbal skills in problem solving, mediating empathetic and socially appropriate responses [35], and planning complex motor programs [18, 19, 56]. The non-motor functions of the BG are implemented by circuits that originate in the prefrontal and limbic regions of the cortex and that engage specific areas of the basal ganglia. Damage to any portions of these circuits would be associated with a variety of behavioral abnormalities.

Interestingly, the BG circuits related to non-motor functions share the same internal neuronal architectures with those related to movement control. Therefore, it is not surprising to see that there are some kind of similarities between the basic cognitive functions and motor functions of the BG. For example, it has been proposed that the BG have evolved as a centralized selection device, for both cognitive and motor activities, and this device is specialized to resolve conflicts over access to limited cognitive and motor resources [155].

Although the relationship of the BG with movement and cognitive activities is well recognized and many general conceptions of BG function and operation have been formulated [34, 57, 123], computational models to date have focused on selected aspects of BG function (see reviews in [7, 10, 54]), and have not sought to reconcile the operation of both motor and complex levels, together with pathophysiological features of BG related disorders.

For example, Suri et al. [182] attempted to explain the role of the BG in the control of movement speed, and therefore focused on the dynamic behavior of the BG circuitry in the millisecond range. These investigators pointed out that small differences in timing between the direct and indirect pathways could produce pulse-step output from motor cortex needed to rapidly drive a limb. The bradykinesia, or slowness of movement, one of the characteristics of parkinsonism, can be simulated by reducing the gain of the direct and indirect pathways and increasing the trans-basal ganglionic signal transit time. As these authors acknowledge, however, considerable experimental evidence [193] indicates that indirect pathway function is generally increased in PD. Contreras-Vidal et al. [31] describe the basis of reduced writing size in PD in terms of weakened activation of thalamic (specifically, VL0) neurons. The model also suggests that variability is due to the overactivation of GPi neurons generated by
dopamine depletion in the basal ganglia. The model, however, does not address the bases of rigidity and tremor, the two other cardinal signs of PD (although Contreras-Vidal et al. have suggested that tremor and micrographia are produced by different impaired neural mechanisms). Several other detailed computational models seek to account for the importance of the interaction between globus pallidus and subthalamic nucleus in normal BG function, with particular attention to how its dysfunction may lead to parkinsonian tremor [55, 68, 136, 159]. These models do not address the normal role of the structure in movement control and motor learning.

Cognitive, behavioral level models have described how BG networks may implement decision making [3, 12, 63, 64, 128] and motor sequence generation [13, 17], assist the encoding of sensory contexts in FC [9], or mediate prediction of behavioral reward [77]. Berns and Sejnowski [12, 13] explain how simple competition between direct pathway units can select a “winner-lose-all” unit in the internal segment of the globus pallidus (GPi) that will in turn activate a specific thalamic unit while adjacent thalamic units are suppressed. The winning striatal unit is the one that is first activated by the cerebral cortex. The mechanism is quite effective, but it attributes learning to the indirect pathway and does not describe how multiple thalamic units may be selected or deselected in parallel. Amos [3] has described the operation of BG in Wisconsin Card Sorting task. The clinically interesting model emphasizes BG cognitive functions and accounts for perseverative errors in PD, Huntington’s disease and schizophrenia. However, this model does not consider the role of the indirect pathway or of the BG in movement control. Similarly, the model of Monchi and Taylor [128] describes the possible role of the BG in working memory storage and retrieval. The model shows correlation with fMRI data in PD and schizophrenia. This model proposes a hard-wired neural network model for the cortico-BG interaction in working memory in both health and disorders associated with BG dysfunctions, but also does not address the indirect pathway or motor control derangements in these disorders. Brown, Bullock, and Grossberg [17] have developed a detailed model of BG learning and participation in gating saccadic eye movements across a range of visual fixation tasks. Parts of this model are also consistent with recent data describing frontal circuit performance during sequential hand movements [5]. This model appears to
address BG facilitation of the execution of sequences that are already learned rather than the implicit ("rote") learning of motor sequences. Beiser and Houk [9] show that loops involving FC and BG could support sparse encoding of serial sensory events. In so doing, the BG and FC units display seemingly random switching between on and off states as has been observed in the caudate nucleus [95, 96]. Still, the model does not specifically address how this encoding is acquired or how it may be used to guide executive behavior. Work by O’Reilly, Frank and colleagues [49, 50, 141] has proposed how BG-FC cooperation can gate working memory so that subgoals can be pursued while other goals are maintained against distractors. The authors demonstrate that this could be used to make contingent decisions, and has predicted features of the decision making in patients with PD [50]. This model like the one to be presented here emphasizes the storage of state information in register-like working memory to influence behavior at a later time. Still, the model assumes toggle-like action of the BG on frontal circuits and does not address the steady motor function derangement characteristic of parkinsonism. Each of these models is consistent with the widely accepted view that the BG fundamentally switch the activities of selected cortical circuits. Several models also suggest how this may be acquired through experience. However, in contrast to the lower-level motor control models reviewed above, these models emphasize the more discrete operation of the BG in the programming of behavior. Little detail, if any, is provided regarding lower-level control of movement speed and posture.

Thus, a range of models already appears to give insights into many important aspects of BG and FC operation. However, precisely because of the wealth of models, a more integrated picture would be useful.

1.3 Thesis Contribution and Outline

With the above concern in mind, we propose in this thesis a uniform architecture for various functions of the BG, and try to figure out the universal functionality of the BG under such an architecture.

The modeling presented in the thesis is generally consistent with many current
views and models of BG function, but contributes additional specificity, simplification and coherence. It will be argued that BG circuitry is well suited to implement context-dependent logical gating of frontal circuits whose dynamics are in general (following Mink [123]) also modulated by other systems. This process is felt to underlie BG support of procedural, or “rote” execution and learning, considered here to be the fundamental function of the BG. Such function of the BG helps to free cortical attention for other tasks. When higher-level executive actions are involved, presumably via trans-caudate pathways [35], the procedural mechanism forms the basis of behavioral programming and automated decision making. More complex behavioral sequences depend upon the BG-facilitated formation of higher order state registers in frontal cortex. On the other hand, lower-level motor functions such as the flattening of speed profile can be achieved by the steady read-out of a sequence of finely spaced position targets. This can also be considered a rudimentary form of procedural program execution. Similarly, triggered reactions [37] can be considered single-step rote procedures.

The model assumes that context-to-control mapping is achieved through learned long term changes in connection strengths between corticofugal inputs and striatal neurons. The long term changes are mediated by dopamine as a reward signal as has been frequently considered [50, 77, 168, 196]. Because this Multi-Input Multi-Output Adaptive Switching (MIMOAS) mechanism is not required to generate pulse-step-like output with carefully calibrated amplitudes (e.g. [182]), its movement control is relatively insensitive to added noise. Nevertheless, gross system operation is dependent upon the crispness of BG-FC network switching, and such crispness of switching can be degraded by changes in the BG internal gains. These effects apparently account for several features of the compromised behavioral learning, posture and movement control seen in PD. Thus, the MIMOAS model provides a unified conception of FC and BG function in behavioral and motor control.

This thesis also explores some theoretical issues concerning the computational and learning capabilities as well as dynamics of the basal ganglia. It is proven that the proposed BG architecture ensures the cortico-BG-thalamocortical circuit to perform any transformation that associates cortical contexts with corresponding thalamocor-
tical actions. Under learning rules that are compatible with the known lasting effect of dopamine and synaptic normalization in the striatum, the BG behavior is shown to converge with expected output in response to specific combinations of cortical inputs. The dynamics of the BG network, specifically the striatal network, is investigated with an emphasis on the winner-take-all (WTA) competition among neurons that are reciprocally connected with lateral inhibition. It is demonstrated that the characteristics of both the neural network equilibria and WTA competition among neurons depend upon the steepness of neuronal activation functions and strength of lateral inhibition. The underlying mechanism of winner-take-all is illustrated in terms of augmented separation of supra-threshold activities of competing neurons. It is then suggested that parkinsonian rest tremor and to some extent rigidity can result from decreased striatal selection efficiency, which is a direct consequence of reduced strength of lateral inhibition or decreased activation function steepness of the striatal projection neurons due to the deficiency of dopamine in PD.

The rest of this thesis is organized as follows: First, Chapter 2 presents an overall picture of the MIMOAS model based upon the most widely recognized or strongly suspected features of BG neurocircuitry. The model views the BG function as realizing general mappings between binary cortical context and response vectors. Following Chapter 2, Chapter 3 considers the pattern classification and learning issues of the BG under the framework of the MIMOAS model. This chapter proves the convergence of the BG learning process and the capability of the basal ganglia as universal cortical pattern classifiers. Then Chapter 4 studies the dynamics of striatal networks and the mechanism of winner-take-all competition among striatal projection neurons. Theoretical analysis and results in this chapter also apply to reciprocally inhibitory neural networks of more general forms. In Chapter 5, the role of the BG and BG-FC interaction in procedural learning is investigated. It is shown that the MIMOAS model reproduces the experimentally-observed sensitivity of procedural learning efficiency to sequence complexity and to integrity of BG circuits. Next Chapter 6 studies the role of the BG in lower-level movement control. An example of cruise movement generation under both normal and PD conditions is presented to demonstrate the lower-level motor function of the basal ganglia. Based on the results of simulations
and experiments presented in the previous two chapters, Chapter 7 further discusses the robustness and assertions of the MIMOAS model as well as the FC-BC interactions. The last Chapter is the conclusion of the thesis.
Chapter 2

Multi-input Multi-output Adaptive Switching Model of Basal Ganglia

This chapter proposes a functional neuroanatomic model of the basal ganglia (BG), called the *multi-input multi-output adaptive switching (MIMOAS) model*. The model attempts to provide a unified explanation for the role of the BG in motor control and cognitive decision making. In the model, the BG circuitry effectively implements a large set of parallel noncompetitive logical OR and NOR circuits that can be driven by specific patterns of cerebral cortical activity. These afford selective gating of target frontal thalamocortical neurons. Such logic-like operation of the BG is the model’s core proposal, which is mainly summarized in Sections 2.2.2 and 2.2.4. The BG operation can also be viewed as a general mapping between binary context and response vectors. The mapping is shown to be learnable via a reinforcement process, which will be discussed in greater detail in Chapter 3.

A large number of neuroanatomical pathways have been identified within the basal ganglia [123, 147]. However, the precise function of most of the pathways has not been established definitively. Still, significant consensus regarding certain features of the architecture continues to build. The MIMOAS model is based upon the smallest number of the most widely recognized connections that we have found to be sufficient to account for the operations under investigation. The model does not attempt to account for all recognized connections.

The commonly accepted, or strongly suspected, macroscopic features of BG neurocircuitry include:
• Action of the BG forms a side path that parallels direct connections between cortical context/command areas and executive areas. Executive areas produce output and potentially recurrent activity to context/command areas and to the striatum. For the moment, these areas are intentionally broadly and loosely defined.

• Action of the BG is mediated by net excitatory “direct” and net inhibitory “indirect” pathways (see Section 1.1.1).

• A large ratio of neuron numbers exists between the BG input and output, and between the input to the external segment of globus pallidus (GPe) and its output (see Section 2.2.5).

• The convergence of neuron numbers from BG input to output begins with a convergence of excitatory input from cortex (typically from neurons in lamina 6 [147] throughout the cortex) onto mutually inhibitory striatal networks in striatum (see Fig. 2-1).

• GPe sends inhibitory output to both subthalamic nucleus (STN) and internal segment of globus pallidus (GPI) (see Section 2.2.1).

• A comparatively small number of STN neurons conveys excitation from cortex diffusely to both GPI and GPe (see Section 2.2.5).

• Thalamic neurons typically engage in reciprocal excitatory interaction with neurons in the cortex [118].

• Striatal neurons receive dopaminergic input form the compact cellular portion of the substantia nigra (SNC).

Given these features of anatomy, the MIMOAS model makes ten principal functional assumptions (Assumptions 2.1-2.10). These assumptions are of different levels of importance to model behavior and have various levels of circumstantial support.
2.1 Basal Ganglia Input and Output

2.1.1 Selection and Winner-take-all Competition in Striatum

Assumption 2.1 *The cortical input activity to a local striatal network, usually a striosome/matrix complex, is able to select a unique set of neurons (or a “unit” here the term unit refers to an arbitrary sized group of similarly functioning neurons) within the network.*

The selection is via lateral inhibition and a winner-take-all (WTA) mechanism. The WTA mechanism enables the activity of a local striatal network to become biased toward a particular unit. Most units are usually suppressed unless they began to “win” due to favorable initial weights in relation to a specific input of cortical context (see Fig. 2-1). In this case, collateral inhibitory input to the unit declines to near zero and the unit activity increases quickly, becoming a full “winner” (with output “1”). Hence, striatal output is usually approaches unity or stays near zero. The dynamics of striatal networks have been partially modeled previously and argued to be subject to reciprocal inhibition between striatal projection neurons [197]. Such winner-take-all competition and selection have been further demonstrated in [111] and Chapter 4 based on both theoretical analysis and simulations. It is suggested that the winner-take-all competition among neurons depend upon the steepness of neuronal activation functions and strength of lateral inhibition. Under conditions of high dopaminergic activities, lateral inhibition is strong and neuronal activation function becomes effectively steep. This produces a higher activation threshold for an individual striatal neuron, but enables very rapid secure selection and rejection of most competitor striatal neurons once the threshold is superseded. Under the opposite conditions, individual neurons are more easily activated and competitors may remain active simultaneously for longer periods. The MIMOAS model adopts a version of these dynamics and proposes how particular neurons become selected by a particular pattern of cortical input.

The proposed high specificity selection in the input stage of the BG is consistent with the binary-like neuronal activities in the striatum. The membrane potential of striatal spiny projection neurons fluctuates between two subthreshold levels, the
Figure 2-1: Winner-take-all competition in the corticostriatal network showing the selection of unique striatal units by two different patterns of cortical input represented by binary vectors at right. In (a) a direct pathway neuron is activated, in (b) an indirect pathway unit. Symbols are explained in the text following (2.6).
depolarized “up” state and the hyperpolarized “down” state. These two states are separated by 15-30 mV, and the mean potential of the up state is usually 3-5 mV below spike threshold [99]. Spiny projection neurons fire action potential almost only during the up state. Therefore, the up states have been perceived as temporal gates invoked by excitatory input, during which the spiny projection neurons are enabled to translate afferent cortical activities into patterns of action potentials and thus allow information processing through cortico-BG circuits [90, 139, 180].

### 2.1.2 Reverberatory Thalamocortical Interaction and Cerebellar Coprocessing

**Assumption 2.2** The cortical target of thalamic neurons is assumed to be a “patch” [17] or “stripe” [49] of layer 5 cortical output neurons. These output neurons are accessed via focal inputs to layers 4 and 5, and perhaps more diffusely through layers 1 and 3 [118]. At least a number of frontal cortical (FC) areas including primary motor (M1) area, supplementary motor area (SMA) and area 46, frequently send excitatory projections to the thalamic nuclei from which they receive excitation [118].

The above assumption makes possible reverberatory self-excitatory thalamocortical interactions. Such interactions have been highlighted by Houk and colleagues and have been assumed by several models [9, 78]. This mechanism has been questioned by some [49] who favor a toggling action of the thalamus on bistable cortical neurons or circuits. A concern regarding the sustained reverberation model is that because of the relatively small number of thalamic neurons in relation to cortical neurons, the reverberation would tend to involve an entire patch or stripe and therefore lack requisite topographic focality. Nonetheless, the MIMOAS model assumes reverberatory interaction for several reasons. In addition to the data and arguments presented by Houk et al. [9, 78], it has been shown that lesions of thalamic nucleus VPLo (the oral part of the ventral posterior lateral nucleus) that connects reciprocally to M1 [127] result in profound paralysis [191], while lesions of the cerebellar input to VPLo do not paralyze. So at least in terms of movement generation, it is arguable that ongoing reverberation may be of great importance. Second, we consider that there are a number of intracortical inhibitory neuronal networks [53, 91]. The MIMOAS model
assumes that it is likely that important topographic focus is provided by intracortical projections to these inhibitory networks in the cortical area being gated as a whole by the BG-thalamic mechanism (e.g. [118]) (Figs. 2-2 and 2-3). The model also considers that target cortical modules are functional units that might be topographically co-mingled.

In addition to Assumption 2.2, the MIMOAS model makes the following assumption concerning a specific function of thalamocortical modules.

**Assumption 2.3** Some thalamocortical modules can serve as working memory in FC [6], or free registers (in the computer architecture sense) for the binary representation of interim states if needed.

As shown in Chapter 5, the interaction between the BG and thalamocortical modules that serve as working memory plays an essential role in procedural learning and generation of sequences of actions.

**Assumption 2.4** The reverberatory connections between cerebral cortex and cerebellum may provide signal processing that can sharpen and perhaps scale signals in the cortex.

Houk and Wise [76, 78] have highlighted the presence of reverberatory connections between cerebral cortex and cerebellum. The MIMOAS model assumes that such
Figure 2-3: Thalamocortical modules: their relationship to BG and cortical inputs and interchangeable schematic representations emphasizing net integrator-like behavior. “s” is Laplace transform frequency variable. When BG output from GPi is 0, reciprocal excitation is enabled. When GPi output is 1 the cortical neuron is maintained well below threshold and therefore relatively unresponsive to cortical inputs. The cerebellum is argued to provide dynamic feedback support. CB(\omega) is given by (2.1). a: When the neuronal activation function is less saturated, the thalamocortical module functions more like an integrator. b: When the neuronal activation function is more saturated, the module functions more like a bi-stable register. In this case, we use brackets on the schematic representations of the thalamocortical module.
pathways may provide coprocessing that can sharpen and perhaps scale signals in the cortex (see [115] for details). As discussed in Chapter 5, this action may be important for FCBG interaction. In the MIMOAS model, cerebellar action is represented by a linear filter with frequency response:

\[
\text{CB}(j\omega) = \left( \beta_1 - \frac{\beta_2}{0.01j\omega + 1} \right) \frac{\beta_3 e^{-0.01j\omega}}{0.01j\omega + 1}.
\] (2.1)

This is very similar to the proposals of [86, 114] that are based on a simple linear interpretation of Ito's functional microcomplexes [83]. Here, the dentate and inhibitory Purkinje neurons are represented as leaky integrators, each having a leak time constant of 0.01 sec. The direct, trans-dentate path is assigned a weight of $\beta_1$ and the trans-Purkinje cell path a weight of $\beta_2$, $\beta_3 = \pm 1$ to indicate a positive or negative net connection. The cerebellar influence is considered to operate with a 10 ms round trip delay. A simulated effect of cerebellar recurrent facilitation on the responsiveness of a thalamocortical model can be found in the appendix of [115].

### 2.1.3 Permissive Role of Basal Ganglia Output

**Assumption 2.5** The BG thalamic neurons act to enable or disable cortical neurons that are being driven by other inputs, rather than to drive cortical neurons directly.

This assumption follows Mink [123] and indicates the binary nature of the BG output: The BG implement facilitation ("1") of a desired action and suppression or inhibition ("0") of an un-wanted action. The assumption is considered realistic because activity in many [187] (but not all, see [66]) BG output units show poor correlation with body kinematics, dynamics, or EMG [16]. This is in contrast to those of many units in the cerebellar system and motor cortex. Rather, BG output unit activity relates more closely to cues and state of motion [16], and is less finely sculpted than in other areas. Moreover, lesions of BG output do not result in paralysis. These observations are most consistent with the logical gating view presented in Section 2.2. On the other hand, it may also be suspected that in certain disease states such as dystonia and chorea, BG thalamic input to cortex may be sufficiently large to activate cortical neurons directly or to allow them to become active with
smaller and hence potentially spurious inputs. This results in involuntary output that is characteristically crude and poorly formed in comparison with normal. The above argument is consistent with the possibility that the BG are in general not concerned with detailed shaping of motor commands (contrast with [182]).

2.2 Logic-like Operation of Basal Ganglia

2.2.1 Indirect Networks and Role of STN

Recent anatomical studies have argued the existence of the “classical” indirect pathway. The GPe-STN projection is considered as a component of the classical indirect pathway. However, it is reported [148] that the GPe projects principally to the rostralateral region of STN (which contains neurons that project back to the GPe), not to the caudomedial region of STN (which contains neurons that project to the GPi or SNr, i.e. substantia nigra pars reticulata). And it is argued that there is no sufficient evidence for a direct link between the GPe-STN fibers and the STN-GPi cells [148]. (Note that in the following, when referring to the output part of the BG, we only consider the GPi and ignore the SNr.) For such reason, the continuity of the classical indirect pathway seems to be questioned due to the above anatomical findings. On the other hand, however, a more recent tracing study [172] directed findings in the squirrel monkey supporting the continuity of the GPe-STN-GPi projection of the classical indirect pathway. It is found that groups of neurons of the GPe are reciprocally connected to groups of neurons in the STN which projects to groups of functionally related neurons in the GPi [172].

The GPe may act on several neuronal systems other than those forming the classical indirect pathway. For example, experiments [147, 148] have shown that the GPe projects massive fibers to GPi. These findings place the GPe in a crucial position to control directly the output components of the basal ganglia. Since the GPe exerts an inhibitory control upon GPi, the GPe-GPi link plays a role similar to the GPe-STN-GPi link in the classical indirect pathway. Furthermore, the GPe neurons and their reciprocally connected STN neurons project to common groups of neurons in the BG output nuclei [172]. Due to the existence of multiple connections among the
GPe, STN, and GPi, Smith et al. have proposed a concept of “indirect network” to enrich the classical indirect pathway by considering the highly interconnected neural system in GPe, STN, and GPi [177].

Assumption 2.6 The GPe-GPi connection exerts stronger inhibitory effect than the GPe-STN-GPi connection. Therefore, this thesis considers the GPe-GPi connection rather than GPe-STN-GPi connection as the primary connection linking GPe and GPi in the indirect pathway. In addition, this thesis considers the STN as another input station of the BG rather than only a component of the indirect pathway.

Although both GPe-GPi connection and GPe-STN-GPi connection have similar effect of inhibiting the BG output nuclei, the former connection exerts stronger inhibitory effect. This is because the GPi neurons receive a high proportion of their inputs directly from the GPe, and those GPe inputs tend to synapse more on the sites of proximal dendrites and cell bodies than on the sites of distal dendrites of the GPi neurons. In comparison, terminals from the STN on GPi are relatively evenly distributed.

Moreover, the inhibitory efficacy of GPe on STN in the GPe-STN-GPi pathway has also been questioned recently. So far, there is no consensus on the role of GPe in the STN overactivity in Parkinson’s disease (PD). The hypokinesia in PD has been interpreted as a consequence of the GPi overactivity, which can be explained as the result of a reduced inhibitory input from the striatum and an increased excitatory input from the STN. The STN overactivity was thought to be due to the reduced inhibitory control exerted by GPe. However, it has been reported that there is no noticeable GPe hypoactivity in MPTP-intoxicated monkeys or in PD patients [105]. Further, removal of pallidal input to STN caused only a slight increase in firing rate (19.5%), compared to an increase of 105% followed by a lesion of SNc [71]. The overactivity of STN may be related to impairment of the dopaminergic projection to the STN. In MPTP-intoxicated macaque monkeys, dopamine levels in the STN fall to 13% of normal [25, 153]. Some studies have proposed that dopamine denervation facilitates STN output [24, 70, 108]. However, other studies including more recent results have pointed to a functional innervation of the STN by dopamine [103, 126,
Therefore, the increased firing of STN in PD may result in part from complex interactions between glutamatergic hyperactive fibers originating in the thalamus and the pedunculopontine nucleus and hypoactive fibers originating in cerebral cortex [142].

In addition, this thesis shares the view of the studies [105, 123, 133] and considers the STN as another input station of the basal ganglia. The STN receives direct cortical projections from wide areas of the frontal lobe [194]. Recent studies have shown that monkey STN receives substantial somatotopically organized projections from the primary motor cortex, supplementary motor area, and dorsal and ventral divisions of premotor cortex [132, 133]. Further, cortical stimulation evokes strong and short-latency excitatory responses in STN neurons, and STN then conveys powerful excitatory effects from cortical areas to the GP, with shorter conduction time than the cortico-striato-GP pathway [98, 133]. Therefore, the STN shares some characteristics of the striatum as an input component of the basal ganglia.

### 2.2.2 Internal Switching

The BG internal circuitry favors a logic operation of binary information representations by neurons. The majority of the BG cells are GABAergic; these cells provide inhibitory projections. Both experimental and simulation results have shown that inhibitory postsynaptic potentials can play a functional role in realizing synchronization of neuronal firing [106], which is essential for binary ("0/1") information processing in the basal ganglia. Furthermore, the MIMOAS model makes the following assumption.

**Assumption 2.7** *In BG and thalamic units, individual inhibitory inputs are ordinarily sufficiently powerful to dominate when both excitatory and inhibitory inputs are active.*

Thus, under normal physiological conditions, the output $Y_{\text{unit}}$, as indicated by firing probability or average firing rate, of the globus pallidus (GP) unit depicted in Fig. 2-4 is more usefully described by

$$Y_{\text{unit}} = \overline{A} \land \overline{B} \land (C \lor D)$$

(2.2)
Figure 2-4: Logic-like processing by principal internal connections of the basal ganglia. a: Assumed elemental computation by each unit in globus pallidus (GP). b: Connections of the $k$-th BG channel and target corticothalamic module. For simplicity, the inhibitory connection between GPe and STN is omitted. Parallel FC-BG channels controlled by a single input context. c: Parallel FC-BG channels controlled by a single input context. Multiple channels are supported by single STN input $Z$. See Section 2.2.4 for details.
than by $y_{\text{unit}} = -a - b + c + d$, where $A, B, C,$ and $D$ are binary signals taking values on $\{0, 1\}$ and $y_{\text{unit}}, a, b, c,$ and $d$ are continuously valued signals not constrained to zero or unity. In (2.2), overbar, $\wedge$, and $\vee$ represent respectively complementation, “and,” and “or.” This assertion is central to the MIMOAS model when it is not modeling disease.

Equation (2.2) emphasizes the model’s assertion that the BG normally operate in a strongly nonlinear, switching fashion. This is consistent with the ideas of [9, 13, 63, 64], but stands in contrast to several other proposals such as [182]. The actual implementation of individual neuronal output used in the model for greater physiological realism is given by (2.3) (see Fig. 2-5):

$$y_{\text{unit}} = f_{\omega c}(\lambda_- (-a - b) + \lambda_+(c + d)), \quad (2.3)$$

where $f_{\omega c}$ is a combined low pass filter and input-output activation function/operator with input scaling constants $\lambda_- > \lambda_+ > 0$ and low pass filter characteristics with corner frequency $\omega_c$. For the globus pallidus, the neuronal activation function may be approximated by a saturation function, $\text{sat}_\epsilon(\cdot)_0^2$, which has an effective input threshold $\epsilon$ and saturation at $\gamma$ (see Fig. 2-5). The corner frequency of the low pass filter of $f_{\omega c}$ is about 100 rad/s. It turns out that under these circumstances, for signals $a, b, c, \text{ and } d$ switching at frequencies up to 10 Hz, (2.3) can be considered functionally normalized to

$$y_{\text{unit}} = [-\lambda a - \lambda b + c + d]_0^1, \quad (2.4)$$

where $[x]_0^1$ denotes $\min\{\max\{x, 0\}, 1\}$ and $\lambda = \lambda_-/\lambda_+ > 1$. For the simulations shown in Chapters 5 and 6,

$$\text{sat}_\epsilon(x)_0^2 = \left[\frac{\gamma}{1 + e^{-\epsilon-(\epsilon)}} - \frac{\gamma}{2}\right]_+$$

with $\gamma = 3, \epsilon = 0$, and $[x]_+ = \max\{x, 0\}$, and $\lambda_-$ and $\lambda_+$ are set to 5 and 3, respectively, for (2.3). As a result, most neurons in the globus pallidus behave according to (2.2).

The GPe and GPi units (Fig. 2-4) have single excitatory inputs from STN with value $Z$. Because $\lambda > 1$, any inhibitory input from striatum shuts off these units. If all striatal units are inactive, then GPe is active with value $Z$ and GPi is off
irrespective of the value of $Z$. The BG output is potentially graded (non binary) only if striatal input to GPe is active and there is no inhibition of GPi via the direct pathway. This occurs if $0 < Z < 1$. Under such circumstances, one might expect a failure of BG suppression of thalamic transmission. The practical significance of this case is not clear and for the moment is not addressed by the model. Rather, the model asserts that under most conditions the BG operation can be described in terms of logical switching. This does not necessarily apply to circumstances in which the BG operation is abnormal. In this case, internal signals may become weaker or sluggish in transitions that degrade system operation.

### 2.2.3 Parallel Modular Architecture

**Assumption 2.8** The BG are organized as hundreds of thousands of independently operating, non-competing, parallel channels from striatum to cortex. Specifically, the input to each channel is a striatal striosome/matrix complex and the output is to a relatively small number of thalamic neurons in “specific” nuclei [140] that share a common cortical target. And within each channel, striatal neurons project either to GPi or to GPe, but not to both (Fig. 2-4b and c).

This assumption is similar to those of [63, 64] where Gurney et al. assume that the striatal neurons projecting to GPi and GPe are from different cell population, but
contrasts with those of [9, 13] where Beiser and Houk did not consider the indirect pathway while Berns and Sejnowski assume that individual striatal projection neuron projects to both the GPi and GPe. At present, the fine structure of BG architecture is not sufficiently well determined to evaluate these conjectures definitively. However, cell population estimates are consistent with this architecture [140, 202]. Also, it is recognized that matrix neurons that project to the direct pathway contain a relative preponderance of D1 dopamine receptors over D2 dopamine receptors, while those that project to indirect pathway contain the reverse [177]. Presumably therefore, striatal neurons leading to direct and indirect pathways are distinct. This assumption is the most critical to the operation of MIMOAS model.

2.2.4 Effectively Binary Context Input and Net Logic-like Operation

Assumption 2.9 From the perspective of BG function, behavioral states or cerebral cortical contexts that serve as inputs \( (C) \) can be viewed as being represented by a collection of \( m \) cortical units \( C_i, i = 1, \ldots, m, \) of which some are comparatively active, and the others are much less active for some nontrivial time period. The BG input could then be represented as an \( m \)-dimensional binary vector \( C = [C_1, \ldots, C_m]^T, C_i \in \{0, 1\}. \)

This formalism has been used profitably by Beiser and Houk [9]. The winner-take-all mechanism in striatum responds most efficiently to such inputs. If cortical input vector signals are too similar in amplitude, winners may be selected slowly or spurious winners may be chosen. It may be anticipated therefore, that especially where speed of BG processing is important, processing would be facilitated by sharp transitions between widely separated levels of input activity. Experimental evidence suggests that BG processing may take time on the order of 100 ms [13, 199], and this is consistent with the internal cumulative phase lags in the MIMOAS model circuitry. Therefore, the model proposes that the BG are ideally suited to process strong cortical switching processes that occur on the order of 10 Hz (i.e. alpha range) or slower, and to demonstrate substantial insensitivity to more subtle or higher frequency transients and noise.
Under the preceding assumptions, the operation of the $k$-th BG channel can be viewed as a binary valued mapping $\text{BG}^k(\cdot)$ from an arbitrary number of $m$-dimensional context vectors to one of $q$ possible GPi output targets $X^k$. That is,

$$X^k = \text{BG}^k(C) \text{ or } X^k = \text{BG}^k(C),$$

where $C \in \{0,1\}^m$, $X^k \in \{0,1\}$, and $k = 1,...,q$. Noted that here $\text{BG}^k(\cdot)$ does not represent the GPi output of the $k$-th BG channel, but rather the complementation of the GPi output—under such convention, $\text{BG}^k(C) = 1$ implies that the cortical context $C$ is associated with an operation of facilitation of the $k$-th BG-thalamocortical channel.

The thalamocortical output targets $Y^k$ can then be expressed as

$$Y^k = U^k \land \text{BG}^k(C),$$

where $Y^k \in \{0,1\}$, $k = 1,...,q$. As depicted in Fig. 2-4b, the net input-output mapping can be expressed more specifically as

$$Y^k = U^k \land (S^k_{D_1} \land S^k_{D_2} \land \cdots \land (S^k_{I_1} \land S^k_{I_2} \land \cdots \land Z) \land Z)$$

$$= U^k \land (S^k_{D_1} \lor S^k_{D_2} \lor \cdots \lor (S^k_{I_1} \lor S^k_{I_2} \lor \cdots \lor Z) \lor Z),$$

where the second expression follows from two applications of De Morgan’s law [82]. Here, $U^k$ represents the intended cortical output of the $k$-th channel, and $Z$ the activation of the STN. The influence of the $j$-th context vector, denoted $C^{(j)}$ (numbered arbitrarily) on the $k$-th BG channel is represented by $S^k_{D_j}$ or $S^k_{I_j}$ for some positive integer $D_j$ or $I_j$, respectively. If $S^k_{D_j} = 1$, the context influences via the direct pathway, and if $S^k_{I_j} = 1$, the context influences via the indirect pathway. Importantly, the proposed learning mechanism (see Chapter 3) ensures that for each $k$, i.e. within any channel, the activation of direct and indirect pathways by a given context is disjoint.

Consider a more general form of (2.7):

$$y^k = u^k \times (S^k_{D_1} \lor S^k_{D_2} \lor \cdots \lor (S^k_{I_1} \lor S^k_{I_2} \lor \cdots \lor Z) \lor Z)$$

$$= \left\{ \begin{array}{ll} u^k \times (S^k_{D_1} \lor S^k_{D_2} \lor \cdots \lor (S^k_{I_1} \lor S^k_{I_2} \lor \cdots)) & \text{for } Z = 1, \\ u^k & \text{for } Z = 0. \end{array} \right.$$
In (2.8), $y^k$ is used instead of $Y^k$ to include the case, as in motor cortex, where the intended cortical output is a continuously-valued signal $u^k$, rather than a binary valued $U^k$. The symbol $\times$ represents product, and $y^k$ takes continuous value rather than binary value. In other cases where the intended output is essentially binary, as with thalamocortical registers (see Chapter 5), the expression can be in fully logical form as in (2.7). Equation (2.8) indicates that the BG can in principle be activated or inactivated according to input $Z$, the activation of the STN. Whether or not this global switch is used physiologically is not clear. However, it conceivably corresponds to the notion of allowing rote mechanisms to take over control or not. Assuming that $Z = 1$, the equation states that each BG channel provides focused inhibition for any nonzero $S^k_{I_j}$, that can be overridden by any nonzero $S^k_{D_j}$. Alternatively, each channel provides focused enabling that can be withdrawn by zeroing all $S^k_{D_j}$ and instituting any $S^k_{I_j}$. The possible relevance of this formal asymmetry is discussed later. As a whole, control of the BG can be considered to implement $q$ independent, parallel mappings from $m$-dimensional binary context vectors to each element within a $q$-dimensional potentially binary output vector of thalamocortical module activities $Y = [y^1, ..., y^q]^T$.

Chapter 3 will further show that the BG operation in the MIMOAS model can realize universal logic mappings.

### 2.2.5 Volume Convergence in Basal Ganglia

Anatomical findings have shown a significant decrease of neuronal tissue volume from the cerebral cortex to the deepest portions of the basal ganglia. It was estimated in human BG that the volume of striatum was 12 times larger than that of the GPe, 20 times larger than that of GPi and SNr, and 60 times larger than that of the STN [202]. In rat basal ganglia, it was found that the right BG consists of approximately 2,790,000 striatal neurons, 46,000 GPe neurons, 29,500 entopeduncular (GPi) and SNr neurons, and 13,600 STN neurons [140]. In other words, the rat striatal neuron number is about 60 times more than that of the GPe, 90 times more than that of GPi and SNr, and 200 times more than that of the STN.

The ratios among the unit numbers of various BG nuclei in the MIMOAS model
qualitatively fit the anatomical data. As will be shown in Chapter 3, in order to realize an $m$-input $q$-output logic function that is compatible with $p$ input-output pairs (i.e., $p$ cortical contexts and their associated BG actions), the implementing circuit of the BG needs $p \cdot q$ striatal units, $q$ GPe units, $q$ GPI units, and 1 STN unit. Therefore, the striatal unit number is about $p$ times more than that of the GPe or GPI, and $p \cdot q$ times more than that of the STN. This is consistent with the convergence of neuron numbers from striatum to GPe/GPI and from GPe/GPI to STN.

2.3 Adaptation Mechanism

The BG also play a critical role in the acquisition of rote behavior [57, 151]. Considerable evidence has related adaptation of the BG to the presence of increased or decreased dopamine as an internal indicator of behavioral reward [167] or punishment [189], respectively. Given the large supply of dopaminergic pathways to the striatum [168], it is natural to consider that the most important plasticity occurs there. A significant number of dopaminergic fibers also travel to STN, so that plasticity there is conceivable as well [108].

Assumption 2.10 Each striosome-matrix complex is supplied by three particular inputs (see Fig. 3-2 in Chapter 3): (i) diffuse, uniform and strong input to the matrix neurons that represents a copy of the actual activity (may be considered as “teaching signals”) of the cortical unit to be controlled by the basal ganglia; (ii) diffuse and necessarily topographically nonuniform input from the cortex (described above) that represents the current behavioral context; (iii) diffuse, uniform dopamine input via the striosomal neurons.

As described above, these input sources are consistent with known neuroanatomical connections. Using these inputs, the BG learn to associate context vectors with on/off control of FC circuits. The learning mechanism is detailed in Chapter 3.
Chapter 3

Pattern Classification and Competitive Learning of Basal Ganglia

Following Chapter 2, this chapter explores some theoretical issues concerning the computational and learning capabilities of the basal ganglia (BG) under the framework of the MIMOAS model. Specifically, this chapter studies the pattern classification and competitive learning of the striatal network in the basal ganglia. The striatal networks can be modeled after simplification as single layer winner-take-all networks with converging excitatory inputs from cortical neurons. It is proven that appropriate setting of the network parameters enables the striatal networks to possess pattern classification capability. Embedded with neural networks of such architecture, the BG are able to perform any transformation for binary vectors. Furthermore, under simple rules mimicking the long term potentiation (LTP) and long term depression (LTD) of synaptic plasticity and synaptic normalization in the striatum during learning, the network behavior is proven to converge, so that the striatal neurons learn to fire with expected patterns in response to specific combinations of cortical inputs.

3.1 Simplified Equations for Basal Ganglia Operations

We assume that each cortical neuron that projects to the striatum initially synapses on each striatal projection neuron. As mentioned previously, the striatum, the input
part of the BG, receives projections from almost all areas of the cerebral cortex. At the single cell level, it is suggested that any striatal projection neuron receives convergent inputs from multiple cortical neurons [147]. Meanwhile, a cortical column extensively innervating widely-distributed, discrete striatal regions, and it is further suggested that single cortical neuron contacts several striatal neurons [147, 170]. Thus, there are both a “one-to-many” and a “many-to-one” patterns of connectivity between the cortex and striatum. Although in a mature cortico-striatum system (which has already experienced considerable learning during development), the cortical and striatal projection neurons are far from fully interwoven with each other [171], we follow Plenz and Kitai [154] and assume that initially every corticostriatal projection neuron contacts every striatal projection neuron.

Let $S_1, S_2, ..., S_n$ be a group of striatal projection neurons in some local area of the striatum. Each striatal neuron receives inputs from $m$ cortical neurons, $C_1, C_2, ..., C_m$. Without causing confusion, we use the same symbols $C_i$ and $S_j$ to represent the outputs or firing rates of the neurons $C_i$ and $S_j$, $i = 1, ..., m$, $j = 1, ..., n$, respectively. The synaptic strength or weight of the corticostriatal connection from $C_i$ to $S_j$ is denoted $w_{ij}$. A cortical signal at the input of a synapse connected to a striatal neuron is then multiplied by the synaptic weight. Generally, the synaptic weight may lie in a range that includes positive as well as negative values, indicating an excitatory synapse or inhibitory synapse, respectively. For the excitatory nature of corticostrial connections, $w_{ij}$ takes non-negative values. Let $d_j$ be the sum of cortical inputs to the $j$-th striatal projection neuron:

$$d_j = \sum_{i=1}^{m} w_{ij} C_i, \quad j = 1, ..., n. \quad (3.1)$$

In vector form, denote $C = [C_1, ..., C_m]^T$ and $w_j = [w_{1j}, ..., w_{mj}]^T$. Then (3.1) can be written as

$$d_j = w_j^T C.$$

So as to simplify the analysis of computational capability of the striatal networks and the basal ganglia, we only consider the network steady-state responses in this chapter, and further assume that the corticostrial network realizes a winner-take-
all gate computing a multi-input multi-output function $\text{WTA}_n$: \[
[S_1, \ldots, S_n] = \text{WTA}_n(d_1, \ldots, d_n) = \text{WTA}_n(w_1^TC, \ldots, w_n^TC),
\] (3.2)
which satisfies
\[
S_j = \begin{cases} 
1 & \text{if } d_j > d_k \text{ for all } k \neq j, \\
0 & \text{if } d_j < d_k \text{ for some } k \neq j.
\end{cases}
\] (3.3)
That is, in the case of inputs that lead to pairwise different $d_1, \ldots, d_n$, only output $S_j$ has value 1, which marks the position of the largest $d_j = w_j^TC$. In the above definition of function $\text{WTA}_n$, we ignored the zero measure set of inputs with $d_j = d_k$ for some $j \neq k$.

Let $D_S$ be the set of index $j$ with which the striatal neuron $S_j$ projects to the direct pathway (GPI), and $I_S$ the set of $j$ with which $S_j$ projects to the indirect pathway (GPe). According to Assumption 2.8, a striatal neuron projects either to GPI or to GPe, but not to both, so we have $D_S \cap I_S = \emptyset$ and $D_S \cup I_S = \{1, \ldots, n\}$.

According to Chapter 2 and the above definitions, the modeled binary input-output operation of a BG channel is equivalent to
\[
\text{BG}(C) = \frac{1}{2} \left(1 + \sum_{j \in D_S} S_j - \sum_{j \in I_S} S_j \right),
\] (3.4)
which can also be realized after simplification by a three-layer network with a winner-take-all (WTA) network as the hidden layer (see Fig. 3-1). It should be noted that
Table 3.1: Main notation used in this chapter.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG(C)</td>
<td>Complementation of GPi output when cortical input is C</td>
</tr>
<tr>
<td>C</td>
<td>[C₁, ..., Cₘ]ᵀ</td>
</tr>
<tr>
<td>Cᵢ</td>
<td>i-th cortical neuron or the output (firing rate) of Cᵢ</td>
</tr>
<tr>
<td>C(j)</td>
<td>j-th cortical context vector</td>
</tr>
<tr>
<td>D_c</td>
<td>{j</td>
</tr>
<tr>
<td>D_j</td>
<td>\arg \max_{k={1,...,n}} w_kᵀC(j) for j ∈ D_c, and 0 for j ∉ D_c</td>
</tr>
<tr>
<td>d_j</td>
<td>Sum of cortical inputs to the j-th striatal projection neuron</td>
</tr>
<tr>
<td>D_s</td>
<td>Set of index j with which the striatal neuron S_j projects to the direct pathway</td>
</tr>
<tr>
<td>I_c</td>
<td>{j</td>
</tr>
<tr>
<td>I_j</td>
<td>\arg \max_{k={1,...,n}} w_kᵀC(j) for j ∈ I_c, and 0 for j ∉ I_c</td>
</tr>
<tr>
<td>I_s</td>
<td>Set of index j with which the striatal neuron S_j projects to the indirect pathway</td>
</tr>
<tr>
<td>L(i)</td>
<td>{i</td>
</tr>
<tr>
<td>m</td>
<td>Number of cortical neurons under consideration</td>
</tr>
<tr>
<td>n</td>
<td>Number of striatal neurons under consideration</td>
</tr>
<tr>
<td>p</td>
<td>Number of cortical contexts (or patterns)</td>
</tr>
<tr>
<td>S_j</td>
<td>j-th striatal projection neuron or the output (firing rate) of S_j</td>
</tr>
<tr>
<td>w_ij</td>
<td>Weight of connection from C_i to S_j</td>
</tr>
<tr>
<td>w_j</td>
<td>[w₁j, ..., wₘj]ᵀ</td>
</tr>
<tr>
<td>δ(j)</td>
<td>Expected value of BG(C(j))</td>
</tr>
</tbody>
</table>

here BG(·) does not represent the output of GPi (i.e. \( \mathbf{X} \), as indicated in Fig. 2-4), but rather the complementation of the GPi output (i.e. \( \mathbf{X} \))—under such convention, BG(C) = 1 implies that the cortical context C is associated with an operation of facilitation of the considered BG-thalamo-cortical channel.

We further introduce the following notation. Given a cortical context C(j), let \( δ(j) \) be a binary assignment associated with C(j): \( δ(j) \) equals 1 if C(j) is associated with an operation of facilitation of the considered BG-thalamo-cortical channel; in other words, \( δ(j) \) equals 1 if the expected BG(C(j)) equals 1. Similarly, \( δ(j) \) equals 0 if the expected BG(C(j)) equals 0, i.e., C(j) is associated with an inhibitory operation of the considered BG-thalamo-cortical channel. Let \( D_c \) be the set of index j with which...
\[ \delta^{(j)} \text{ equals } 1, \text{i.e.,} \]
\[ D_C = \{ j \mid \delta^{(j)} = 1, \ 1 \leq j \leq p \}, \quad (3.5) \]

where \( p \) is the total number of cortical contexts (or patterns) under consideration. And define \( I_C \) as
\[ I_C = \{ j \mid \delta^{(j)} = 0, \ 1 \leq j \leq p \}. \quad (3.6) \]

For each \( j \in D_C \), denote \( D_j \) the index of striatal neuron that wins when the input cortical context is \( C^{(j)} \), i.e.,
\[ D_j = \arg \max_{k \in \{1, \ldots, n\}} w_k^T C^{(j)} \quad \text{for } j \in D_C. \]

Set \( D_j = 0 \) for \( j \not\in D_C \). Similarly, denote \( I_j \) the index of striatal neuron that wins when \( C = C^{(j)} \), i.e.,
\[ I_j = \arg \max_{k \in \{1, \ldots, n\}} w_k^T C^{(j)} \quad \text{for } j \in I_C. \]

And set \( I_j = 0 \) for \( j \not\in I_C \). According to the definition of \( D_S \) and \( I_S \), we have that \( D_j \in D_S \) and \( I_j \in I_S \) for any nonzero \( D_j \) and \( I_j \). Note that even for nonzero \( D_k \) and \( D_j \) (or \( I_k \) and \( I_j \)) where \( k \neq j \), \( D_k \) may equal \( D_j \) (or \( I_k \) may equal \( I_j \)), which implies that the same striatal neuron wins when the cortical context is either \( C^{(j)} \) or \( C^{(k)} \). Denote \( L^{(j)} \) the set of location index of an element in context vector \( C^{(j)} \) that equals 1, i.e.,
\[ L^{(j)} = \{ i \mid C^{(j)}_i = 1, \ 1 \leq i \leq m \}. \]

The above notation is also summarized in Table 3.1.

### 3.2 Basal Ganglia as Universal Pattern Classifier

In this section, we study the computational capability of the basal ganglia as a cortical pattern classifier. As discussed in the previous section, the operation of a BG channel equivalently realizes a binary function \( BG : \{0, 1\}^m \rightarrow \{0, 1\} \), mapping from the state of \( m \) cortical neurons to a thalamo-cortical action (either facilitation or inhibition). For the binary nature of its output, the function \( BG(\cdot) \) can also be considered as a cortical pattern classifier. The pattern classifier \( BG(\cdot) \) is specified not only by the BG
architecture but also by the tunable weights of the corticostriatal connections, \(w_{ij}\), \(i = 1, \ldots, m\), \(j = 1, \ldots, n\). In this section, we are specifically interested in the following problem: Given a set of input vectors, say \(C(1), \ldots, C(p)\), each of which belongs to \(\{0,1\}^m\), and for any possible binary assignments \(\delta(1), \ldots, \delta(p)\), each of which belongs to \(\{0,1\}\), does there always exist a \(BG(\cdot)\) with appropriately chosen weights \(w_{ij}\) such that \(BG(C(j)) = \delta(j)\) for any \(j = 1, \ldots, p\)? Or in other words, is \(BG(\cdot)\) a universal pattern classifier? This problem will be addressed by Propositions 3.1-3.3.

**Proposition 3.1** Let \(C(1), \ldots, C(p)\) be \(p\) pairwise different nonzero vectors and \(\delta(1), \ldots, \delta(p)\) be any binary assignments: \(C(j) \in \{0,1\}^m\) and \(\delta(j) \in \{0,1\}\) for \(j = 1, \ldots, p\). If for any \(j\), \(L(j) \not= L(k)\) for any \(k \not= j\) where \(j, k \in \{1, \ldots, p\}\), then there exists a \(BG\) channel with \(p\) striatal neurons and \(m \times p\) nonnegative weights so that the input-output mapping of this \(BG\) channel satisfies \(BG(C(j)) = \delta(j)\) for any \(j = 1, \ldots, p\).

**Proof:** We prove by construction. Construct a \(BG\) network with \(m\) inputs, single output, \(p\) striatal neurons satisfying that \(S_j\) projects to the direct pathway if \(\delta(j) = 1\) and to the indirect pathway if \(\delta(j) = 0\), \(j = 1, \ldots, p\), and nonnegative weights \(w_{ij}\) determined by

\[
w_{ij} = \frac{1}{\sum_{k=1}^{m} C_k^{(i)}} C_k^{(j)}, \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, p. \tag{3.7}
\]

Next we verify that under the above construction, \(BG(C(j))\) equals \(\delta(j)\) for any \(j = 1, \ldots, p\).

When \(C = C(j)\), we have \(d_j = \sum_{i=1}^{m} w_{ij} C_i^{(j)} = \frac{1}{\sum_{k=1}^{m} C_k^{(i)}} \sum_{i=1}^{m} (C_i^{(j)})^2 = 1\) (note that \((C_i^{(j)})^2 = C_i^{(j)}\) since \(C_i^{(j)}\) takes binary value). Meanwhile, for any \(l \not= j\), \(d_l = \sum_{i=1}^{m} w_{il} C_i^{(j)} = \frac{1}{\sum_{k=1}^{m} C_k^{(l)}} \sum_{l \in L(i)} C_i^{(j)} < \frac{1}{\sum_{k \in L(i)}} \sum_{l \in L(i)} 1 = 1\), because \(L(l)\) is not a subset of \(L(j)\). Therefore, we have \(d_j > d_l\) for any \(l \not= j\), and thus only \(S_j\) equals 1 and other \(S_l\) equal 0 for \(C = C(j)\). Since \(S_j\) projects to the direct pathway if \(\delta(j) = 1\) and to the indirect pathway if \(\delta(j) = 0\), we have that \(BG(C(j)) = \delta(j)\). Q.E.D.

The network construction presented in the proof of Proposition 3.1 needs \(p\) striatal neurons for the \(BG\) network to realize an input-output relation compatible with the \(p\) given input-output pairs, \((C(j), \delta(j))\), \(j = 1, \ldots, p\). In the following, we further show
that under certain conditions a more compact BG network can be constructed to realize the same set of input-output pairs as expected.

**Condition 3.1** Given $D_C$ and $I_C$ defined by (3.5) and (3.6), let $D^1, \ldots, D^r$ be a partition of $D_C$ (i.e., $D^1 \cup \cdots \cup D^r = D_C$ and $D^k \cap D^l = \emptyset$ for any $k \neq l$) and $I^1, \ldots, I^s$ a partition of $I_C$. These partitions further satisfy:

- For any $k = 1, \ldots, r$, $\bigcap_{j \in D^k} L(j) \not\subset L(l)$ for any $l \in I_C$.
- For any $k = 1, \ldots, r$, $\bigcap_{j \in D^k} L(j) \not\subset L(l)$ for any $l \in D_C - D^k$.
- For any $k = 1, \ldots, s$, $\bigcap_{j \in I^k} L(j) \not\subset L(l)$ for any $l \in D_C$.
- For any $k = 1, \ldots, s$, $\bigcap_{j \in I^k} L(j) \not\subset L(l)$ for any $l \in I_C - I^k$.

**Proposition 3.2** Let $C^{(1)}, \ldots, C^{(p)}$ be $p$ pairwise different nonzero binary vectors and $\delta^{(1)}, \ldots, \delta^{(p)}$ be any binary assignments. Assume that nonempty sets $D^1, \ldots, D^r$ and $I^1, \ldots, I^s$ constitute a partition of $D_C$ and a partition of $I_C$, respectively, and they satisfy Condition 3.1. Then there exists a BG channel with $r + s$ striatal neurons, $r$ of which project to the direct pathway and $s$ of which project to the indirect pathway, so that the input-output mapping of this BG channel, characterized by function $\text{BG}(\cdot)$, satisfies $\text{BG}(C^{(j)}) = \delta^{(j)}$ for any $j = 1, \ldots, p$.

**Proof:** We construct a BG network with $r + s$ striatal neurons, $r$ neurons (denoted $S_1, \ldots, S_r$) projecting to the direct pathway and $s$ neurons (denoted $S_{r+1}, \ldots, S_{r+s}$) to the indirect pathway. The corticostriatal connection weights $w_{ik}$, $i = 1, \ldots, m$, $k = 1, \ldots, r + s$, are given as follows:

- For $k = 1, \ldots, r$,
  
  $$w_{ik} = \begin{cases} 
  \frac{1}{|\bigcap_{j \in D^k} L(j)|} & \text{if } i \in \bigcap_{j \in D^k} L(j), \\
  0 & \text{if } i \not\in \bigcap_{j \in D^k} L(j),
  \end{cases}$$

  where $|L|$ represents the number of elements in a set $L$.  

51
• For $k = r + 1, \ldots, r + s$,

$$w_{ik} = \begin{cases} 
\frac{1}{\bigcap_{j \in I^{k-r}} L^{(j)}} & \text{if } i \in \bigcap_{j \in I^{k-r}} L^{(j)}, \\
0 & \text{if } i \not\in \bigcap_{j \in I^{k-r}} L^{(j)}.
\end{cases}$$

When the cortical context is $C^{(j)}$ and $j$ belongs to some $D^k \subset D_C$ (which implies that $k \leq r$ and $\delta_j = 1$), we have $d_k = \sum_i w_{ik} C_i^{(j)} = \sum_{i \in L^{(j)}} w_{ik} = 1$. It can be verified that $d_i$ is strictly less than 1 for any $l \neq k$ due to Condition 3.1. Hence $S_k = 1$ when $C = C^{(j)}$ and $j \in D^k$. So $BG(C^{(j)}) = 1 = \delta_j$.

Similarly, in the case that the cortical context is $C^{(j)}$ and $j$ belongs to some $I^{k-r} \subset I_C$ and $k \in \{r + 1, \ldots, r + s\}$ (which implies that $\delta_j = 0$), and we have $d_k = \sum_i w_{ik} C_i^{(j)} = 1$. It can be verified that $d_i$ is strictly less than 1 for $l \neq k$ due to Condition 3.1. Hence $S_k = 1$ when $C = C^{(j)}$ and $j \in I^{k-r}$. So $BG(C^{(j)}) = 0 = \delta_j$.

Therefore, we proved that based on the above construction of weights $BG(C^{(j)})$ equals $\delta_j$ for any $j = 1, \ldots, p$. Q.E.D.

Let us see an example which demonstrates the realization of a mapping from seven cortical contexts to seven binary assignments using a BG network. The seven cortical contexts are

$$C^{(1)} = [0, 0, 1, 1, 0, 0]^T,$$  
$$C^{(2)} = [1, 1, 1, 0, 0, 0]^T,$$  
$$C^{(3)} = [0, 1, 1, 0, 1, 0]^T,$$  
$$C^{(4)} = [0, 1, 1, 0, 0, 1]^T,$$  
$$C^{(5)} = [0, 1, 0, 1, 1, 1]^T,$$  
$$C^{(6)} = [0, 0, 0, 1, 1, 1]^T,$$  
$$C^{(7)} = [1, 0, 0, 1, 1, 1]^T,$$

and the corresponding binary assignments are

$$\left[\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}, \delta^{(5)}, \delta^{(6)}, \delta^{(7)}\right] = [1, 1, 1, 1, 0, 0, 0].$$

Then we have $D_C = \{1, 2, 3, 4\}$ and $I_C = \{5, 6, 7\}$. Consider the following partitions of $D_C$ and $I_C$: $D^1 = \{1\}$, $D^2 = \{2, 3, 4\}$, and $I^1 = I_C = \{5, 6, 7\}$. Since

$$L^{(1)} = \{3, 4\}, \quad L^{(2)} = \{1, 2, 3\},$$  
$$L^{(3)} = \{2, 3, 5\}, \quad L^{(4)} = \{2, 3, 6\},$$  
$$L^{(5)} = \{2, 4, 5, 6\}, \quad L^{(6)} = \{4, 5, 6\},$$  
$$L^{(7)} = \{1, 4, 5, 6\}.$$
we have \( \cap_{j \in D^1} L^{(j)} = \{3, 4\} \), \( \cap_{j \in D^2} L^{(j)} = \{2, 3\} \), and \( \cap_{j \in I} L^{(j)} = \{4, 5, 6\} \). Obviously, 
\( \cap_{j \in D^1} L^{(j)} \) is not a subset of \( L^{(l)} \) for any \( l \in I_C \) and for any \( l \in D_C - D^1 = D^2 \);
\( \cap_{j \in D^2} L^{(j)} \) is not a subset of \( L^{(l)} \) for any \( l \in I_C \) and for any \( l \in D_C - D^2 = D^1 \);
\( \cap_{j \in I} L^{(j)} \) is not a subset of \( L^{(l)} \) for any \( l \in D_C \). So Condition 3.1 is satisfied, and
following Proposition 3.2, we need only three neurons in the striatal network (WTA layer) with two (denoted \( S_1 \) and \( S_2 \)) projecting to the direct pathway and one (denoted \( S_3 \)) projecting to the indirect pathway so as for the BG channel to realize the mapping from \( C^{(j)} \) to \( \delta^{(j)} \), \( j = 1, \ldots, 7 \), and the corticostriatal weights may be chosen as
\[
\begin{align*}
w_1 &= [0, 0, 1/2, 1/2, 0, 0]^T, \\
w_2 &= [0, 1/2, 0, 0, 0, 0]^T, \\
w_3 &= [0, 0, 1/3, 1/3, 1/3]^T.
\end{align*}
\]

It can be seen from Propositions 3.1 and 3.2 (the former can be considered as a special case of the latter) and their proofs that these two propositions do not apply to the situations where \( L^{(j)} \) is a subset of \( L^{(k)} \) for some \( j \in D_C \) and \( k \in I_C \) or \( (j \in I_C \) and \( k \in D_C) \). For example, \( C^{(1)} = [1, 1, 0, 0, 0]^T \) and \( C^{(2)} = [1, 1, 1, 0, 0]^T \) are associated with \( \delta^{(1)} = 1 \) and \( \delta^{(2)} = 0 \), respectively, and in this case, \( L^{(1)} \) is a subset of \( L^{(2)} \).
Following the construction of weights as (3.7), the WTA layer of the BG network will have problem to judge which one wins—the direct pathway or indirect pathway, when the input cortical context is \( C^{(2)} \). In the following, we propose a construction of weights that can deal with any pairwise different cortical contexts.

**Proposition 3.3** Let \( C^{(1)}, \ldots, C^{(p)} \) be \( p \) pairwise different nonzero vectors and \( \delta^{(1)}, \ldots, \delta^{(p)} \) be any binary assignments: \( C^{(j)} \in \{0, 1\}^m \) and \( \delta^{(j)} \in \{0, 1\} \) for \( j = 1, \ldots, p \). Then there exists a BG channel with \( p \) striatal neurons and \( m \times p \) nonnegative weights so that the input-out mapping of this BG channel satisfies \( \text{BG}(C^{(j)}) = \delta^{(j)} \) for any \( j = 1, \ldots, p \).

Before proceeding with the proof, let us see a particular choice of weights \( w_{ij} \) for the corticostriatal connections, with which the BG network may realize universal logic mapping from \( \{0, 1\}^m \) to \( \{0, 1\} \). Given \( p \) nonzero cortical contexts \( C^{(1)}, \ldots, C^{(p)} \),
let

\[ w_{ij} = \frac{1}{\sqrt{|L(j)|}} C^{(j)}_i, \quad i = 1, \ldots, m \text{ and } j = 1, \ldots, p, \quad (3.8) \]

where \(|L(j)|\) denotes the number of elements in \(L(j)\), which equals \(\sum_i C^{(j)}_i > 0\). In vector form, \((3.8)\) is equivalent to

\[ w_j = C^{(j)}/ \| C^{(j)} \|, \quad j = 1, \ldots, p, \quad (3.9) \]

where \(\| x \|\) represents the Euclidean norm of a vector \(x\), defined by \((x^T x)^{\frac{1}{2}}\). Note that \(\| C^{(j)} \|\) equals \((\sum_{i=1}^m C^{(j)}_i)^{\frac{1}{2}} = (\sum_{i=1}^m C^{(j)}_i)^{\frac{1}{2}} = \sqrt{|L(j)|}\) since \(C^{(j)}\) is a binary vector. Obviously, \((3.9)\) implies

\[ \| w_j \| = 1, \quad j = 1, \ldots, p. \quad (3.10) \]

Such choice of weights is important for the basal ganglia, because the corticostriatal connection weights, regardless of their initial values, tend to converge to \((3.8)\) or \((3.9)\) following the learning of the striatal networks. Convergence of the BG learning process will be proved in the next section.

**Proof of Proposition 3.3:** We construct a BG network with \(m\) inputs, single output, \(p\) striatal neurons satisfying that \(S_j\) projects to the direct pathway if \(\delta^{(j)} = 1\) and to the indirect pathway if \(\delta^{(j)} = 0, \quad j = 1, \ldots, p\), and nonnegative weights \(w_{ij}\) satisfying \((3.8)\) or \((3.9)\). Next we verify that under the above construction, \(BG(C^{(j)})\) equals \(\delta^{(j)}\) for any \(j = 1, \ldots, p\).

When \(C = C^{(j)}, \quad j \in \{1, \ldots, p\}\), we have \(d_j = w_j^T C^{(j)} = (C^{(j)T} C^{(j)})^{\frac{1}{2}} \| C^{(j)} \| = \| C^{(j)} \| = \sqrt{|L(j)|}\). To estimate \(d_k, \quad k \neq j\), consider the following three cases.

Case 1: \(|L^{(k)}| > |L^{(j)}|\). In this case, using inequality \(C^{(k)}_i C^{(j)}_i \leq C^{(j)}_i\), we have \(d_k = w_k^T C^{(j)} = \sum_i \frac{1}{\sqrt{|L^{(k)}|}} C^{(k)}_i C^{(j)}_i \leq \sum_i \frac{1}{\sqrt{|L^{(k)}|}} C^{(j)}_i = \frac{1}{\sqrt{|L^{(k)}|}} |L^{(j)}| < \frac{1}{\sqrt{|L^{(j)}|}} |L^{(j)}| = \sqrt{|L^{(j)}|}\). Hence, \(d_k < d_j\).

Case 2: \(|L^{(k)}| < |L^{(j)}|\). In this case, using inequality \(C^{(k)}_i C^{(j)}_i \leq C^{(k)}_i\), we have \(d_k = \sum_i \frac{1}{\sqrt{|L^{(k)}|}} C^{(k)}_i C^{(j)}_i \leq \sum_i \frac{1}{\sqrt{|L^{(k)}|}} C^{(k)}_i = \frac{1}{\sqrt{|L^{(k)}|}} |L^{(k)}| = \sqrt{|L^{(k)}|} < \sqrt{|L^{(j)}|}\). Thus \(d_k < d_j\).

Case 3: \(|L^{(k)}| = |L^{(j)}|\). Since \(C^{(k)} \neq C^{(j)}\), \(\sum_i C^{(k)}_i C^{(j)}_i\) must be strictly smaller than \(\sum_i C^{(j)}_i = |L^{(j)}|\). Therefore, \(d_k = \sum_i \frac{1}{\sqrt{|L^{(k)}|}} C^{(k)}_i C^{(j)}_i < \frac{1}{\sqrt{|L^{(j)}|}} |L^{(j)}| = \sqrt{|L^{(j)}|}\), so \(d_k < d_j\).

|\(j\)|  | \(\delta^{(j)}\)  |
|-----|-----------------|
| 1   | 1               |
| 2   | 0               |
| 3   | 1               |
| 4   | 0               |
| 5   | 1               |
| 6   | 0               |

|\(k\)|  | \(d_k\)  |
|-----|---------|
| 1   | 0.5     |
| 2   | 1.5     |
| 3   | 2.5     |
| 4   | 3.5     |
| 5   | 4.5     |
| 6   | 5.5     |

Therefore, \(d_k < d_j\) for all \(k \neq j\).
Therefore $d_j > d_k$ for any $k \neq j$ when $C = C^{(j)}$, and thus only $S_j$ equals 1 and other $S_k$ equal 0 for $C = C^{(j)}$. Since $S_j$ projects to the direct pathway if $\delta^{(j)} = 1$ and to the indirect pathway if $\delta^{(j)} = 0$, we have that $BG(C^{(j)}) = \delta^{(j)}$. Q.E.D.

The geometry underlying the proof of Proposition 3.3 is that the neuron whose weight vector is the most similar to the input cortical context wins the competition. Write each $d_j$ in the following form

$$d_j = w_j^T C = \| w_j \| \cdot \| C \| \cos \theta_j = \| C \| \cos \theta_j,$$

where $\theta_j$ is the angle between the weight vector $w_j$ and the cortical context vector $C$. Since the weights are determined by (3.9), $\theta_j$ becomes 0 and $\cos \theta_j$ equals 1 when the cortical input is $C^{(j)}$, while other $\cos \theta_k$ is less than 1 for any $k \neq j$. So $d_j$ wins and other $d_k$ lose when $C = C^{(j)}$.

Not restricted to the choice of $w_{ij}$ as given in (3.8) or (3.9), we can show that a BG network may act as a universal pattern classifier based on any construction of $w_{ij}$ that satisfies the condition below, and the proof follows the similar steps as in the proof of Proposition 3.3.

**Condition 3.2** Given $p$ nonzero cortical contexts $C^{(1)}, ..., C^{(p)}$, we have

- $w_{ij} = 0$ if $i \notin L^{(j)}$,
- $w_{ij} > w_{ik}$ if $|L^{(j)}| < |L^{(k)}|$ (i.e., $\sum_i C_i^{(j)} < \sum_i C_i^{(k)}$) and $i \in L^{(j)} \cap L^{(k)}$,
- $w_{ij} = w_{ik}$ if $|L^{(j)}| = |L^{(k)}|$ and $i \in L^{(j)} \cap L^{(k)}$,
- $\sum_i w_{ij} < \sum_i w_{ik}$ if $|L^{(j)}| < |L^{(k)}|$, where $j, k \in \{1, ..., p\}$.

It can be verified that (3.9) is a special case of weight constructions satisfying Condition 3.2.
3.3 Competitive Learning of Corticostriatal Networks

It has been widely accepted that the basal ganglia and the dopamine system provide a crucial brain mechanism for mediating learning by reinforcement [11, 77, 195]. Although reinforcement learning has many advantageous properties, it has an important limitation, called the credit assignment problem [8, 125]. This problem includes the temporal credit assignment problem, i.e., how to get reinforcement signals at the right time, and spatial credit assignment problem, i.e., how to get reinforcement signals at the right place (synapses). Here we revisit the “adaptive switching” part of the MIMOAS model (see Chapter 2) and address the spatial credit assignment problem in the basal ganglia, specifically, the corticostriatal networks.

We propose for the corticostriatal networks a competitive learning rule, which combines a three-factor rule [122, 195] and synaptic normalization. We use the term of competitive learning to emphasize that under the proposed architecture of learning the spatial credit assignment problem is solved by the BG via competition among striatal neurons and synapses: The strengthening or weakening of corticostriatal connections primarily happen at synapses whose postsynaptic neuron wins the competition with other projection neurons—this is revealed by the three-factor rule. At the same time, we introduce the concept of homeostatic synaptic plasticity or synaptic normalization, which is shown to be essential to stabilize the corticostriatal network during the process of reinforcement learning. We prove that the proposed learning rule ensures the network to converge with expected input-output relations.

3.3.1 Three-Factor Rule of Competitive Learning

Dopamine contributes significantly to the synaptic long-term potentiation (LTP) and long-term depression (LTD) in the striatum. It is recognized that simultaneous presence of cortical glutamatergic input activity, phasic increase of dopamine concentration, and striatal neuron activity cause a corticostriatal synapse to undergo LTP [156, 195]. This learning rule involves “three factors”: phasic release of dopamine, presynaptic activity, and postsynaptic activity [195]. Wickens proposed
Table 3.2: Effects of the three factors and dopamine receptor ratio on corticostriatal synaptic plasticity, where $\leftrightarrow$ represents no change, $\downarrow$ LTD, $\uparrow$ LTP, and * can be either 0 or 1.

<table>
<thead>
<tr>
<th>Presynaptic activity</th>
<th>Postsynaptic activity</th>
<th>Dopamine</th>
<th>Direct pathway (higher D1:D2 ratio)</th>
<th>Indirect pathway (lower D1:D2 ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>*</td>
<td>*</td>
<td>$\leftrightarrow$</td>
<td>$\leftrightarrow$</td>
</tr>
<tr>
<td>*</td>
<td>0</td>
<td>*</td>
<td>$\leftrightarrow$</td>
<td>$\leftrightarrow$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$\downarrow\downarrow$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$\uparrow\uparrow$</td>
<td>$\uparrow$</td>
</tr>
</tbody>
</table>

that LTP occurs at a conjunction of all three factors, and repeated conjunctions of presynaptic and postsynaptic activations in the absence of phasic release of dopamine leads to weakening of synaptic connections [195]. It should be noted that the lasting effects of dopamine in the striatum are complex and experimental results contradictory to the three-factor rule have been seen [149, 176, 178].

The MIMOAS model integrates the findings of many experiments relating dopamine to synaptic long-term plasticity [23, 27, 28, 156, 192] and extends Wickens’ learning rule to incorporate plausible differences in the synaptic changes that may be mediated by D1 and D2 receptors. Specifically, it has been proposed by Silkis [173, 174] that activity of NMDA (glutamatergic) receptors with additional increased levels of intracellular $Ca^{2+}$ promotes LTP in striatal projection neurons, while activity of NMDA receptors with decreased $Ca^{2+}$ levels reduces LTP or even reverses LTP into LTD. In contrast, inactivity of NMDA receptors with increased levels of intracellular $Ca^{2+}$ (or, respectively, with decreased $Ca^{2+}$ levels) promotes LTD (or reduces LTD, respectively) [173, 174]. It has also been seen that both D1 and D2 receptors mediate intracellular $Ca^{2+}$ levels, but the former are more potent in increasing $Ca^{2+}$ levels than the latter [173, 174]. The details of the interactions of these processes are complex and have not been fully clarified. However, the major effects are consistent with the possibility that LTP and LTD occur strongly in striatal neurons projecting to the direct pathway because of the greater preponderance of D1 over D2 receptors there. And weaker plasticity occurs in the striatal neurons projecting to the indirect
Figure 3-2: Striatal neuronal connections and neurotransmitters. Striatal projection neurons receive glutaminergic (Glu) input from cortical neurons, SNc neurons release dopamine (DA, dashed), striatal inhibitory interneurons are cholinergic (Ach) or GABAergic, and striatal neuron collateral inhibitory projections are GABAergic.

pathway because of the lower ratio of D1 receptors to D2 receptors. The assumed dynamics of dopamine-dependent synaptic adaptation are summarized in Table 3.2.

The rule of synaptic formation involving “three factors” enables Hebbian-like learning to be modulated by a reward/punishment input. Specifically, with respect to this system it has been observed in vivo that the dopamine release undergoes phasic increase during positive reinforcement of rewarded behaviors [167], while dopamine level drops below baseline during aversive events or when the choices of actions do not lead to reward [189].

The effect of these synaptic dynamics on BG reinforcement learning can be seen as follows: Referring to Fig. 3-2, assume that the context is $C^{(j)}$ (with only two digits of the context vector, $C_1^{(j)}$ and $C_2^{(j)}$, shown in Fig. 3-2), and consider two striatal projection neurons, $S_{D_j}$ and $S_{I_j}$. Suppose “go” is rewarded and “no go” is punished (or not rewarded) and that the initial strengths of the corticostriatal connections enable movement because $S_{D_j} = 1$ and $S_{I_j} = 0$ result from winner-take-all competition. In this case, dopamine phasically increases in the presence of context of $C^{(j)}$ and activation of $S_{D_j}$. According to Table 3.2, the connections between those active digits (neurons) in the context vector and the striatal neuron $S_{D_j}$ are strengthened. The result is that the context becomes more strongly associated with $S_{D_j}$ in the direct pathway. On the other hand, if the initial corticostriatal connections cause $S_{I_j} = 1$, then the output of the BG channel favors “no go,” resulting in punishment or no reward. In this case, release of dopamine drops during the
presence of context $C^{(j)}$ and activation of $S_{I_j}$ attenuating the connections between the context vector and the striatal neuron $S_{I_j}$ while those between the context and $S_{D_j}$ undergo no significant change. Neuron $S_{I_j}$ gradually loses the competition with $S_{D_j}$, which then becomes reinforced as described above. A complementary analysis shows that reward for “no go” or punishment for “go” cause $C^{(j)}$ to become associated with $S_{I_j}$. Thus, the correct channel assignment develops irrespective of initial connection conditions.

Because adaptation in both directions is accelerated by activity in the striatal neuron, additional excitatory “teaching signals” (e.g., signal $T$, as shown in Fig. 3-2) could enhance the learning within certain domains of striatal neurons.

### 3.3.2 Learning Rule Combining Three-factor Rule and Synaptic Normalization

Although the three-factor rule can successfully adapt the BG network to associate a single cortical context to an action (like “go” or “no go”), we still do not know if the rule can handle the situation where arbitrary set of input-output pairs need to be learned by the network; although Proposition 3.3 shows that there always exists a set of corticostriatal connection weights that ensure the BG network to realize any mapping from $\{0,1\}^m$ to $\{0,1\}$, we still do not know if the weights may converge to the expected ones during the learning process. In fact, computer simulations of some learning tasks of the BG (not shown here) have suggested that direct implementation of the three-factor rule as described in Table 3.2 does not always stabilize the connection weights of the network: The weights may oscillate during the learning of multiple input-output pairs and the network may not be guaranteed to converge.

In the following we introduce a modified learning rule, which may stabilize the learning process and which is more physiologically plausible than the original three-factor rule. The modification is made based on the concept of “homeostasis.” Homeostasis has recently been brought to synaptic physiology [22, 121]. When the activity of neurons is beyond the normal operating range, a form of adaptation occurs that changes the gain of all synapses [22]. Scanziani et al. [163] found that in hippocampus LTP of some inputs leads to LTD of inputs in a neighboring volume, which could keep
rough conservation of synaptic weight over a postsynaptic cell. Turrigiano et al. [188] showed that a change at a single synapse can cause the change of efficacy of the whole cell, suggesting that local change in the synaptic strength scales the strength of the other synapses of the same neuron. The homeostatic synaptic plasticity has also been suggested to explain the post-traumatic hyperexcitability and epileptogenesis in chronically isolated neocortex [79]. With homeostasis, neurons can stabilize their excitability while retaining relative differences in strength among synapses. It seems logical that evolution would endow a cell with a way of controlling the strength of its input to ensure that the cell remains within a meaningful operating range [121].

The homeostatic synaptic plasticity is not a part of our current understanding of LTP and LTD [121], and its mechanisms are still not completely understood [198]. To our best knowledge, the homeostatic synaptic plasticity has not been applied to the study of the effect of LTP and LTD in the corticostriatal synapses. However, we postulate that the mechanism of homeostasis may be essential to interpret the discrepancy between the three-factor rule and some experiment findings about the lasting effect of dopamine in the striatum [149, 176, 178].

Homeostatic synaptic plasticity and synaptic resource redistribution have been introduced to computational models and considered as physiological basis for the technique of synaptic weight normalization [29, 40]. Here normalization means to impose some global constraint to synaptic weights. Two types of constraints are commonly used: One is to hold the sum of all the weights of synapses onto a given postsynaptic neuron to a constant value, and the other is to constrain the sum of squares of the weights instead of their linear sum [33]. In the analysis presented in this study, we propose to use the latter form of normalization.

Before normalization, the three-factor rule can be expressed as follows. Let $w_{ij}(T)$ be the updated weight of a corticostriatal connection (linking the $i$-th cortical neuron and the $j$-th striatal neuron) right after the $T$-th event of learning, where $T$ is a non-negative integer. By “event of learning,” we mean a discrete event during which the cortico-BG network is presented with a cortical context, say $C^{(j)}$, and is rewarded or punished with respectively increased or decreased levels of dopamine release associated with the action of the network in response to the cortical context. Without loss
of generality, let the $j$-th striatal projection neuron $S_j$ be the representative neuron for the cortical context $C^{(j)}$, i.e., the network will get rewarded if $S_j$ wins the competition with other striatal neurons when the network is presented with $C^{(j)}$. Let $w_{ij}(0)$ be the initial value of $w_{ij}$. Then according to the three-factor rule, when $C$ equals $C^{(j)}$, $S_j$ equals 1, and dopamine release undergoes phasic increase, we have

$$w_{ij}^0(T) = \begin{cases} w_{ij}(T-1) + \eta_{\text{LTP}}^D & \text{if } i \in L^{(j)} \text{ and } S_j \text{ projects to direct pathway,} \\ w_{ij}(T-1) + \eta_{\text{LTP}}^I & \text{if } i \in L^{(j)} \text{ and } S_j \text{ projects to indirect pathway,} \\ w_{ij}(T-1) & \text{if } i \not\in L^{(j)}, \end{cases}$$

(3.11)

where $\eta_{\text{LTP}}^D$ and $\eta_{\text{LTP}}^I$ are positive constants characterizing the weight strengthening (LTP) of a corticostriatal connection in the direct pathway and in the indirect pathway, respectively, during the reinforcement learning (according to Table 3.2, we have $\eta_{\text{LTP}}^D > \eta_{\text{LTP}}^I$), and we use the superscript 0 to differentiate these weight update equations without normalization from those with normalization. When $C$ equals $C^{(k)}$, $k \neq j$, $S_j$ equals 1, and dopamine release undergoes decrease, we have

$$w_{ij}^0(T) = \begin{cases} \max \left\{ 0, w_{ij}(T-1) - \eta_{\text{LTD}}^D \right\} & \text{if } i \in L^{(j)} \text{ and } S_j \text{ projects to direct pathway,} \\ \max \left\{ 0, w_{ij}(T-1) - \eta_{\text{LTD}}^I \right\} & \text{if } i \in L^{(j)} \text{ and } S_j \text{ projects to indirect pathway,} \\ w_{ij}(T-1) & \text{if } i \not\in L^{(j)}, \end{cases}$$

(3.12)

where $\eta_{\text{LTD}}^D$ and $\eta_{\text{LTD}}^I$ are positive constants characterizing the weight weakening (LTD) of a corticostriatal connection during the reinforcement learning (according to Table 3.2, we have $\eta_{\text{LTD}}^D > \eta_{\text{LTD}}^I$). When $C$ equals $C^{(k)}$, $k \neq j$, and $S_j$ equals 0, we have

$$w_{ij}^0(T) = w_{ij}(T-1).$$

(3.13)

Now taking into account the homeostatic synaptic plasticity, we assume the sum of all the weights of corticostriatal connections targeting on a given striatal neuron is fixed. Without loss of generality, we assume

$$\sum_i w_{ij}(T)^2 = 1 \quad \text{or} \quad \|w_j(T)\| = 1$$

(3.14)

for any $j$ and any $T$. Then we introduce the modified learning rule based on the
three-factor rule and synaptic normalization:

\[ w_{ij}(T) = \frac{w_{ij}^0(T)}{\| w_{ij}^0(T) \|} \quad (3.15) \]

where \( w_{ij}^0(T) \) is determined by (3.11), (3.12), or (3.13), and \( w_{ij}^0(T) \) denotes the vector consists of \( w_{1j}^0(T), \ldots, w_{mj}^0(T) \).

In the next subsection, we prove the convergence of the modified learning rule.

### 3.3.3 Convergence of Competitive Learning

**Proposition 3.4** Let the corticostriatal network be repeatedly presented with a nonzero cortical context, \( C^{(j)} \), and let the network be rewarded with phasically increased dopamine release associated with \( S_j = 1 \). Then the learning rule of (3.11) and (3.15) ensures \( w_j(T) \to C^{(j)} \| C^{(j)} \|, \) as \( T \to \infty \).

**Proof:** We only need to prove for the case that \( S_j \) projects to the direct pathway. The proof for the case that \( S_j \) projects to the indirect pathway follows the same steps.

According to (3.11) and (3.15), we have

\[
\begin{align*}
\frac{w_{ij}(T) - 1}{w_{ij}(T)} &= \frac{\eta_{\text{LTP}}^D}{\sum_{k \in L^{(i)}} [w_{kj}(T - 1) + \eta_{\text{LTP}}^D]^2 + \sum_{k \notin L^{(i)}} w_{kj}(T - 1)^2} \\
\frac{w_{ij}(T) - 1}{w_{ij}(T)} &= \frac{\eta_{\text{LTP}}^D}{\sum_{k \in L^{(i)}} [w_{kj}(T - 1) + \eta_{\text{LTP}}^D]^2 + \sum_{k \notin L^{(i)}} w_{kj}(T - 1)^2} \\
\text{for } i \in L^{(j)} \text{ and } \\
\text{for } i \notin L^{(j)}. 
\end{align*}
\]

According to \( (3.16) \) and \( (3.17) \), we have

\[
\frac{w_{ij}(T) - 1}{w_{ij}(T)} \to 0 \text{ for } i \notin L^{(j)} \text{ as } T \to \infty. \quad (3.18)
\]

According to (3.16) and (3.17), we have

\[
\| w_j(T) \| = 1 \text{ for any } T. \quad (3.19)
\]

62
Therefore,
\[
\left\{ \sum_{k \in L(j)} \left[ w_{k,j}(T-1) + \eta_{\text{LTP}}^D \right]^2 + \sum_{k \not\in L(j)} w_{k,j}(T-1)^2 \right\}^{\frac{1}{2}}
\]
\[
= \left\{ \sum_k w_{k,j}(T-1)^2 + \sum_{k \in L(j)} \left[ 2w_{k,j}(T-1)\eta_{\text{LTP}}^D + \eta_{\text{LTP}}^D \right]^2 \right\}^{\frac{1}{2}}
\]
\[
\geq \left( 1 + \sum_{k \in L(j)} \eta_{\text{LTP}}^D \right)^{\frac{1}{2}} > 1,
\]

since \( w_{k,j} \) is non-negative, \( \eta_{\text{LTP}}^D \) is greater than 0, and \( L(j) \) is nonempty. Hence, for \( i \not\in L(j) \)
\[
w_{ij}(T) \leq \left( \frac{1}{1 + \sum_{k \in L(j)} \eta_{\text{LTP}}^D} \right)^{\frac{1}{2}} w_{ij}(T-1) \leq \left( \frac{1}{1 + \sum_{k \in L(j)} \eta_{\text{LTP}}^D} \right)^{\frac{1}{2}} w_{ij}(0),
\]

which implies that \( w_{ij}(T) \to 0 \) for \( i \not\in L(j) \) as \( T \to \infty \).

Next we show
\[
w_{ij}(T) \to 1 / \| C^{(j)} \| \text{ for } i \in L(j) \text{ as } T \to \infty.
\] (3.20)

If we can prove
\[
w_{ij}(T)/w_{k,j}(T) \to 1 \text{ for any } i, k \in L(j) \text{ as } T \to \infty,
\] (3.21)

then (3.20) must be true, following (3.18) and (3.19). So we only need to prove (3.21).

Note that for any \( i, k \in L(j) \), we have
\[
\left| \frac{w_{ij}(T) - 1}{w_{k,j}(T)} \right| = \left| \frac{w_{ij}(T-1) + \eta_{\text{LTP}}^D}{w_{k,j}(T-1) + \eta_{\text{LTP}}^D} - 1 \right|
\]
\[
= \left| \frac{w_{ij}(T-1)}{w_{k,j}(T-1)} - 1 \right| \left| \frac{w_{k,j}(T-1)}{w_{k,j}(T-1) + \eta_{\text{LTP}}^D} \right|
\]
\[
\leq \left| \frac{w_{ij}(T-1)}{w_{k,j}(T-1)} - 1 \right| \left( 1 + \frac{\eta_{\text{LTP}}^D}{1 + \eta_{\text{LTP}}^D} \right),
\]

since \( w_{k,j} \) is nonnegative and no greater than 1 (according to (3.19)) and \( \frac{w_{k,j}}{w_{k,j} + \eta_{\text{LTP}}^D} \) is
a monotone increasing function of $w_{kj}$ for $w_{kj} > 0$. Therefore,

$$\left| \frac{w_{ij}(T)}{w_{kj}(T)} - 1 \right| \leq \left| \frac{w_{ij}(0)}{w_{kj}(0)} - 1 \right| \left( \frac{1}{1 + W_{LTP}} \right)^T,$$

which approaches 0 as $T \to \infty$. So we proved (3.21). Q.E.D.

It can be seen that under the same scenario as presented in Proposition 3.4 the three factor rule without synaptic normalization simply leads to the unconstrained growth of the strengths of some corticostrial connections. Meanwhile, the modified learning rule with normalization ensures the network to converge to (3.9). As demonstrated in the proof of Proposition 3.3, the corticostriatal connection weights determined by (3.9) guarantee the BG network to realize universal logic mapping from \{0, 1\}^m to \{0, 1\}. Proposition 3.4, however, does not mention how the corticostriatal network evolves to find an appropriate “representative” striatal neuron $S_j$ (that projects to the expected pathway, either direct or indirect pathway) for a given cortical input $C^{(j)}$. Nevertheless, an appropriate striatal neuron associated with a cortical context may be figured out via the procedure involving both reward and punishment as described in Section 3.3.1, and may also be determined following the guidance of some cortical “teaching signals” (see Fig. 3-2).
Chapter 4

Dynamics of Winner-take-all Competition in Striatal Networks

This chapter studies the dynamics of striatal networks and the mechanism of winner-take-all competition among striatal projection neurons. The study is not restricted to the striatal networks in the basal ganglia, however. It is approached in a more general framework of recurrent neural networks with lateral inhibition. Theoretical analysis is then applied to understand the operations of the striatal networks under both physiological and pathophysiological conditions. This consequently explains the significance of dopamine and acetylcholine in the modulation of selectivity of competing corticostriatal activities.

Winner-take-all via lateral inhibition is an important mechanism for many fundamental computational abilities of animal neural systems. Since the finding of lateral inhibition in the eye of Limulus [69], lateral inhibition has been discovered in many neural circuits, such as those in vertebrate somatosensory cortex [135], auditory cortex [4], visual system [89], hippocampus [162], and basal ganglia [123]. In sensory systems, the process of lateral inhibition improves signal resolution via a simple competitive or winner-take-all mechanism in local neural networks [135]; in some cortical areas, the winner-take-all competition among excitatory neurons with delayed inhibition from interneurons has been suggested to play a key role in the generation of certain natural rhythms such as the 4-7 Hz theta rhythm in hippocampus [32]; and in the basal ganglia, winner-take-all has been proposed to be the underlying mechanism for the focused selection of competing motor programs [63, 64, 123]. Therefore, as
a common functional feature of neural systems, winner-take-all mechanisms deserve further investigation.

Considerable effort has already been made to explain how lateral inhibition can lead to winner-take-all competition among neurons (e.g. [32, 41, 51, 65, 201]). Coultrip et al. [32] simulated and analyzed physiological interactions among excitatory and inhibitory neurons in a modeled network architecture of hippocampal field CA1. They demonstrated the generation of a simple winner-take-all mechanism, which allows only the most strongly-activated cell in a group to respond with spiking activity. Ermentrout [41] studied the complex dynamics in winner-take-all neural networks, and showed that as the inhibition slows down, the winner-take-all networks may exhibit oscillatory behaviors. In the above two insightful studies ([32] and [41]), however, the network architectures were relatively simple: A layer of excitatory neurons coupled with a single inhibitory interneuron were considered. Hahnloser [65] provided both quantitative analysis and simulations, showing that global inhibition may give rise to multi-stable winner-take-all mechanism in a recurrent networks of neurons. Again, this study was restricted to single, global inhibition and, in addition, Hahnloser considered for each neuron only linear activation function above threshold. Fukai and Tanaka [51] considered a neural network with uniform lateral inhibition and self-inhibition. They found that the strength of lateral inhibition relative to that of self-inhibition is crucial for determining the steady states of the network. Different ratios between the strengths of lateral inhibition and self-inhibition may lead to either winner-take-all or winners-share-all behaviors. Fukai and Tanaka’s work provided profound mathematical basis for understanding neural selection mechanisms; however, their analysis was subject to a specific choice of activation function, i.e., Lotka-Volterra equation [51], for neuronal firing activity. Xie et al. [201] extended the grouping of potential winners in the winner-take-all networks beyond single neuron or uniformly arranged groups of neurons. They showed that competition between arbitrary groups of neurons can be realized by organizing lateral inhibition in networks. However, as the authors acknowledged, the activation function considered in their paper was linear threshold function, which fails to characterize the saturation of neuronal response to large input.
Beyond the scope of neural systems, other efficient realizations of winner-take-all (not so biologically plausible, however) were also proposed [45, 102, 109, 181, 200, 203, 204]. Analog circuits for winner-take-all have been implemented in real applications as well [205].

This chapter focuses on the realization of winner-take-all mechanisms in biologically plausible recurrent neural networks with lateral inhibition such as the modeled striatal networks in the basal ganglia. A broad range of neuronal activation functions are considered, including the discontinuous threshold function, which models the activity of neurons with infinite gain in limiting situations. Conditions are derived for the network to present winner-take-all competition among neurons, and the underlying mechanism of winner-take-all is illustrated in terms of augmented separation of supra-threshold activities of competing neurons. In addition to winner-take-all, this chapter also discusses other characteristics of the networks with lateral inhibition, such as the effect of hysteresis and the number and stability of the network equilibria. Then the results obtained from theoretical analysis are applied to investigate the mechanism of selection and switching in the striatal networks of the basal ganglia under both normal and pathophysiological conditions.

4.1 Equations for Neuronal Networks with Lateral Inhibition

Consider a recurrent network of \( n \) neurons with lateral inhibition. Denote \( x_i \) the postsynaptic membrane potential and \( y_i \) the firing rate of neuron \( i \), where \( i \) ranges from 1 to \( n \). Let \( d_i \) be the input (or sum of inputs if there are more than one inputs) to neuron \( i \) from the outside of the \( n \)-neuron network, and \( v_{ik} > 0 \) be the synaptic strength of the lateral inhibitory connection from neuron \( i \) to neuron \( k \) for any \( k \neq i \). Here we assume that the inhibitory connections originating from the same neuron take the same strength, and we do not consider the case of self-inhibition, where a neuron forms inhibitory connection from its output to its own input. Then the
dynamics of the neural network can be modeled by

\[
\frac{dx_i}{dt} = -x_i - \sum_{k \neq i} v_k y_k + d_i, \quad (4.1)
\]

\[
y_i = f_i(x_i), \quad (4.2)
\]

where \(i\) takes \(1, \ldots, n\), \(\tau\) is a time constant, and \(f_i(\cdot)\) is a nonnegative activation function characterizing the relation between the membrane potential and firing rate of neuron \(i\). In matrix-vector form, denote \(x = [x_1, \ldots, x_n]^T\), \(y = [y_1, \ldots, y_n]^T\), \(d = [d_1, \ldots, d_n]^T\), and \(V = [v_{ik}]_{n \times n}\) with \(v_{ii} = 0\) and \(v_{ik} = v_k\) for \(i \neq k\). Then (4.1) can be written in a compact form as

\[
\tau \frac{dx}{dt} = -x - V y + d. \quad (4.3)
\]

Especially, the dynamics of a group of reciprocally connected projection neurons in some local area of the striatum can be modeled by (4.1) and (4.2), where the network receives inputs from a variety of cortical neurons (\(d_i\) then represents the sum of cortical inputs to the \(i\)-th striatal projection neuron) and sends outputs \((y_i)\) to the next stages of the basal ganglia, i.e., GPi, SNr, and GPe.

An *equilibrium* of (4.1) and (4.2) is a solution to \(-x - V y + d = 0\), i.e.,

\[
x_i = d_i - \sum_{k \neq i} v_k f_k(x_k), \quad i = 1, \ldots, n. \quad (4.4)
\]

An equilibrium \(x^*\) has the property that whenever the state of the system starts at \(x^*\) it will remain at \(x^*\) for all future time. Due to the nonnegativity of \(f_i(\cdot)\), (4.4) implies

\[
x_i^* \leq d_i, \quad i = 1, \ldots, n. \quad (4.5)
\]

An equilibrium \(x^*\) is *stable* if, for each \(\epsilon > 0\), there is \(\delta > 0\) such that

\[
\| x(0) - x^* \| < \delta \Rightarrow \| x(t) - x^* \| < \epsilon \quad \text{for any} \quad t \geq 0,
\]

where \(\| x \|\) can be any norm, but usually we use the Euclidean norm \(\| x \| = (x_1^2 + \cdots + x_n^2)^{1/2} = (x^T x)^{1/2}\). An equilibrium \(x^*\) is *asymptotically stable* if it is stable
and \( \delta \) can be chosen such that

\[
\| x(0) - x^* \| < \delta \quad \Rightarrow \quad \lim_{t \to \infty} \| x(t) - x^* \| = 0.
\] (4.7)

The equilibrium \( x^* \) is \emph{globally asymptotically stable} if it is stable and \( x(t) \) approaches \( x^* \) as \( t \to \infty \) for any \( x(0) \).

### 4.2 Existence, Uniqueness, and Stability of Equilibrium

This section presents some general properties of the equilibria of (4.1) and (4.2). Without further declaration, in this section the neuronal activation function \( f_i(\cdot) \) can be any function satisfying the following condition.

**Condition 4.1** The neuronal activation \( f_i(u) \), where \( i = 1, \ldots, n \), is continuous and nonnegative for any \( u \in (-\infty, +\infty) \). Furthermore, \( f_i(u) \) has continuous first derivative, and \( \dot{f}_i(u) \) is nonnegative for any \( u \in (-\infty, +\infty) \).

#### 4.2.1 Existence of Equilibrium

**Proposition 4.1** The network of (4.1) and (4.2) has at least one equilibrium.

**Proof:** Denote \( h_i(x) = d_i - \sum_{k \neq i} v_k f_k(x_k) \) and \( h(x) = [h_1(x), \ldots, h_n(x)]^T \). Consider a compact convex set \( D = [d_1 - \sum_{k \neq 1} v_k f_k(d_k), d_1] \times \cdots \times [d_n - \sum_{k \neq n} v_k f_k(d_k), d_n] \). For any \( x \in D \), \( h_i(x) \) is no greater than \( d_i \) due to \( f_k(x_k) \geq 0 \) for any \( k \), and \( h_i(x) \) is no less than \( d_i - \sum_{k \neq i} v_k f_k(d_k) \) due to \( f_k(x_k) \leq f_k(d_k) \) for any \( k \) (this is because \( f_k(\cdot) \) is monotone nondecreasing and \( x_k \leq d_k \)). Therefore, \( h(x) \in D \) for any \( x \in D \). According to Brouwer Fixed Point Theorem [82], \( x = h(x) \) has a fixed point in \( D \), which implies that the network of (4.1) and (4.2) has at least one equilibrium. Q.E.D.

It can be seen from the proof, Proposition 4.1 holds for any activation functions that are continuous, nonnegative, and monotone nondecreasing, not just restricted to the activation functions described in Condition 4.1.
4.2.2 Isolated Equilibria and Uniqueness of Equilibrium

Except for assertion of equilibrium existence, Proposition 4.1 does not give any further information about the equilibrium or equilibria of (4.1) and (4.2). In the following, Propositions 4.2 and 4.3 address more detailed properties of the system equilibria. Before showing those propositions, we first present an intermediate result, which will be used in the proofs of the two propositions.

**Lemma 4.1** For

\[
A = \begin{bmatrix}
1 & a_2 & \cdots & a_{n-1} & a_n \\
 a_1 & 1 & \cdots & a_{n-1} & a_n \\
 \vdots & \vdots & \ddots & \vdots & \vdots \\
 a_1 & a_2 & \cdots & 1 & a_n \\
 a_1 & a_2 & \cdots & a_{n-1} & 1
\end{bmatrix}
\]

where \( n \geq 2 \), the determinant of \( A \) equals

\[
\det A = \prod_{k=1}^{n} (1 - a_k) + \sum_{i=1}^{n} a_i \prod_{k \neq i} (1 - a_k).
\] (4.8)

**Proof:** It can be easily verified that (4.8) holds for \( n = 2 \) and \( n = 3 \). In the following, we consider \( n \geq 4 \).

Subtracting the first row from the other rows of \( A \), we get a new matrix

\[
A_1 = \begin{bmatrix}
1 & a_2 & a_3 & \cdots & a_{n-2} & a_{n-1} & a_n \\
 a_1 - 1 & 1 - a_2 & 0 & \cdots & 0 & 0 & 0 \\
 a_1 - 1 & 1 - a_3 & 0 & \cdots & 0 & 0 & 0 \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
a_1 - 1 & 0 & 0 & \cdots & 1 - a_{n-2} & 0 & 0 \\
a_1 - 1 & 0 & 0 & \cdots & 0 & 1 - a_{n-1} & 0 \\
a_1 - 1 & 0 & 0 & \cdots & 0 & 0 & 1 - a_n
\end{bmatrix}.
\]

According to the property of the determinant of a matrix [75], \( \det A \) equals \( \det A_1 \).

Further using the Laplace expansion by minors along the first row of \( A_1 \) [75], we have

\[
\det A = \det A_1 = (-1)^{1+1} \begin{bmatrix}
1 - a_2 & 0 & \cdots & 0 & 0 \\
0 & 1 - a_3 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 - a_{n-1} & 0 \\
0 & 0 & \cdots & 0 & 1 - a_n
\end{bmatrix}
\]
\[ +(-1)^{1+2}a_2 \begin{vmatrix} a_1 - 1 & 0 & \cdots & 0 & 0 \\ a_1 - 1 & 1 - a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 - 1 & 0 & \cdots & 1 - a_{n-1} & 0 \\ a_1 - 1 & 0 & \cdots & 0 & 1 - a_n \\ \end{vmatrix} \]

\[ + \sum_{i=3}^{n-1} (-1)^{1+i}a_i \begin{vmatrix} a_1 - 1 & 1 - a_2 & \cdots & 0 & 0 \\ a_1 - 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 - 1 & 0 & \cdots & 0 & 0 \\ a_1 - 1 & 0 & \cdots & 0 & 1 - a_{i+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 - 1 & 0 & \cdots & 0 & 0 \\ \end{vmatrix} + (-1)^{1+n}a_n \begin{vmatrix} a_1 - 1 & 1 - a_2 & \cdots & 0 & 0 \\ a_1 - 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 - 1 & 0 & \cdots & 0 & 0 \\ a_1 - 1 & 0 & \cdots & 0 & 1 - a_{n-1} \\ \end{vmatrix} \]

For each determinant except the first one in the above equation, we use the Laplace expansion by minors along the first column of the determinant. Then we have

\[
\det A = \prod_{k=2}^{n} (1 - a_k) + (-1)a_2(1 - a_1 - 1) \prod_{k=3}^{n} (1 - a_k) + \sum_{i=3}^{n-1} (-1)^{1+i}a_i(-1)^{1+i-1}(1 - a_i - 1) \prod_{k=2}^{i-1} (1 - a_k) \prod_{k=i+1}^{n} (1 - a_k) + (-1)^{1+n}a_n(-1)^{1+n-1}(1 - a_1 - 1) \prod_{k=2}^{n-1} (1 - a_k) \]

\[
= \prod_{k=2}^{n} (1 - a_k) + \sum_{i=2}^{n} a_i \prod_{k \neq i}^{n} (1 - a_k) \]

\[
= \prod_{k=1}^{n} (1 - a_k) + \sum_{i=1}^{n} a_i \prod_{k \neq i}^{n} (1 - a_k). \]

Q.E.D.

**Proposition 4.2** Let \( x^* \) be an equilibrium of (4.1) and (4.2). If

\[
\prod_{k=1}^{n} [1 - v_k \dot{f}_k(x_k)] + \sum_{i=1}^{n} v_i \dot{f}_i(x_i^*) \prod_{k \neq i}^{n} [1 - v_k \dot{f}_k(x_k^*)] \neq 0, \tag{4.9}
\]

71
then \( x^* \) is an isolated equilibrium.

**Proof:** Since \( \dot{f}_i(x_i) \) is continuous for any \( i = 1, \ldots, n \), Inequality (4.9) implies that there exists an \( \epsilon > 0 \) and an open set \( S(x^*, \epsilon) \equiv \{ x \mid |x_i - x_i^*| < \epsilon \text{ for any } i = 1, \ldots, n \} \) such that \( \prod_{k=1}^n [1 - v_k \dot{f}_k(x_k)] + \sum_{i=1}^n v_i \dot{f}_i(x_i) \prod_{k \neq i}^n [1 - v_k \dot{f}_k(x_k)] \) is not equal to 0 for any \( x \in S(x^*, \epsilon) \).

Next we prove that \( x^* \) is the only equilibrium in \( S(x^*, \epsilon) \). Assume that there exists another equilibrium \( x' \in S(x^*, \epsilon) \). Apparently, \( x' \) satisfies \( x'_i = d_i - \sum_{k \neq i} v_k f_k(x'_k) \) for any \( i = 1, \ldots, n \). Let \( u = x' - x^* \), and substitute \( x' = x^* + u \) into the above equilibrium equation for \( x' \): \( x_i^* + u_i = d_i - \sum_{k \neq i} v_k f_k(x_k^* + u_k) \). According to the mean value theorem [82] and the fact that \( \dot{f} \) is continuous, there exist \( \eta_i, i = 1, \ldots, n \), such that \( f_i(x_i^* + u_i) = f_i(x_i^*) + \dot{f}_i(x_i^* + \eta_i u_i) u_i \) where \( 0 \leq \eta_i \leq 1 \). Therefore, we have \( x_i^* + u_i = d_i - \sum_{k \neq i} v_k [f_k(x_k^*) + \dot{f}_k(x_k^* + \eta_k u_k) u_k] \), which becomes \( u_i = -\sum_{k \neq i} v_k \dot{f}_k(x_k^* + \eta_k u_k) u_k^* \), \( i = 1, \ldots, n \). Denote \( a_i = v_i \dot{f}_i(x_i^* + \eta_i u_i) \), and write the above equations in matrix form \( Au = 0 \) where

\[
A = \begin{bmatrix}
1 & a_2 & \cdots & a_{n-1} & a_n \\
a_1 & 1 & \cdots & a_{n-1} & a_n \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_1 & a_2 & \cdots & 1 & a_n \\
a_1 & a_2 & \cdots & a_{n-1} & 1
\end{bmatrix}.
\]

Since \( x' = x^* + u \in S(x^*, \epsilon) \) and \( 0 \leq \eta_i \leq 1 \) for any \( i = 1, \ldots, n \), we have \((x_1^* + \eta u_1, \ldots, x_n^* + \eta u_n) \in S(x^*, \epsilon) \). Thus \( \prod_{k=1}^n (1 - a_k) + \sum_{i=1}^n a_i \prod_{k \neq i}^n (1 - a_k) \neq 0 \). According to Lemma 4.1, \( \det A \) is not equal to 0. Hence \( Au = 0 \) has only one solution, i.e., \( u = 0 \), which implies \( x' = x^* \). So there is only one equilibrium in the open set \( S(x^*, \epsilon) \) containing \( x^* \). In other words, \( x^* \) is an isolated equilibrium. Q.E.D.

**Corollary 4.1** Let \( x^* \) be an equilibrium of (4.1) and (4.2). If

\[
v_i \dot{f}_i(x_i^*) < 1 \text{ for any } i = 1, \ldots, n, \tag{4.10}
\]

or

\[
v_i \dot{f}_i(x_i^*) > 1 \text{ for any } i = 1, \ldots, n, \tag{4.11}
\]

then \( x^* \) is an isolated equilibrium.

**Proof:** We only need to show that Condition (4.10) or (4.11) implies (4.9).
If Condition (4.10) is satisfied (note that \(v_i \hat{f}_i(x_i^*)\) is also nonnegative for any \(i = 1, \ldots, n\), then we have \(\prod_{k=1}^{n} [1 - v_k \hat{f}_k(x_k^*)] + \sum_{i=1}^{n} v_i \hat{f}_i(x_i^*) \prod_{k \neq i} [1 - v_k \hat{f}_k(x_k^*)] > 0\), so Inequality (4.9) is satisfied.

If Condition (4.11) is satisfied, then \(\frac{v_i \hat{f}_i(x_i^*)}{1 - v_i \hat{f}_i(x_i^*)}\) is less than \(-1\) for any \(i = 1, \ldots, n\). Hence

\[
1 + \frac{v_1 \hat{f}_1(x_1^*)}{1 - v_1 \hat{f}_1(x_1^*)} + \cdots + \frac{v_n \hat{f}_n(x_n^*)}{1 - v_n \hat{f}_n(x_n^*)} < 0,
\]

and thus

\[
\prod_{k=1}^{n} [1 - v_k \hat{f}_k(x_k^*)] + \sum_{i=1}^{n} v_i \hat{f}_i(x_i^*) \prod_{k \neq i} [1 - v_k \hat{f}_k(x_k^*)] = \{(\prod_{k=1}^{n} [1 - v_k \hat{f}_k(x_k^*)])(1 + \frac{v_1 \hat{f}_1(x_1^*)}{1 - v_1 \hat{f}_1(x_1^*)} + \cdots + \frac{v_n \hat{f}_n(x_n^*)}{1 - v_n \hat{f}_n(x_n^*)}) \neq 0.
\]

So Inequality (4.9) is also satisfied. Q.E.D.

**Proposition 4.3** The network of (4.1) and (4.2) has only one equilibrium if there exists \(M_i\) such that \(\hat{f}_i(x_i) \leq M_i\) for any \(x_i\) and

\[
v_i M_i < 1 \tag{4.12}
\]

for any \(i = 1, \ldots, n\).

**Proof:** Let \(x^*\) be an equilibrium, and assume that there exists another equilibrium \(x'\). Then following similar steps as in the proof of Proposition 4.2, we may prove \(x' = x^*\). Q.E.D.

The result presented in Proposition 4.3 can be extended to more general activation functions, which are not even differentiable. This will be discussed in Proposition 4.5 and Condition 4.2 in the subsection concerning the stability and convergence of an equilibrium.

Proposition 4.3, as well as Proposition 4.5 to be shown in the following subsection, presents a sufficient condition of uniqueness of equilibrium for the network of (4.1) and (4.2). To demonstrate how conservative this condition is, consider a simple scenario where \(v_i\) equals \(v\) and \(M_i\) equals \(M\) for \(i = 1, \ldots, n\), and see what the existing results from general recurrent (Hopfield) network theory can offer us concerning the
uniqueness of equilibrium. Using the results presented in [48, 61], for example, it can be derived that the network of (4.1) and (4.2) has only one equilibrium if

$$(n - 1)vM < 1. \tag{4.13}$$

This condition is the same as what we can get by using the contract mapping theorem [82]. As a comparison, Proposition 4.3 gives the following condition

$$vM < 1,$$

which is much less conservative than (4.13) especially for large $n$.

It will be further demonstrated in an example presented in Fig. 4-3 that the right side of (4.12) can not be relaxed to any number greater than 1 and (4.12) can not even be relaxed to include equality.

### 4.2.3 Stability of Equilibrium

**Proposition 4.4** Let $x^*$ be an equilibrium of (4.1) and (4.2). If $v_i\hat{f}_i(x_i^*) < 1$ for any $i = 1, ..., n$, then $x^*$ is an asymptotically stable equilibrium.

**Proof:** Given $x$, denote $u = x - x^*$ and $u_i = x_i - x_i^*$ for $i = 1, ..., n$. Furthermore, denote $g_i(u_i) = f_i(x_i) - f_i(x_i^*) = f_i(x_i^* + u_i) - f_i(x_i^*)$ and

$$M_v = \frac{1}{2} + \frac{\max\{v_i\hat{f}_i(x_i^*)\}}{2}.$$

Apparently, $M_v$ is strictly greater than $v_i\hat{f}_i(x_i^*)$ for any $i = 1, ..., n$ and strictly less than 1, since $v_i\hat{f}_i(x_i^*)$ is less than 1.

Because $\hat{f}_i(x_i)$ is continuous and $M_v$ is strictly greater than $\max\{v_i\hat{f}_i(x_i^*)\}$, there exists an $\epsilon > 0$ and a domain containing $x^*$, $S(x^*, \epsilon) \equiv \{x \mid |x_i - x_i^*| < \epsilon \text{ for any } i = 1, ..., n\}$, such that $v_i\hat{f}_i(x_i)$ is less than $M_v$ for any $x \in S(x^*, \epsilon)$ and for any $i = 1, ..., n$. Correspondingly, denote $D = \{u \mid u + x^* \in S(x^*, \epsilon)\}$. Therefore,

$$v_i\hat{g}_i(u_i) < M_v \text{ for any } u \in D. \tag{4.14}$$
We use a Lure-type Lyapunov function \([48, 97]\)

\[
V(u) = \frac{1}{2} \sum_{i=1}^{n} cu_i^2 + \sum_{i=1}^{n} \int_{0}^{u_i} v_i g_i(s) ds,
\]

(4.15)

where \(c\) is a positive constant satisfying

\[
c < \frac{4}{n(1 - M_v)}.
\]

(4.16)

Obviously, \(V(0)\) equals 0 and \(V(u)\) is greater than 0 for any \(u \in D\) except for \(u = 0\).

In the following, we want to show \(dV/dt < 0\) for any \(u \in D\) and \(u \neq 0\). If this is true, then we can claim that \(u = 0\), i.e. \(x = x^*\), is asymptotically stable, according to [97] (Theorem 3.1, Page 100).

Let us first calculate \(\tau \frac{du_i}{dt}\). \(\tau \frac{du_i}{dt} = \tau \frac{dx_i}{dt} = -x_i - \sum_{k \neq i} v_k f_k(x_k) + d_i = -x_i^* + u_i - \sum_{k \neq i} v_k [f_k(x_i^*) + g_k(u(k))] + d_i = -u_i - \sum_{k \neq i} v_k g_k(u(k)).\) Then we have

\[
\tau \frac{dV}{dt} = \sum_{i=1}^{n} cu_i \left( \tau \frac{du_i}{dt} \right) + \sum_{i=1}^{n} v_i g_i(u_i) \left( \tau \frac{du_i}{dt} \right)
\]

\[
= \sum_{i=1}^{n} cu_i [-u_i - \sum_{k \neq i} v_k g_k(u(k))] + \sum_{i=1}^{n} v_i g_i(u_i) [-u_i - \sum_{k \neq i} v_k g_k(u(k))]
\]

\[
= \sum_{i=1}^{n} [cu_i + v_i g_i(u_i)] [-u_i - \sum_{k \neq i} v_k g_k(u_k)]
\]

\[
= \sum_{i=1}^{n} [cu_i + v_i g_i(u_i)] [-u_i - v_i g_i(u_i)] - \sum_{k=1}^{n} v_k g_k(u_k)
\]

\[
= - \sum_{i=1}^{n} [cu_i + v_i g_i(u_i)] [u_i - v_i g_i(u_i)] - \sum_{k=1}^{n} [cu_i + v_i g_i(u_i)] \sum_{k=1}^{n} v_k g_k(u_k).
\]

Note that

\[
\sum_{i=1}^{n} [cu_i + v_i g_i(u_i)] \sum_{k=1}^{n} v_k g_k(u_k)
\]

\[
= \sum_{i=1}^{n} cu_i \sum_{k=1}^{n} v_k g_k(u_k) + \left( \sum_{k=1}^{n} v_k g_k(u_k) \right)^2
\]

\[
= \sum_{i=1}^{n} \left\{ cu_i \sum_{k=1}^{n} v_k g_k(u_k) + \frac{1}{n} \left( \sum_{k=1}^{n} v_k g_k(u_k) \right)^2 \right\}
\]

\[
= \sum_{i=1}^{n} \left\{ \frac{c u_i^2}{4} + cu_i \sum_{k=1}^{n} v_k g_k(u_k) + \frac{1}{n} \left( \sum_{k=1}^{n} v_k g_k(u_k) \right)^2 \right\} - \sum_{i=1}^{n} \frac{c u_i^2}{4}
\]

\[
= \sum_{i=1}^{n} \left\{ \frac{c u_i^2}{2} + \frac{1}{\sqrt{n}} \left[ \sum_{k=1}^{n} v_k g_k(u_k) \right] \right\}^2 - \sum_{i=1}^{n} \frac{c u_i^2}{4}
\]

\[
\geq - \sum_{i=1}^{n} \frac{c u_i^2}{4}.
\]

Therefore,

\[
\tau \frac{dV}{dt} \leq - \sum_{i=1}^{n} [cu_i + v_i g_i(u_i)] [u_i - v_i g_i(u_i)] + \sum_{i=1}^{n} \frac{c u_i^2}{4}.
\]
According to (4.14) and the definition of $M_v$, for any $u \in D$ we have (i) $0 \leq v_i g_i(u_i) \leq M_v u_i$ when $u_i > 0$ and (ii) $M_v u_i \leq v_i g_i(u_i) \leq 0$ when $u_i < 0$. Since $M_v < 1$, we have $0 \leq v_i g_i(u_i) < u_i$ for $u_i > 0$ and $u_i < v_i g_i(u_i) \leq 0$ for $u_i < 0$. Thus $v_i g_i(u_i)[u_i - v_i g_i(u_i)]$ is no less than 0 for any $u \in D$. Therefore,

$$\frac{dV}{dt} \leq - \sum_{i=1}^{n} [c u_i + v_i g_i(u_i)]|u_i - v_i g_i(u_i)| + \sum_{i=1}^{n} \frac{c_n}{4} u_i^2$$

$$= - \sum_{i=1}^{n} c u_i [u_i - v_i g_i(u_i)] - \sum_{i=1}^{n} v_i g_i(u_i)[u_i - v_i g_i(u_i)] + \sum_{i=1}^{n} \frac{c_n}{4} u_i^2$$

$$\leq - \sum_{i=1}^{n} c u_i (u_i - M_v u_i) + \sum_{i=1}^{n} \frac{c_n}{4} u_i^2$$

$$= - \sum_{i=1}^{n} [c (1 - M_v) - \frac{c_n}{4}] u_i^2.$$

According to (4.16), $c (1 - M_v) - \frac{c_n}{4}$ is strictly greater than 0, so $\frac{dV}{dt}$ is strictly negative for any $u \in D$ except for $u = 0$. Q.E.D.

Using the Lure-type Lyapunov function as constructed in the proof of Proposition 4.4, we may extend the conclusion of uniqueness of equilibrium in Proposition 4.3 to more general activation functions. We may relax the constraint of “differentiable $f_i(u)$ with nonnegative $\dot{f}_i(u)$” in Condition 4.1 to the following one.

**Condition 4.2** The neuronal activation $f_i(u)$, where $i = 1, \ldots, n$, is nonnegative for any $u \in (-\infty, +\infty)$. In addition, $f_i(u)$ is globally Lipschitz continuous, i.e.,

$$0 \leq \frac{f_i(u_1) - f_i(u_2)}{u_1 - u_2} \leq M_i,$$

for any two different $u_1, u_2$.

It is obvious that a globally Lipschitz continuous activation function may be non-differentiable and unbounded. Then we have the following proposition.

**Proposition 4.5** Let $f_i(\cdot)$ satisfy Condition 4.2 and the constant $M_i$ in (4.17) is constrained by $v_i M_i < 1$ for any $i = 1, \ldots, n$. Then the network of (4.1) and (4.2) has a unique equilibrium, which is globally asymptotically stable.

**Proof:** Following the proof of Proposition 4.1, the network of (4.1) and (4.2) with activation functions described as above has at least one equilibrium. Let $x^*$ be such
an equilibrium. Denote \( u_i = x_i - x_i^* \) and \( g_i(u_i) = f_i(x_i) - f_i(x_i^*) = f_i(x_i^* + u_i) - f_i(x_i^*) \) for \( i = 1, \ldots, n \).

Define an energy function \( V(u) \) in the form of \((4.15)\), where \( c \) is a positive constant satisfying \( c < \min_i \left\{ \frac{4}{n} (1 - v_i M_i) \right\} \). Obviously, \( V(0) \) equals 0, \( V(u) \) is greater than 0 for any \( u \neq 0 \), and \( V(u) \) approaches positive infinity as \( \| u \| \to \infty \). Furthermore, following the similar steps as in the proof of Proposition 4.4, we may show that \( \frac{dV}{dt} \) is strictly negative for any \( u \neq 0 \). Then we can claim that \( u = 0 \), i.e., \( x = x^* \), is globally asymptotically stable, according to [97] (Theorem 3.2, Page 110). Obviously, \( x^* \) has to be the unique equilibrium of the network of \((4.1)\) and \((4.2)\). \textbf{Q.E.D.}

### 4.2.4 Examples of Equilibria

Let us see some examples of equilibria. Consider two-neuron networks, where the neuronal activation functions of both neurons in a network are the same, i.e., \( f_1(\cdot) = f_2(\cdot) \equiv f(\cdot) \), and the lateral inhibitory connections between the two neurons are symmetric, i.e., \( v_1 = v_2 \equiv v \).

Fig. 4-1 presents an example of a two-neuron network with unique equilibrium. The activation function considered here is a sigmoidal function, specifically a logistic function of the following form (see the left picture of Fig. 4-1):

\[
f(u) = \frac{1}{1 + e^{-\frac{u}{a}}}, \tag{4.18}
\]

where \( b \) is a translation factor, called \textit{threshold} or \textit{bias}, and \( a \) is a dilation factor: The smaller \( a \) is, the steeper the logistic function appears to be. The right picture of Fig. 4-1 presents several trajectories, each of which shows how the dynamics of the two-neuron network evolves from a specific initial state. In this example, \( d_1 = 1.5, d_2 = 1, a = 1/3, b = 1, \) and \( v = 1 \). It can be calculated that \( \max_u \hat{f}(u) \) equals \( \frac{1}{4a} \). Then \( v \max_u \hat{f}(u) \) equals \( 3/4 \), less than 1. According to Propositions 4.3 and 4.5, the network should have a unique equilibrium, which in addition should be globally asymptotically stable. This is consistent with what is shown by Fig. 4-1 (Right).

Fig. 4-2 presents two examples with multiple equilibria. The activation functions considered in both examples are logistic functions of the form of \((4.18)\) with \( b = 0.5 \), and the value of \( v \max_u \hat{f}(u) \) equals \( 7/4 > 1 \). The network inputs are \( d_1 = d_2 = 1 \).
Figure 4-1: Left: A sigmoidal function as neuronal activation function. Right: Dynamics of a two-neuron network, where $v \max_u \hat{f}(u) = 3/4 < 1$. The network has only one equilibrium, which is globally asymptotically stable.

Figure 4-2: Dynamics of a two-neuron network, where $v \max_u \hat{f}(u) = 7/4 > 1$. The network has three equilibria, two of which are asymptotically stable and the other one unstable. Left: $v = 1$ and $\max_u \hat{f}(u) = 7/4$. Right: $v = 2$ and $\max_u \hat{f}(u) = 7/8$. 
for both examples. In Fig. 4-2 (Left), $a$ and $v$ are set $1/7$ and $1$, respectively. The network has three equilibria $(0.96, 0.038), (0.038, 0.96)$, and $(0.50, 0.50)$, the first two of which are asymptotically stable and the other unstable. At the unstable equilibria, $v \dot{f}(0.50)$ is approximately $1.75$, greater than $1$. At the two stable equilibria, $v \dot{f}(0.038)$ and $v \dot{f}(0.96)$ approximately equal $0.26 < 1$. This is consistent with Proposition 4.4: Proposition 4.4 predicts that $(0.96, 0.038)$ and $(0.038, 0.96)$ are asymptotically stable based on $v \dot{f} < 1$ at the equilibria. This example is also consistent with Corollary 4.1, which predicts that all the three equilibria are isolated. Similarly, in the example presented in Fig. 4-2 (Right), where $a = 2/7$ and $v = 2$, the network has three equilibria $(0.97, -0.67), (-0.67, 0.97)$, and $(0.31, 0.31)$, the first two being asymptotically stable and the other unstable. At the unstable equilibria, $v \dot{f}(0.31)$ is approximately $1.58 (> 1)$, while at the two stable equilibria, $v \dot{f}(-0.67)$ and $v \dot{f}(0.97)$ approximately equal $0.11$ and $0.95$, respectively. This example is also consistent with Proposition 4.4 and Corollary 4.1.

Fig. 4-3 shows an example with an infinite number of stable equilibria. The activation function considered here is a smoothed linear threshold function (see the
left picture of Fig. 4-3):

\[ f(u) = \begin{cases} 
  1/(1 + e^{-4u}) & \text{if } u \leq 0, \\
  u + \frac{1}{2} & \text{if } u > 0.
\end{cases} \]

Note that \( \dot{f}(u) \) equals 1 for any \( u \geq 0 \). The network inputs are \( d_1 = d_2 = 1 \). With \( v = 1 \), \( v_{\max} \dot{f}(u) \) equals 1. It can been seen from the right picture of Fig. 4-3 that the network state trajectories, starting from various initial states, converge to a line segment between \((0, 0.5)\) and \((0.5, 0)\). Any point on that line segment is an equilibrium, which is stable but not asymptotically stable. Let \( x^* \) be an equilibrium on the line segment between \((0, 0.5)\) and \((0.5, 0)\). Obviously, \( x^* \) is not an isolated equilibrium, since there are an infinite number of equilibria in any neighborhood of \( x^* \) no matter how small the neighborhood is. And it can be seen that \( v \dot{f}(x_1^*) = v \dot{f}(x_2^*) = 1 \), which fails to satisfy (4.9). This is consistent with Proposition 4.2. This example further demonstrates that the conditions presented in Propositions 4.3 and 4.5 are tight: The right side of (4.12) can not be relaxed to any number greater than 1 and (4.12) can not even be relaxed to include equality.

### 4.3 Winner-take-all Competition

In the study of winner-take-all competition, we want to be “fair” with each individual neuron: We assume that the network is symmetric for each neuron, i.e., all neurons have the same activation function and same synaptic strength of lateral inhibitory connection between neurons. Specifically, let \( f_i(\cdot) = f(\cdot) \) and \( v_i = v \) for \( i = 1, \ldots, n \). Then (4.1) becomes

\[
\frac{dx_i}{dt} = -x_i - \sum_{k \neq i} v f(x_k) + d_i, \quad i = 1, \ldots, n,
\]

where \( f(\cdot) \) satisfies the following condition (similar to Condition 4.2).

**Condition 4.3** The neuronal activation \( f(u) \) is nonnegative and globally Lipschitz continuous, i.e.,

\[
0 \leq \frac{f(u_1) - f(u_2)}{u_1 - u_2} \leq M,
\]

for any two different \( u_1, u_2 \).
4.3.1 Order Preserving Equilibrium

An equilibrium of (4.19), $x^*$, is said to be order preserving with respect to the network inputs, if $x_i^*$ is less than $x_j^*$ whenever $d_i$ is less than $d_j$, where $i,j \in \{1,...,n\}$.

Proposition 4.6 The network of (4.19) always has an equilibrium $x^*$ that is order preserving with respect to the network inputs.

Proof: Without loss of generality, let $d_1 \leq d_2 \leq \cdots \leq d_n$.

Denote $h_i(x) = d_i - \sum_{k \neq i} v f(x_k)$ and $h(x) = [h_1(x),...,h_n(x)]^T$. Consider a compact convex set $D = \{x | d_i - \sum_{k \neq i} v f(d_k) \leq x_i \leq d_i, \ i = 1,...,n, \ and \ x_1 \leq \cdots \leq x_n \}$. Similar to the proof of Proposition 4.1, it can be shown that for any $x \in D$, $h_i(x)$ is no greater than $d_i$ and no less than $d_i - \sum_{k \neq i} v f(d_k)$. Furthermore, for any $i < j$ (and thus $d_i \leq d_j$), we have $h_i(x) - h_j(x) = d_i - d_j + v f(x_i) - v f(x_j) \leq 0$, because $d_i \leq d_j$ and $f(x_i) \leq f(x_j)$ (since $f$ is monotone nondecreasing and $x_i$ is no greater than $x_j$ due to $i < j$ and $x \in D$). Therefore, $h(x) \in D$ for any $x \in D$. According to Brouwer Fixed Point Theorem [82], $x = h(x)$ has a fixed point in $D$, which implies that (4.19) has at least one equilibrium, say $x^*$, such that $x_1^* \leq x_2^* \leq \cdots \leq x_n^*$.

For any $d_i < d_j$, we have $x_i^* \leq x_j^*$ and thus $f(x_i^*) \leq f(x_j^*)$. Therefore, $x_i^* - x_j^* = d_i - d_j + v f(x_i^*) - v f(x_j^*) \leq d_i - d_j < 0$, so $x_i^*$ is strictly less than $x_j^*$. Hence we proved that $x^*$ is order preserving with respect to the network inputs. Q.E.D.

It can be seen from the proof of Proposition 4.6, (4.19) is guaranteed to have an order preserving equilibrium as long as $f(\cdot)$ is continuous, nonnegative, and monotone nondecreasing. In other words, Proposition 4.6 may also apply to some situation where $f(\cdot)$ is not globally Lipschitz continuous.

Proposition 4.6 does not mention the stability of an order preserving equilibrium $x^*$. From the example presented in Fig. 4-3, it can be seen that an order preserving equilibrium may not necessarily be asymptotically stable. However, following Proposition 4.4, $x^*$ is asymptotically stable if $f(\cdot)$ satisfies Condition 4.1 and $v \dot{f}(x_i^*) < 1$ for any $i = 1,...,n$. 

81
Proposition 4.7  The network of (4.19) has only one equilibrium, \( x^* \), if \( f(\cdot) \) satisfies Condition 4.3 and the constant \( M \) in (4.20) is constrained by

\[
vM < 1. \tag{4.21}
\]

This equilibrium is globally asymptotically stable. Furthermore, \( x^* \) is order preserving with respect to the network inputs, and \( x_i^* \) equals \( x_j^* \) whenever \( d_i \) equals \( d_j \).

**Proof:** First show that an equilibrium of (4.19), say \( x^* \), has to be order preserving. Prove this by contradiction. If for some \( d_i < d_j \) we have \( x_i^* \geq x_j^* \), then we can get

\[
0 \leq x_i^* - x_j^* = d_i - d_j + vf(x_i^*) - vf(x_j^*) \leq d_i - d_j + vM(x_i^* - x_j^*),
\]

which implies

\[
0 \leq (1 - vM)(x_i^* - x_j^*) \leq d_i - d_j < 0,
\]

causing contradiction. Therefore, \( x^* \) has to be order preserving. Similarly, we may prove \( x_i^* = x_j^* \) if \( d_i = d_j \).

Next show that (4.19) has unique equilibrium and the equilibrium is globally asymptotically stable. This directly follows Proposition 4.5. Here we want to show an alternative proof of the uniqueness of equilibrium without using the Lure-type Lyapunov function. This proof is via contradiction. Assume that the network has more than one equilibria, and let \( x^{(1)} \neq x^{(2)} \) be two of the equilibria. Without loss of generality, assume \( d_1 \leq d_2 \leq \cdots \leq d_n \). According to the above argument, we have \( x_1^{(1)} \leq \cdots \leq x_n^{(1)} \) and \( x_1^{(2)} \leq \cdots \leq x_n^{(2)} \). Let \( i \) be the smallest integer such that \( x_i^{(1)} \neq x_i^{(2)} \), which implies \( x_j^{(1)} = x_j^{(2)} \) for any \( j < i \) if \( i > 1 \). Without loss of generality, let \( x_i^{(1)} < x_i^{(2)} \). Thus \( x_i^{(1)} = d_i - \sum_{k \neq i} vf(x_k^{(1)}) < x_i^{(2)} = d_i - \sum_{k \neq i} vf(x_k^{(2)}) \), and therefore \( \sum_{k \neq i} f(x_k^{(1)}) > \sum_{k \neq i} f(x_k^{(2)}) \). Then there exists at least one \( k > i \) such that \( f(x_k^{(1)}) > f(x_k^{(2)}) \). Obviously, \( x_k^{(1)} > x_k^{(2)} \). Now we have \( x_1^{(1)} < x_1^{(2)} \leq x_k^{(2)} < x_k^{(1)} \).

Consider simple calculation as follows:

\[
x_k^{(1)} - x_k^{(2)} + x_i^{(2)} - x_i^{(1)} = (d_k - \sum_{j \neq k} vf(x_j^{(1)})) - (d_k - \sum_{j \neq k} vf(x_j^{(2)}))
\]
\[
+ (d_i - \sum_{j \neq i} vf(x_j^{(2)})) - (d_i - \sum_{j \neq i} vf(x_j^{(1)}))
\]
\[
= -vf(x_i^{(1)}) + vf(x_i^{(2)}) - vf(x_k^{(2)}) + vf(x_k^{(1)})
\]
\[
= v(f(x_i^{(2)}) - f(x_i^{(1)})) + v(f(x_k^{(1)}) - f(x_k^{(2)}))
\]
\[
\leq vM(x_i^{(2)} - x_i^{(1)}) + vM(x_k^{(1)} - x_k^{(2)})
\]
\[
< x_i^{(2)} - x_i^{(1)} + x_k^{(1)} - x_k^{(2)}.
\]

82
which leads to contradiction and thus completes the proof of uniqueness. Q.E.D.

4.3.2 Increased Differences Between Neuronal Activities

**Proposition 4.8** Let $x^*$ be an order preserving equilibrium of (4.19). If $f(\cdot)$ is strictly monotone increasing, then for any $d_i < d_j$ we have

$$x_j^* - x_i^* > d_j - d_i > 0,$$

(4.22)

i.e., expressed in another form,

$$d_i - x_i^* > d_j - x_j^* \geq 0.$$

(4.23)

**Proof:** Since $x^*$ is an order preserving equilibrium, we have $x_i^* < x_j^*$ for any $d_i < d_j$ and thus $f(x_j^*) < f(x_j^*)$ because $f(\cdot)$ is strictly monotone increasing. Therefore, $x_j^* - x_i^* = d_j - d_i + v f(x_j^*) - v f(x_i^*) > d_j - d_i$. So we get (4.22). Expressing this inequality in another form, we have $d_i - x_i^* > d_j - x_j^*$. Inequality (4.5) guarantees $d_j - x_j^* \geq 0$. Then we get (4.23). Q.E.D.

The significance of Proposition 4.8 is that it reveals the underlying mechanism of winner-take-all competition in a recurrent network with lateral inhibition. Inequality (4.22) shows the expansion of neuronal postsynaptic membrane potentials in the presence of lateral inhibition: The lateral inhibition between every pair of neurons plays a role of increasing the differences between neuronal potentials at any order preserving equilibrium (without the lateral inhibition, $x_i^*$ equals $d_i$ for any $i$ and thus $x_j^* - x_i^* = d_j - d_i$). Inequality (4.23) shows the asymmetric effect of lateral inhibition: Although the lateral inhibition decreases the postsynaptic membrane potential of each neuron, such effect is not symmetric; at an order preserving equilibrium, the potential of a neuron with larger input is decreased less than that of a neuron with smaller input. This is demonstrated in Fig. 4-4.

For our particular interest in the winner-take-all networks, we now restrict our attention to a family of sigmoidal activation functions, with which the networks may effectively implement winner-take-all competition among neurons. The activation function $f(u)$ considered in the following satisfies Condition 4.4.
Figure 4-4: Distribution of the network inputs versus individual elements of an order preserving equilibrium $x^*$. If $d_i < d_j$, then $d_i - x_i^* > d_j - x_j^*$.

**Condition 4.4** The neuronal activation $f(u)$ is continuous, nonnegative, and strictly monotone increasing: $f(u)$ is continuous; $\dot{f}(u)$ is monotone decreasing on $[b, \infty)$ and monotone increasing on $(-\infty, b]$ this implies that $\dot{f}(u)$ achieves its maximum at $u = b$ and that $f(u)$ is concave on the right side of $u = b$ and convex on the left side of $u = b$.

In the above condition, $b$ is called the bias or membrane potential threshold of a neuron. Neuron $i$ is said to be active at time $t$ if $x_i(t) > b$ and inactive if $x_i(t) \leq b$.

**Proposition 4.9** Consider a network of (4.19) with neuronal activation function satisfying Condition 4.4. Let $d_i$ and $d_j$ be two inputs such that $b < d_i < d_j$, and let $x^*$ be an order preserving equilibrium of (4.19). If neurons $i$ and $j$ are both active at $x^*$, i.e., $b < x_i^* < x_j^*$, then there must be

\[ v \dot{f}(d_j) < 1, \quad (4.24) \]

\[ x_j^* - x_i^* \geq \frac{d_j - d_i}{1 - v f(d_j)}, \quad (4.25) \]

and

\[ f(x_j^*) - f(x_i^*) \geq f(d_j) - f(d_i) + \frac{v \dot{f}(d_j) f(d_i)}{1 - v f(d_j)} (d_j - d_i). \quad (4.26) \]
Proof: First prove (4.24) and (4.25). Since \( \dot{f}(\cdot) \) is continuous, we have

\[
x^*_j - x^*_i = d_j - d_i + v f(x^*_j) - v f(x^*_i) = d_j - d_i + v \dot{f}(\xi)(x^*_j - x^*_i),
\]

where \( \xi \in [x^*_i, x^*_j] \subset (b,d_j) \), according to the mean value theorem [82]. Hence \( \dot{f}(\xi) \geq \dot{f}(d_j) \) (\( \dot{f}(\cdot) \) is monotone decreasing on \([b,\infty)\)). Because \( x^*_i < x^*_j \) and \( d_i < d_j \), there must be \( v \dot{f}(\xi) < 1 \), otherwise we may derive \( x^*_j - x^*_i \geq d_j - d_i + x^*_j - x^*_i \) and thus \( d_j \leq d_i \), causing contradiction. Therefore, we have \( v \dot{f}(d_j) \leq v \dot{f}(\xi) < 1 \), which leads to (4.24). Following (4.27), we further have

\[
x^*_j - x^*_i = \frac{d_j - d_i}{1 - v \dot{f}(\xi)} \geq \frac{d_j - d_i}{1 - v \dot{f}(d_j)},
\]

which leads to (4.25).

Next prove (4.26). According to (4.25), the fact that \( b < x^*_i < x^*_j \leq d_j \), and that \( f(\cdot) \) is strictly monotone increasing and \( \dot{f}(\cdot) \) is monotone decreasing on \([b,\infty)\), we have

\[
f(x^*_j) - f(x^*_i) \geq f(d_j) - f(d_i - (x^*_j - x^*_i)) \geq f(d_j) - f(d_i) - \frac{d_j - d_i}{1 - v \dot{f}(d_j)} = f(d_j) - f(d_i) - \frac{v \dot{f}(d_j)(d_j - d_i)}{1 - v \dot{f}(d_j)} \geq f(d_j) - f(d_i) + \dot{f}(d_i) \frac{v \dot{f}(d_j)(d_j - d_i)}{1 - v \dot{f}(d_j)} = f(d_j) - f(d_i) + \frac{v \dot{f}(d_j) \dot{f}(d_i)}{1 - v \dot{f}(d_j)} (d_j - d_i),
\]

completing the proof. Q.E.D.

Proposition 4.9 indicates that the lateral inhibition increases both the differences between postsynaptic membrane potentials and the differences between outputs of active neurons at an order preserving equilibrium. Fig. 4-5 presents an example showing the outcome of competition among neurons in a nine neuron network under different strength of lateral inhibition \( v \). With appropriate choice of \( v \), the network may present winner-take-all competition among neurons such that there can be at
Figure 4-5: Winner(s)-take-all competition in a network with lateral inhibition. The network has 9 neurons, with logistic function \((4.18)\) as neuronal activation function, where \(a = 1/8\) and \(b = 0.5\). The initial state is \(x(0) = 0\). a: Network inputs \(d = [0.6, 1.0, 0.8, 1.2, 0.7, 1.1, 0.9, 0.4, 0.5]^T\). b: Four active neurons (Neurons 2, 4, 6, and 7) at the equilibrium when \(v = 0.1\). c: Two active neurons (Neurons 4 and 6) at the equilibrium when \(v = 0.5\). d: Single winner, i.e. Neuron 4 as the single active neuron, at the equilibrium when \(v = 1\).

Most a single neuron active at the equilibrium.

Corollary 4.2 Consider a network of \((4.19)\) with neuronal activation function satisfying Condition 4.4. Let \(d_j\) and \(d_i\) be the largest and second largest inputs, respectively, with \(b < d_i < d_j\), and \(x^*\) an order preserving equilibrium of \((4.19)\). If

\[
v f(d_j) \geq 1, \tag{4.28}
\]

or if

\[
v f(d_j) < 1 \text{ and } d_j - b < \frac{d_j - d_i}{1 - v f(d_j)}, \tag{4.29}
\]

then the network has at most one active neuron at \(x^*\).

**Proof:** The proof directly follows Proposition 4.9. Q.E.D.
In the following section, we consider a limiting situation where the neuronal activation functions have infinite gain. This special situation approximates those involving steep neuronal activation functions, and provides us with a more detailed picture of winner-take-all competition.

### 4.4 Dynamics of Winner-take-all in a Limiting Situation

This section investigates the winner-take-all competition under a limiting situation where neurons have discontinuous activation functions. We are primarily interested in the situations with activation functions that possess high-slope nonlinearities. However, rather than studying the case that the slope is high but of finite value, we choose to model the network here as a system of differential equations with discontinuous right hand side (with infinite gain). The advantage of analyzing the ideal discontinuous case, as pointed out by Forti and Nistri [47], is that such analysis can usually demonstrate a clear picture of the salient features of motion, such as the presence of possibility that trajectories be confined for some time on discontinuity surfaces.

Neural networks with discontinuous activation functions frequently arise in both practical and theoretical studies. For example, Hopfield has showed that networks consisting of simple two-state neurons produce content-addressable memory [73]. Kennedy and Chua have used neural networks with diode-like activation functions to solve linear and nonlinear programming problems [94]. Forti and Nistri have studied global convergence of neural networks possessing neurons with infinite gain [47]. In this section, we consider the dynamics of winner-take-all networks with discontinuous neuronal activation and lateral inhibition.

The discontinuous neuronal activation function considered in this section is of the following form

$$\sigma(u - b) = \begin{cases} 
1 & \text{if } u > b, \\
0 & \text{if } u \leq b,
\end{cases} \quad (4.30)$$

where $b$ represents the bias or threshold of a neuron. Let $v_i = v$ for $i = 1, ..., n$. Then
the network of (4.1) and (4.2) becomes

\[ \tau \frac{dx_i}{dt} = -x_i - v \sum_{k \neq i} y_k + d_i, \quad (4.31) \]
\[ y_i = \sigma(x_i - b), \quad (4.32) \]

where \( i = 1, \ldots, n. \)

This system fails to satisfy conventional existence results of differential equation theory. Let us see a simple example:

\[ \frac{du}{dt} = \frac{1}{2} - \sigma(u), \quad u(0) = 0, \quad (4.33) \]

where \( \sigma(u) \) is defined by (4.30). Following the dynamics of (4.33), \( u(t) \) has to keep the value 0 for every \( t > 0 \) (thus \( \frac{du}{dt} = 0 \) for every \( t > 0 \)), and obviously there exists no classical (i.e. almost everywhere) solution to the above scalar equation. A generalization of the concept of solution is thus required.

### 4.4.1 Concepts of Filippov Solutions

A seminal contribution in the analysis of differential equations with discontinuous right hand sides was made by Filippov [46]. He developed a new solution concept for differential equations whose right hand sides were only required to be Lebesgue measurable in the state and time variables. Consider the vector differential equation in the form

\[ \frac{dx}{dt} = f(x, t), \quad (4.34) \]

where \( f : R^n \times R \rightarrow R^n \) is measurable and essentially locally bounded. A vector function \( x(t) \), defined on the interval \([t_0, t_1]\), is called a Filippov solution of (4.34) if it is absolutely continuous and if for almost all \( t \in [t_0, t_1] \) the vector \( \frac{dx(t)}{dt} \) belongs to \( K[f](x, t) \), which is defined as the smallest convex closed set containing the limiting values of vector field \( f(\cdot) \) in progressively smaller neighborhoods of \( x(t) \) in the space of \( x \) except for a set of measure zero [46]. Please see Appendix A for a precise expression of \( K[f](x, t) \) in mathematical notation.

Comparing Filippov’s definition with the usual definition of the solution in the case of continuous right hand side of differential equations, we see that \( K[f](x, t) \)
contains a single point coinciding with \( f(x,t) \) when \( f(x,t) \) is continuous. Thus for this case the solution of (4.34) in the sense of Filippov is the same as the solution in the ordinary sense.

Referring back to the example of (4.33). When \( u \neq 0 \), the differential equation’s right hand side \( f(u) = 0.5 - \sigma(u) \) is continuous, and \( K[f](u) = \{-0.5\} \) for \( u > 0 \) and \( K[f](u) = \{0.5\} \) for \( u < 0 \). At point \( u = 0 \), consider a neighborhood of 0, say \((-\delta, \delta)\). For any \( \delta > 0 \), the values of \( f(u) \) on \((-\delta, \delta)\) are \(-0.5\) and \(0.5\). The values of \( f(u) \) on \((-\delta, \delta) - N\) are still \(-0.5\) and \(0.5\) for any set \( N \) with measure zero, where \((-\delta, \delta) - N\) means the set of points that belong to \((-\delta, \delta)\) but not belong to \(N\). Thus the limiting values of \( f(u) \) on \((-\delta, \delta) - N\) as \( \delta \) approaches 0 are \(-0.5\) and \(0.5\). Therefore, \( K[f](0)\) equals \([-0.5, 0.5]\), which is the smallest convex closed set containing \(-0.5\) and \(0.5\). It is clear that \( u(t) = 0 \) satisfies \( du(t)/dt = 0 \in K[f](0) \), and thus \( u(t) = 0 \) is a Filippov solution to (4.33).

It is important in Filippov’s definition that sets of measure zero are discarded. This technical detail allows solutions to be defined at points even where the vector field itself is not defined, such as at the interface of two regions in a piecewise defined vector field.

Existence of Filippov solutions of (4.31) and (4.32) is guaranteed by the Existence Theorem shown in Appendix A. In the rest of this section, any trajectories of \( x(t) \) of (4.31) and (4.32) are meant to be Filippov solutions of (4.31) and (4.32).

A vector \( x^* \) is said to be an equilibrium of (4.31) and (4.32) if \( x(t) = x^* \) is a Filippov solution of the system. Considering a system subject to (4.34), an equilibrium \( x^* \) is a vector satisfying \( 0 \in K[f](x^*, t) \). Note that this definition is consistent with the previous definition of equilibrium in the situation where \( f(x,t) \) is continuous.

### 4.4.2 Winner-take-all in a Limiting Situation

We show that lateral inhibition among neurons can lead to winner-take-all competition in the network: Only a single neuron is active in the network at an equilibrium. Denote \( d_{\text{max}} = \max\{d_1, ..., d_n\} \).

**Proposition 4.10** Let \( x(t) \) be a state trajectory of (4.31) and (4.32) originating from \( x(0) \). Then for any \( i = 1, ..., n \), \( \lim_{t \to \infty} x_i(t) \) is no greater than \( d_i \). In other words, for
any $\epsilon > 0$, there exists a $T_i \geq 0$ such that $x_i(t) < d_i + \epsilon$ for any $t > T_i$. Furthermore, $T_i$ can be chosen as

$$T_i = \begin{cases} 0 & \text{if } x_i(0) \leq d_i + \epsilon, \\ \tau \ln \left( \frac{x_i(0) - d_i}{\epsilon} \right) & \text{if } x_i(0) > d_i + \epsilon. \end{cases}$$

(4.35)

**Proof:** Since $v \sum_k \sigma(x_k - b) \geq 0$ for any $x$, we have $M_x\{-x_i + d_i - v \sum_k \sigma(x_k - b)\} \leq M_x\{-x_i + d_i\} = -x_i + d_i$, where $M_x\{\cdot\}$ is defined in Appendix A. For $x(t)$ is a Filippov solution of (4.31) and (4.32), $x(t)$ should satisfy (A.4) (see Appendix A): $\tau \frac{dx_i(t)}{dt}$ is no greater than $M_x\{-x_i + d_i - v \sum_k \sigma(x_k - b)\} \leq -x_i + d_i$ for (at least) almost all $t$. According to the Comparison Theorem (see Appendix A), $x_i(t) \leq (x_i(0) - d_i) e^{-\frac{t}{\tau}} + d_i$. It can be seen from this inequality and from (4.35) that $x_i(t)$ is less than $d_i + \epsilon$ for any $t > T_i$. Q.E.D.

The tangent vector of the state trajectory, $\frac{dx(t)}{dt}$, does not necessarily satisfy (4.31) and (4.32) for almost all $t$ (recalling the example in (4.33)), so we can not derive inequalities directly from (4.31) and (4.32); rather than that, we use (A.4) (Appendix A) to estimate bounds for $\frac{dx}{dt}$.

**Corollary 4.3** If $d_{\max} < b$, the network of (4.31) and (4.32) has only one equilibrium $x^*$ with $x^*_i = d_i$ for all $i$, and at the equilibrium all neurons output 0. Furthermore, the equilibrium $x^*$ is globally asymptotically stable.

**Proof:** We show that $x(t)$ converges to $x^*$ no matter where $x(t)$ originates; following this, it has to be true that $x^*$ is the only equilibrium of the system. Since $d_{\max} < b$, according to Proposition 4.10, there exists a $T \geq 0$ ($T = 0$ when $x(0) \in \{x \mid x_i < b, i = 1, \ldots, n\}$) such that for any $t \geq T$ $x(t)$ belongs to $\{x \mid x_i < b, i = 1, \ldots, n\}$, which is an open set. In this open set, $x(t)$ is a solution of the following differential equations with continuous right hand sides: $\tau \frac{dx_i(t)}{dt} = -x_i(t) - d_i$ for $i = 1, \ldots, n$. Consequently, $x_i(t) = (x_i(T) - d_i) e^{-\frac{t}{\tau}} + d_i$ for any $t \geq T$, and further $\lim_{t \to \infty} x_i(t) = d_i$ and $\lim_{t \to \infty} x(t) = x^*$. Due to the convergence of $x(t)$ to $x^*$ no matter where $x(t)$ originates and the exponential convergence of $x(t)$ toward $x^*$ whenever $x(t)$ originates in $\{x \mid x_i < b, i = 1, \ldots, n\}$—an open set containing $x^*$, the equilibrium $x^*$ has to
be globally asymptotically stable. It is obvious that at the equilibrium all neurons output 0. Q.E.D.

**Proposition 4.11** Assume that $q$ is a nonnegative integer, and $q$ and $v$ satisfy $qv > d_{\text{max}} - b$. Let $x(t)$ be a state trajectory of (4.31) and (4.32) originating from $x(0)$. Denote $F = \{ x \mid \sum_i \sigma(x_i - b) \leq q \}$ and $F^c = \{ x \mid \sum_i \sigma(x_i - b) \geq q + 1 \}$. Then for any $x(0) \in F$, $x(t)$ remains in $F$ for any $t > 0$; for any $x(0) \in F^c$, $x(t)$ will converge to $F$ such that $x(t)$ belongs to $F$ for any $t > T$, where

$$T = \max \{ \tau \ln \frac{x_i(0) + qv - d_i}{b + qv - d_i} \mid x_i(0) > b \}.$$  

(4.36)

**Proof:** The proof has the following two parts.

Part 1: Show that $x(t)$ remains in $F$ for any $t > 0$, given $x(0) \in F$. We prove this by contradiction. Assume that there exists a time $T < \infty$ at which $x(T) \in F^c$. Then denote $T_0 = \sup \{ t \mid x(t) \in F, 0 \leq t < T \}$, and introduce a notation $K(t) = \{ k \mid x_k(t) > b \}$. Obviously, $\sum_i y_i(t) = |K(t)|$, where $|K(t)|$ denotes the number of elements in $K(t)$. The outline of the proof for Part 1 is to show first $x(T_0) \in F$ and then to derive $x_k(T) \leq b$ for some $k \in K(T)$, which causes contradiction to the definition of $K(T)$.

First we show $x(T_0) \in F$. Assume this is not true, i.e., $x(T_0)$ belongs to $F^c$. Denote $\epsilon = \min \{ x_k(T_0) - b \mid k \in K(T_0) \} > 0$ and

$$B(x(T_0), \epsilon) = \{ x' \mid \|x' - x(T_0)\| < \epsilon \}.$$  

(4.37)

Since $x(T_0) \in F^c$, we have $\sum_i y_i(T_0) = |K(T_0)| \geq q + 1$. It can be further tested that, for any $k \in K(T_0)$, $x_k$ is greater than $b$ for any $x \in B(x(T_0), \epsilon)$. Hence, for any $x \in B(x(T_0), \epsilon)$, $\sum_i y_k$ is no less than $|K(T_0)| \geq q + 1$, which implies $B(x(T_0), \epsilon) \subset F^c$. Since $B(x(T_0), \epsilon)$ is an open set centered at $x(T_0)$ and $x(t)$ is a continuous function of $t$, there exists a $\delta > 0$ such that, for any $t \in (T_0 - \delta, T_0 + \delta)$, $x(t)$ belongs to $B(x(T_0), \epsilon) \subset F^c$. However, since $T_0 = \sup \{ t \mid x(t) \in F, 0 \leq t < T \}$, there exists at least one $t_\delta \in (T_0 - \delta, T_0]$ such that $x(t_\delta) \in F$ no matter how small $\delta$ is. This leads to contradiction. Therefore, $x(T_0)$ should not belong to $F^c$, and thus $x(T_0) \in F$ and $T_0 < T$ ($T_0$ can not equal $T$ due to $x(T) \in F^c$).
Next we show $x_k(T) \leq b$ for some $k \in K(T)$ (this will lead to contradiction to the
definition of $K(T)$). For $|K(T_0)| \leq q < q + 1 \leq |K(T)|$, there exists at least one $k$
satisfying $k \in K(T)$ but $k \not\in K(T_0)$, i.e., $x_k(T_0) \leq b$ while $x_k(T) > b$. According to the
definition of $T_0$, $x(t)$ belongs to $F^c$ for any $t \in (T_0, T)$. Thus for any $t \in (T_0, T)$ there
exists a $\epsilon(t) > 0$ such that $x(t)$’s neighborhood $B(x(t), \delta)$ is a subset of $F^c$ whenever
$\delta \in (0, \epsilon(t))$ (following the steps similar to the construction of $B(x(T_0), \epsilon)$ in (4.37)).
As a consequence, for any $x' \in B(x(t), \delta)$ we have $-x'_k + d_k - v \sum_{i \neq k} \sigma(x'_i - b) \leq
-x'_k + d_k - v(\sum_{i} \sigma(x'_i - b) - 1) \leq -x'_k + d_{\text{max}} - qv < -x'_k + b$. Referring to (A.4), we
may derive
\[
\tau \frac{dx_k(t)}{dt} \leq M_x \{-x_k + d_k - v \sum_{i \neq k} \sigma(x_i - b)\} \\
\leq M_x \{-x_k + b\} \\
= -x_k(t) + b.
\]
Further, according to the Comparison Theorem (Appendix A), we have
\[
x_k(t) \leq (x_k(T_0) - b)e^{-\frac{t - T_0}{\tau}} + b \leq b
\]
for any $t \in [T_0, T)$. Due to the continuity of $x_k(t)$ at time $T$, we have $x_k(T) \leq b$,
which causes contradiction to $k \in K(T)$ and which completes the proof of Part 1.

Part 2: Show that, given $x(0) \in F^c$, $x(t)$ will converge to $F$, and will be trapped
in $F$ in a time no greater than (4.36). We prove this by contradiction. Denote
$T_f = \inf \{t \mid x(t) \in F, t \geq 0\}$. Assume
\[
T_f > T = \max\{\tau \ln \frac{x_i(0) + qv - d_i}{b + qv - d_i} \mid x_i(0) > b\}.
\]
According to the definition of $T_f$, $x(t)$ remains in $F^c$ for $t < T_f$. Similar to the
argument in Part 1, we may derive that, for almost all $t < T_f$, $\frac{dx_i(t)}{dt} \leq -x_i(t) + d_i - qv$.
Using the Comparison Theorem (Appendix A) again, we have
\[
x_i(t) \leq (x_i(0) + qv - d_i)e^{-\frac{t}{\tau}} + d_i - qv
\]
for any $t \in [0, T_f)$. Considering those $i$ with $x_i(0) \leq b$, if $x_i(0) + qv - d_i \geq 0$, then
$x_i(t) \leq (x_i(0) + qv - d_i)e^{-\frac{t}{\tau}} + d_i - qv \leq x_i(0) \leq b$ for any $t \in [0, T_f)$; if $x_i(0) + qv - d_i < 0$,
then $x_i(t) < d_i - qv \leq d_{\text{max}} - qv < b$ for any $t \in [0, T_f)$. For those $i$ with $x_i(0) > b$ and for any $t \in [T, T_f)$, we have

\[
x_i(t) = \left( x_i(0) + qv - d_i \right) e^{-\frac{t}{\tau}} + d_i - qv \leq (x_i(0) + qv - d_i) e^{-\ln \frac{x_i(0) + qv - d_i}{b - d_i} | x_i(0) > b} + d_i - qv = b.
\]

Therefore, $x_i(t)$ is no greater than $b$ for any $i$ and any $t \in [T, T_f)$. That is to say, $x(t)$ belongs to $F$ for $t \in [T, T_f)$, which is contrary to the fact that “$x(t)$ remains in $F^c$ for $t < T_f$.” Hence, the assumption of $T_f > T$ cannot hold, and thus we have $T_f = \inf \{ t \mid x(t) \in F, t \geq 0 \} \leq T$, which implies $x(t) \in F$ for any $t > T$. This completes the proof of Part 2. Q.E.D.

Consider two special cases of Proposition 4.11: $q = 0$ and $q = 1$. In the former case, $d_{\text{max}} < b$, Proposition 4.11 tells us: if $y_i(0) = 0$ for all $i$, then $y_i(t)$ will keep being 0 for any $t > 0$; if $y_i(0) = 1$ for some $i$, then all neurons will be quiet (outputting 0) after a time interval no greater than $T = \max\{ \tau \ln \frac{x_i(0) - d_i}{b - d_i} | x_i(0) > b \}$. In the latter case, $v > d_{\text{max}} - b$, Proposition 4.11 states that there is at most one neuron active (outputting 1) in the network for any $t > T$, where $T$ equals 0 if initially there is at most one active neuron in the network and $T$ equals $\max\{ \tau \ln \frac{x_i(0) + qv - d_i}{b + qv - d_i} | x_i(0) > b \}$ otherwise.

**Proposition 4.12** Let $v$ be strictly greater than $d_{\text{max}} - b$, the set $I = \{ i \mid d_i > b \} = \{ i_1, ..., i_p \}$ be nonempty, and $d_j$ be strictly less than $b$ for $j \notin I$. Then the network of (4.31) and (4.32) has the following properties:

- The system has exactly $p$ stable equilibria $x^{(1)}, ..., x^{(p)}$, where $x^{(k)}$ $(k = 1, ..., p)$ is determined by $x_{i_k}^{(k)} = d_{i_k}$ and $x_j^{(k)} = d_j - v$ for $j \neq i_k$; at the equilibrium $x^{(k)}$ only the $i_k$-th neuron outputs 1 and other neurons output 0.

- If $p = 1$, any trajectory of $x(t)$ will converge to the equilibrium $x^{(1)}$ no matter where $x(0)$ is located.

- If $p \geq 2$, any trajectory of $x(t)$ will converge to one of the stable equilibria $x^{(k)}$.  

93
According to Proposition 4.11, there exists a time
we have
\[ x \]
Therefore,
have any equilibrium outside the set
Obviously, there is no stable equilibrium on \( U \) in the sense that some infinitely small perturbation can cause \( x(t) \) on \( U \) to leave \( U \) and converge to one of \( x^{(k)} \), \( k = 1, \ldots, p \).

**Proof:** First introduce the following notations:

\[
\begin{align*}
F &= \{ x \mid \sum_k \sigma(x_k - b) \leq 1 \}, \\
F^c &= \{ x \mid \sum_k \sigma(x_k - b) > 1 \}, \\
F_i &= \{ x \mid x_i > b, x_k \leq b, k \neq i \}, \quad i = 1, \ldots, n, \\
F_0 &= \{ x \mid x_k \leq b, k = 1, \ldots, n \}.
\end{align*}
\]

According to the above definition, \( F_r \cap F_s \) is empty for any \( r \neq s \) and \( r, s \in \{0, 1, \ldots, n\} \), and \( F \) equals \( F_0 \cup F_1 \cdots \cup F_n \), i.e., \( \{ F_r, r = 0, 1, \ldots, n \} \) is a partition of \( F \). Further, denote

\[
\begin{align*}
F^\text{inner}_i &= \{ x \mid x_i > b, x_k < b, k \neq i \}, \quad i = 1, \ldots, n, \\
F^\text{outer}_0 &= \{ x \mid x_k < b, k = 1, \ldots, n \}.
\end{align*}
\]

Obviously, \( F^\text{inner}_r (r \in \{0, 1, \ldots, n\}) \) is the largest open set contained in \( F_r \).

The rest of the proof consists of the following four parts.

Part 1: Prove that, no matter where \( x(0) \) is located, there exists a time \( T \geq 0 \) such that for any \( t \geq T \) we have \( x(t) \in F_0 \cup F_1 \cup \cdots \cup F_i \), where \( i_1, \ldots, i_p \in I \).

According to Proposition 4.11, there exists a time \( T_1 \geq 0 \) (\( T_1 = 0 \) when \( x(0) \in F \)) such that \( x(t) \) belongs to \( F \) for any \( t \geq T_1 \). Further, since \( d_j < b \) for \( j \not\in I \), according to Proposition 4.10, there exists a time \( T_2 \) such that for any \( t \geq T_2 \) and for any \( j \not\in I \) we have \( x_j(t) < b \) and thus \( x_j(t) \not\in F_j \). Let \( T = \max\{T_1, T_2\} \). Then, for any \( t \geq T \), \( x(t) \) belongs to \( F = F_0 \cup F_1 \cup \cdots \cup F_n \), but does not belong to \( F_j \) for any \( j \not\in I \). Therefore, \( x(t) \) belongs to \( F_0 \cup F_1 \cup \cdots \cup F_{i_p} \) for any \( t \geq T \).

The conclusion of Part 1 implies that the network of (4.31) and (4.32) does not have any equilibrium outside the set \( F_0 \cup F_1 \cup \cdots \cup F_{i_p} \), and any trajectories of \( x(t) \), no matter where it originates, will eventually be trapped in \( F_0 \cup F_1 \cup \cdots \cup F_{i_p} \) within a finite time. Therefore, in the following we only need to consider the case where \( x(t) \) originates in \( F_0 \cup F_1 \cup \cdots \cup F_{i_p} \).
Part 2: Prove that \( x(t) \) will converge to the equilibrium \( x^{(k)} \) if \( x(0) \in F_{i_k} \) \( (k = 1, \ldots, p) \), and further show that \( x^{(k)} \) is asymptotically stable. It is obvious that at the equilibrium \( x^{(k)} \) only the \( i_k \)-th neuron outputs 1 and other neurons output 0.

In this part, first consider the case of \( x(0) \in F_{i_k}^{inner} \) \( (i_k \in I) \), and show that \( x(t) \) remains in \( F_{i_k}^{inner} \) for any \( t > 0 \). We prove this by contradiction. Assume that there is some time \( t \) such that \( x(t) \not\in F_{i_k}^{inner} \). Then denote \( T = \inf \{t \mid x(t) \not\in F_{i_k}^{inner} t \geq 0\} \).

Since \( F_{i_k}^{inner} \) is an open set, it has to be true that \( x(T) \not\in F_{i_k}^{inner} \) and \( x(t) \in F_{i_k}^{inner} \) for any \( t \in [0, T) \) (proof of this is similar to that of \( x(T_0) \in F \) in the proof of Proposition 4.11). In the open set \( F_{i_k}^{inner}, x(t), t \in [0, T) \), is a solution of differential equations with continuous right hand side: \( \tau \frac{dx_{i_k}(t)}{dt} = -x_{i_k}(t) + d_{i_k} \) and \( \tau \frac{dx_j(t)}{dt} = -x_j(t) + d_j - v \) for \( j \neq i_k \). Consequently, for any \( t \in [0, T), x_{i_k}(t) = (x_{i_k}(0) - d_{i_k}) e^{-\frac{t}{\tau}} + d_{i_k} \geq \min \{x_{i_k}(0), d_{i_k}\} > b \), and \( x_j(t) = (x_j(0) - d_j + v) e^{-\frac{t}{\tau}} + d_j - v \leq \max \{x_j(0), d_j - v\} < b \) for \( j \neq i_k \). Due to the continuity of \( x(t) \) at \( t = T \), we have \( x_{i_k}(T) \geq \min \{x_{i_k}(0), d_{i_k}\} > b \) and \( x_j(T) \leq \max \{x_j(0), d_j - v\} < b \) for \( j \neq i_k \), i.e., \( x(T) \in F_{i_k}^{inner} \), which is contrary to our previous assumption. So we proved that \( x(t) \) remains in \( F_{i_k}^{inner} \) for any \( t > 0 \), whenever \( x(0) \in F_{i_k}^{inner} \).

Next we show that \( x(t) \) converges to \( x^{(k)} \) whenever \( x(0) \in F_{i_k}^{inner} \), and \( x^{(k)} \) is an asymptotically stable equilibrium. Since \( x(t) \) remains in \( F_{i_k}^{inner} \) for \( t \geq 0 \), as already shown in the previous paragraph, we have \( x_{i_k}(t) = (x_{i_k}(0) - d_{i_k}) e^{-\frac{t}{\tau}} + d_{i_k} \) and \( x_j(t) = (x_j(0) - d_j + v) e^{-\frac{t}{\tau}} + d_j - v \) for \( j \neq i_k \). It is clear that \( \lim_{t \to \infty} x_{i_k}(t) = d_{i_k} \) and \( \lim_{t \to \infty} x_j(t) = d_j - v \) for \( j \neq i_k \). Thus \( \lim_{t \to \infty} x(t) = x^{(k)} \). Obviously, \( x^{(k)} \) is an equilibrium of the system. Due to the exponential convergence of \( x(t) \) toward \( x^{(k)} \) whenever \( x(t) \) originates in \( F_{i_k}^{inner} \), an open set containing \( x^{(k)} \), the equilibrium \( x^{(k)} \) is asymptotically stable.

The rest of Part 2 is to show that \( x(t) \) also converges to \( x^{(k)} \) when \( x(0) \) belongs to \( F_{i_k} \) but does not belong to \( F_{i_k}^{inner} \), i.e., \( x(0) \in F_{i_k} - F_{i_k}^{inner} \). Due to the previous paragraph, we only need to show that there exists some \( t > 0 \) such that \( x(t) \in F_{i_k}^{inner} \).

Let \( \epsilon = x_{i_k}(0) - b \), which is greater than 0 due to \( x(0) \in F_{i_k} \). Consider in the space of \( x \) the \( \epsilon \)-neighborhood of \( x(0) \), \( B(x(0), \epsilon) \). According to the definition of \( \epsilon \), \( x_{i_k} \) is strictly greater than \( b \) for any \( x \in B(x(0), \epsilon) \). Due to the continuity of \( x(t) \), there exists a \( \delta > 0 \) such that \( x(t) \in B(x(0), \epsilon) \) for any \( t \in [0, \delta] \). Obviously
Note that for any $t \in [0, \delta]$ (which implies $x(t) \in B(x(0), \epsilon)$) and any $j \neq i_k$, $M_x\{-x_j + d_j - v \sum_{r \neq j} \sigma(x_r - b)\} \leq M_x\{-x_j + d_j - v\} = -x_j + d_j - v$ (refer to Appendix A for the definition of $M_x\{\cdot\}$). According to (A.4), $\frac{dx_j(t)}{dt}$ should be no greater than $M_x\{-x_j + d_j - v \sum_{r \neq j} \sigma(x_r - b)\} \leq -x_j + d_j - v$ for almost all $t \in [0, \delta]$. Using Comparison Theorem (Appendix A), for any $j \neq i_k$ (which implies $x_j(0) \leq b$) we have $x_j(t) \leq (x_j(0) - d_j + v) e^{-\frac{\delta}{T}} + d_j - v$ for any $t \in [0, \delta]$ and thus $x_j(\delta) \leq (x_j(0) - d_j + v) e^{-\frac{\delta}{T}} + d_j - v < \max\{x_j(0), d_j - v\} \leq b$. In summary, we have shown that $x_{i_k}(\delta) > b$ and $x_j(\delta) < b$ for $j \neq i_k$, i.e., $x(\delta) \in F_{i_k}^\text{inner}$. Starting from $x(\delta)$, an inner point of $F_{i_k}^\text{inner}$, $x(t)$ will converge to $x^{(k)}$ (according to the previous paragraph).

Part 3: Show that $x(t)$ will converge to one of $x^{(k)}$, $k = 1,...,p$, or converge to and reach the set $U$ defined by (4.41) within a finite time, when $x(0) \in F_0$. Since Part 3 is obviously true for the case of $x(0) \in U$, in the following we only need to consider the case of $x(0) \in F_0 - U$. Note that $U$ is empty for the case of $p = 1$, thus when $p = 1$ the statement of Part 3 becomes that any trajectory of $x(t)$ will converge to the equilibrium $x^{(1)}$.

First prove that for any $x(0) \in F_0^\text{inner}$ there exists some $t$ such that $x(t) \not\in F_0^\text{inner}$. This can be shown by contradiction, following the steps similar to those in the proof of Corollary 4.3: Assume that $x(t) \in F_0^\text{inner}$ for any $t > 0$, and then we may derive that $\lim_{t \to \infty} x_i(t) = d_i$ for any $i$, which implies that $d_i \leq b$ for any $i$ and which is contrary to $d_i > b$ for $i \in I$.

Then it makes sense to denote $T_0 = \inf\{t \mid x(t) \not\in F_0^\text{inner}, t \geq 0\}$. Because (i) $F_0^\text{inner}$ is an open set, (ii) $F_0$ is the closure of $F_0^\text{inner}$, and (iii) $x(t)$ is continuous, $x(T_0)$ should stay on the boundary of $F_0$, i.e., $x(T_0) \in F_0 - F_0^\text{inner}$. If $x(T_0)$ belongs to $U$, then we are done. So in the following, we only need to consider the case that $x(t)$ starts from a point in $F_0 - F_0^\text{inner} - U$. Without loss of generality, assume that $x(t)$ originates in $F_0 - F_0^\text{inner} - U$ at time 0, which falls into the following two categories.

Case 3.1: $x_j(0) = b$ for some $j \not\in I$. According to Proposition 4.10, $x_j(t)$ will be strictly less than $b$ for any $j \not\in I$ and any $t > 0$. Arbitrarily pick a $t_0 > 0$, we know from the proof of Part 1 that $x(t_0)$ could only belong to $F_0 \cup F_i_1 \cup \cdots \cup F_i_p$. If $x(t_0) \in F_i_k$, $i_k \in I$, then $x(t)$ will converge to $x^{(k)}$, according to Part 2. According to the previous arguments in Part 3, if $x(t_0) \in F_0$, then there exists a $T \geq t_0$ ($T = t_0$, if
when \( x(t_0) \) is already on the boundary of \( F_0 \) such that \( x(T) \) is on the boundary of \( F_0 \) with \( x_i(T) = b \) for some \( i \in I \) and \( x_j(T) < b \) for any \( j \not\in I \). If \( x(T) \) belongs to \( U \), then we are done. Therefore, we only need to consider the case that \( x_{i_k}(T) = b \) for only one \( i_k \in I \) and \( x_j(T) < b \) for any \( j \neq i_k \). This case will be covered by the argument for Case 3.2.

Case 3.2: \( x_{i_k}(0) = b \) for only one \( i_k \in I \), and \( x_j(0) < b \) for any \( j \neq i_k \). Let \( \epsilon = \min \{ b - x_j(0) \mid j \neq i_k \} \). Consider the \( \epsilon \)-neighborhood of \( x(0) \), \( B(x(0), \epsilon) \). According to the definition of \( \epsilon \), \( x_j \) is strictly less than \( b \) for any \( x \in B(x(0), \epsilon) \). Due to the continuity of \( x(t) \), there exists a \( \delta > 0 \) such that \( x(t) \in B(x(0), \epsilon) \) for any \( t \in [0, \delta] \). Obviously \( x_j(\delta) < b \) for any \( j \neq i_k \). Note that for any \( t \in [0, \delta] \) \( x(t) \in B(x(0), \epsilon) \), \( M_x \left\{ \left( -x_{i_k} + d_{i_k} - v \sum_{j \neq i_k} x_j \right) \right\} = -x_{i_k} + d_{i_k} \). Denote \( z(t) = x_{i_k}(t) \). According to (A.5), \(-\tau(dx_{i_k}(t)/dt) = \tau(dx(t)/dt) \) should be no greater than \( M_x \left\{ \left( -x_{i_k} + d_{i_k} - v \sum_{j \neq i_k} x_j \right) \right\} = -x_{i_k} + d_{i_k} = -z + d_{i_k} \) for almost all \( t \in [0, \delta] \). Using Comparison Theorem, \( z(t) \leq (z(0) + d_{i_k}) e^{-\frac{\tau}{\delta}} - d_{i_k} \) for any \( t \in [0, \delta] \) and thus \( x_{i_k}(\delta) > (x_{i_k}(0) - d_{i_k}) e^{-\frac{\tau}{\delta}} + d_{i_k} \). In summary, we have shown that \( x_{i_k}(\delta) > b \) and \( x_j(\delta) < b \) for \( j \neq i_k \), i.e., \( x(\delta) \notin F_{i_k}^{\text{inner}} \). Starting from \( x(\delta) \), \( x(t) \) will converge to \( x^{(k)} \), according to Part 2.

Part 4: Show that there is no stable equilibrium on \( U \) in the sense that, for any \( x(0) = x_0 \in U \), some infinitely small perturbations can cause \( x(t) \) to deviate from \( U \) and converge to a stable equilibrium, which is one of \( x^{(k)} \), \( k = 1, \ldots, p \). For any \( x_0 \in U \), consider the following perturbation on the initial value of \( x(t) \) starting from \( x_0: x(0) = x_0 + \epsilon e_{i_k} \), where \( \epsilon > 0 \) and \( e_{i_k} = [0, \ldots, 0, 1, 0, \ldots, 0]^T \) with the \( i_k \)-th element of \( e_{i_k} \) being 1 and other elements being 0, \( i_k \in I \). Under such perturbation, no matter how small \( \epsilon \) is, we have \( x(0) \notin F_{i_k} \) and thus \( x(t) \) converges to \( x^{(k)} \), according to Part 2. Therefore, there can not be any stable equilibrium on \( U \).

Part 1 through Part 4 imply that the system has exactly \( p \) stable equilibria \( x^{(1)}, \ldots, x^{(p)} \), where \( i_1, \ldots, i_p \in I \). Q.E.D.

Let us see some examples of equilibria in the networks with neuronal activation functions of the form (4.30). First, consider a two-neuron network with \( v = 1 \) and \( b = 1 \) for both neurons. The left picture of Fig. 4-6 presents an example of a single
Figure 4-6: Dynamics of two neurons with symmetric lateral inhibition $v = 1$ and discontinuous activation function of form (4.30) where $b = 1$. Left: $d_1 = 1.5$ and $d_2 = 0.8$. The network has only one equilibrium, $(1.5, -0.2)$, which is globally asymptotically stable. Right: $d_1 = 1.5$ and $d_2 = 1.2$. The network has three equilibria, two of which—$(0.5, 1.2)$ and $(1.5, 0.2)$—are asymptotically stable, and the other one, $(1, 1)$, unstable.

Figure 4-7: Dynamics of a three-neuron network with discontinuous neuronal activation function of form (4.30). $v = 1$, $b = 1$, and $d = [1.8, 1.5, 1.2]^T$. The network has three stable equilibria, $(1.8, 0.5, 0.2)$, $(0.8, 1.5, 0.2)$, and $(0.8, 0.5, 1.2)$.
equilibrium for the network inputs of $d_1 = 1.5$ and $d_2 = 0.8$. The right picture of Fig. 4-6 shows that the network has two stable equilibria and one unstable equilibrium for the inputs of $d_1 = 1.5$ and $d_2 = 1.2$, both of which are larger than the threshold $b$. This is consistent with the conclusion of Proposition 4.12. The next example is a three-neuron network with lateral inhibition of strength $v = 1$, neuronal threshold $b = 1$, and network inputs $d = [1.8, 1.5, 1.2]^T$. Fig. 4-7 shows that the network has three stable equilibria at $(1.8, 0.5, 0.2)$, $(0.8, 1.5, 0.2)$, and $(0.8, 0.5, 1.2)$. This is also consistent with Proposition 4.12.

4.5 Discussion

4.5.1 Hysteresis

In the study of recurrent networks, one of the most interesting problems is the uniqueness of network equilibrium. From practical point of view, if a network possesses a unique equilibrium which in addition is globally asymptotically stable, this network becomes very attractive to optimization problems, since global convergence prevents a network from the risk of being trapped at some local minimum of the energy function [47]. On the other hand, however, networks with multiple stable equilibria are of significant value in applications as well. For example, they have been widely used as associative memories [73, 74]. Various input patterns can be stored in an associative memory as stable equilibria of the network. Moreover, we show that the recurrent networks with lateral inhibition and possessing multiple stable equilibria can exhibit hysteresis in the response of output state transition to shift in input state. (Hysteresis means the lagging of an effect behind its cause; it is originally referred to as the phenomenon in which the magnetic induction of a ferromagnetic material lags behind the changing magnetic field.) Such behavior is characteristic of the Schmitt Trigger circuit [169] that can be used to guard systems against spurious state switches due to input signal noise.

Fig. 4-8 presents examples of hysteresis realized by networks with lateral inhibition. Each pair of pictures in Fig. 4-8, (a and b), (c and d), or (e and f), shows the trajectories of quasi-steady state responses $x_1$ and $y_1$ of Neuron 1, one neuron in a
Figure 4-8: Hysteresis in two-neuron networks with lateral inhibition. Each pair of pictures in (a,b), (c,d), or (e,f) present the trajectories (with dashed arrows indicating the directions of motions) of $x_1$ and $y_1$ of one neuron as its input $d_1$ is quasi-statically increased from 0.4 to 1.6 and then decreased from 1.6 to 0.4. The other input is fixed to $d_2 = 1$. The network evolves from $x = 0$ as $d_1$ starts to increase from 0.4, but transitional responses are not shown here.  

a,b: The network with single equilibrium does not have the effect of hysteresis. The activation function is of the form of (4.18). $a = 1/20$, $b = 1$, and $v = 0.1$.  

c,d: The network with multiple equilibria may have the effect of hysteresis. The activation function is of the form of (4.18). $a = 1/10$, $b = 0.5$, and $v = 1$.  

e,f: The network clearly shows hysteresis. The neuronal activation function is of the form of (4.30). $b = 0.5$ and $v = 1$.  

100
two-neuron network, as the input to Neuron 1 \( d_1 \) is quasi-statically increased from 0.4 to 1.6 and then decreased from 1.6 to 0.4. The input to the other neuron is fixed to \( d_2 = 1 \). The lower two pairs of pictures, Fig. 4-8 (c, d, e, and f), demonstrate the hysteresis or Schmitt-trigger-like effect of two networks that possess multiple equilibria for some range of inputs. Neuron 1 in either networks changes its output state abruptly when its input rises above a certain reference; however, the output does not switch back automatically when the input level sinks again unless a second, lower reference voltage is crossed. In contrast, the two-neuron network shown in Fig. 4-8 (a and b) does not present such hysteresis effect, since this network has only one stable equilibrium for any network inputs—this is guaranteed by the fact that in this network \( v_{\text{max}} f(u) \) equals 1/2, which is less than 1.

Further simulations (not shown here) suggest that hysteresis is produced and made more prominent by increasing either the steepness of the neural activation functions \( f(\cdot) \) or the strength of lateral inhibition \( v \).

### 4.5.2 Selection Reliability and Efficiency in Striatal Networks

Lateral inhibition has been proposed as a mechanism underlying the neural competition within the basal ganglia [60, 63, 64, 112, 115, 197] that enables them to select or switch motor programs. Recently, a multi-input multi-output adaptive switching (MIMOAS) model of frontocortical and basal ganglionic interaction was proposed [112, 115], in which cortical inputs to the striatum (an input processing component of the basal ganglia), modeled as a single layer, recurrent lateral inhibitory network, must specify effectively binary control outputs. Because the control outputs are postulated to strongly modulate frontocortical activity, it is important that this system does not suffer spurious fluctuations. Rather, changes in control output should occur only with definitive changes in cortical state. Owing to the Schmitt Trigger-like behavior described above, striatal output control appears to be quite robust to noise and fluctuations in cortical input signals.

Although high-speed dynamics were not the focus of this report, it is evident from (4.1) or (4.19) that inhibitory input slows the activation of any given neuron. Thus, in general, strong lateral inhibition within the network would be expected to slow
the transition from one equilibrium state to another. If this would occur within a feedback control loop as postulated by the MIMOAS model [112, 115], such delay could give rise to instability. The model proposes that this is in fact a mechanism underlying the classic 4-6 Hz rest tremor seen in Parkinson's disease, a disorder in which a deficiency in the neurotransmitter dopamine results in decreased peak slope of activation functions of striatal neurons (among other effects) [197]. Shallower functions \( f(\cdot) \) cause lateral inhibition \( v f(\cdot) \) to decline less quickly as a function of neuronal membrane potential when it drops below the bias level \( b \). As a result, when inputs change, currently active neurons delay the emergence of activity in other neurons, resulting in inappropriately sluggish switching. Consistent with this mechanism is the particular efficacy of dopamine replacement and anticholinergic medications in relieving parkinsonian tremor—both medications likely steepen the activation function \( f(\cdot) \) [197]:

Dopamine modulates the steepness of the activation function of the striatal projection neurons via D1 receptors. D1 receptors have important interactions with both potassium channels and L-type calcium channels and can be both excitatory and inhibitory [72, 143]. As a result, dopamine increases the steepness of the activation function via terminals that express D1 receptors. Acetylcholine is thought to play an important role in shaping the responsiveness of the striatal projection neurons as well. Specifically, Wickens and Oorschot [197] suggested that increasing acetylcholine levels flatten the activation function due to the increased \( K^+ \) conductance.

Thus, it appears that both physiological and pathophysiologica l features of the striatal network may be better understood in terms of the model and analysis presented herein.
Chapter 5

Role of Basal Ganglia in Procedural Learning

Long term memory can be divided into declarative (explicit) memory and procedural (implicit) memory. Declarative memory requires conscious recall. It is also called explicit memory, since it consists of information that is explicitly stored and retrieved. In contrast, procedural memory is based on implicit learning of certain patterns about the world. It is revealed when we do better in a given task due only to repetition. In other words, procedural memory is the long-term memory of skills and procedures, such as the skills to swim or play a musical instrument. Procedural memory is often not easily verbalized, while declarative memory can be put into words.

These two types of memories and learning (acquisition of memories) are governed by different mechanisms and different brain circuits [150], and the basal ganglia (BG) have been considered to play a critical role in procedural memory and learning. Impairment of procedural learning has been described in patients with BG related disorders such as Huntington’s disease [101], Parkinson’s disease (PD) [150, 161], and stroke in the basal ganglia [14]. Meanwhile, procedural learning may be preserved in the presence of impaired declarative learning in patients with temporal lobe lesions [179] and Alzheimer’s disease [100].

This chapter studies the role of the BG and the frontal cortico (FC)-BG interaction in procedural learning. It is shown that the MIMOAS model reproduces the experimentally-observed sensitivity of procedural learning efficiency to sequence com-
Figure 5-1: A pulse-like input from cortical channel A as might arise from a cue, is positively integrated to cause saturation of corticothalamic module (register, see Fig. 2-3) RA. In this case, any one of three input activity patterns $C^{(2)}$, $C^{(3)}$, and $C^{(4)}$ activates the direct pathway by causing the upper striatal unit to win. The register may be subsequently zeroed by a pulse from cortical source B (input context pattern $C^{(1)}$) that causes the lower, indirect pathway unit to win.

Complexity and to integrity of BG circuits. Several simulations were designed to evaluate the ability of the model to learn to control the execution of movement. In each case, it is assumed that a deliberative cerebral system exists that can produce some semblance of the desired behavior initially, perhaps requiring considerable attention to do so. The role of the basal ganglia is to monitor contexts (internal and external) and to learn to reproduce the behavior given a parsimonious initial cue (internal or external) with little ongoing attention.

5.1 Frontal Registers Targeting Parietal Circuits

The MIMOAS model assumes the existence and availability of thalamocortical modules that can serve as working memory in FC (Assumption 2.3). Here it is further assumed that these thalamocortical modules are part of cortico-cortical networks that play an important role in their activation, as has been considered by several investigators [9, 78]. In particular, it is assumed that the modules tend to be preferentially activated when certain combinations of cortical units elsewhere are simultaneously active. Thus, we consider the modules to function as detectors and then indicators
Figure 5-2: Operation of several registers with partially overlapping inputs. Here registers RA and RB receive excitation from and are activated only by inputs A and B, respectively (top right traces). Register RAB receives excitation from both. However, RAB may become trained to become active only when input B is received and RA is active (indicating that A has been received earlier). Thus, RAB comes to represent the sequence A-B while RA “holds” the interval between the cues (bottom right traces).
of certain patterns of cortical activity. Once activated, these modules may in turn serve as contextual input to the BG and also may drive other cortical regions. For the encoding of sequential information, we assume that the activity of typical frontal units (Fig. 5-1) keep excited on an order of several tens of milliseconds after the onset of activity decline. That is, as noted in units in area 46 of the prefrontal cortex [5], the offset of unit activity is typically not abrupt. This feature may enable certain FC units to become transiently activated during sequential activity changes in the units that they monitor. Appropriate recurrent activity through the BG then yields a switchable register that detects and indicates the particular transition (Fig. 5-2).

Area 46 is likely to be relevant to the movement control. This area was recently shown by [5] to have signals related to different phases of serial drawing movements. Area 46 is known to have reciprocal connections with lateral posterior parietal areas (e.g. 7a, 7ip) that are involved in visuomotor information processing [26, 184]. It also connects with other prefrontal regions that could be active in declarative sequence knowledge. Area 46 in turn has strong interconnections with pre-SMA (perhaps substantially via PMdr [183]) and from there to SMA proper (M2, F6 within area 6) [107]. SMA then projects reciprocally to area 4.

Fig. 5-3 depicts the FC network used by the MIMOAS model to drive movements (see also [115]). The details of the FC processing are not emphasized, since many aspects of the interconnections of these areas are not known. However, considerable circumstantial evidence supports this architecture. In addition to input from area 46, pre-SMA communicates reciprocally with both the basal ganglia and cerebellum [80]. Both pre-SMA and cerebellum have been implicated in timing functions [1], action response set switching [160], and in implicit motor learning [43]. Moreover, electrical stimulation of pre-SMA has been associated with terminating movements and alternating movements, suggesting a switching function [67]. This makes pre-SMA an attractive convergence site for the BG or Area 46-mediated sequencing and cerebellum-mediated signal timing, sharpening and possibly scaling. PMdr may also have a role in these functions based on its position between area 46 and pre-SMA. The central model proposal is that after procedural learning thalamocortical modules in pre-SMA together with circuits involving just SMA and M1 may be sufficient to
Figure 5-3: Proposed interaction of the BG and FC in the “programming” of single-joint movements under stabilized long-loop proprioceptive control. During initial supervised phase, input from higher areas drives registers in pre-SMA (or PMdr) sequentially. After sufficient procedural learning, the BG are able to drive registers autoregressively. Prototypical signals expected in each area are indicated at left. Signals expected during cruise movement shown in parentheses. In model, pre-SMA units exhibit pulse-like activity with somewhat sharp transitions but precise waveform is not important. In cruise movements, pulses are smaller and sequential across registers in time (see Fig. 6-1) but could also toggle back and forth between two modules (see results). These are integrated in SMA to yield a net input to area 4 that is staircase-like. Integrators SMA and area 4 are presumed to not saturate easily. BG control of SMA is depicted but not used in this simulation. Diodes represent distribution of tracking error signal as strictly non-negative neural commands to agonist and antagonist muscles, $u_{ag}$, $u_{ant}$, respectively. The cerebellum (CB) provides support for registers and internal feedback for stabilization.
drive internally programmed movements. Visual guidance and declarative knowledge become optional. The cortical arrangement is closely analogous to the sequential saccade control architecture of Brown et al. [17].

5.2 Simulations

In this section, we examine the role of the BG in procedural learning based on simulations. The first and second simulated tasks are to reproduce the experimental results of Pascual-Leone et al. [150]. This is to demonstrate how the BG are involved in using cues to accelerate responses in a serial reaction time (SRT) task. The SRT task measures the time between cue and response for a random or fixed, repeating sequence of cues. For a random sequence of cues, the speed of response almost does not improve. However, for a repeating sequence of sufficient length and complexity (such that subjects under testing can not recognize the sequence easily), the time between cue and response gradually decreases before subjects become aware of the sequence.

The third simulated task is to learn to follow a guided sequence of single-joint movements more efficiently, and to eventually develop the ability reproduce the sequence automatically in response to an initial cue. Although involving movement, these tasks are considered programming and therefore elementary cognitive problems: The BG learn to predict and therefore specifically enable sequential action rather than specify the kinematics or dynamics. The identical mechanisms could be applied to sequencing non-motor circuit activation. This point is demonstrated by the fourth simulated task that involves the learning sequences of vectors where any number of elements of the vector may be active simultaneously. This task is a generalization of the example shown in [13] where there is only one active element at any time step and the active element does not repeat during the sequence of actions.

5.2.1 Procedural Learning of a Sequence of Movements

The MIMOAS model was used to reproduce the experimental results of [150]. In [150], Pascual-Leone et al. asked healthy subjects and subjects with PD and cerebellar (CB)
degeneration to quickly strike four targets as indicated by adjacent LEDs. The targets were specified in repeating pseudo-random sequences, and the rates of learning sequences of length 8, 10, or 12 were compared. Each movement started from the same position and was of the same distance, so similar-amplitude, rapid step-like position trajectories were generated in subjects upon perception of the LEDs. In the model, the visual cue was represented as a pulse-like input to pre-SMA thalamo-cortical modules and simultaneously to the striatum. Four pre-SMA modules were considered output units that strongly activated four SMA modules to yield rapid step-like inputs to area 4 (see Fig. 5-3). All commanded movements activated the agonist channel. Return movement from the target was not modeled. Response latency was defined as the interval between appearance of an LED signal and time that the arm reached the target distance. In general, the BG acted to enable both pre-SMA modules and the agonist motor cortical channel thereby accelerating the processing of cues within the system. In particular, response latency was reduced if the BG enabling occurred before the cue was delivered, i.e. predictively on the basis of preceding cues. As in the experiment, repeating sequences were given 60 times, and evaluated in blocks of 10.

Fig. 5-4 shows the simulation results for normal procedural learning. Motor performance as indicated by decreasing latency of response improved progressively until all sequences were performed efficiently. However, shortening of response time was inversely related to the length of the repeating sequence. The learning curves are quite similar to those of the human data [150]. The explanation in terms of the MI-MOAS model is that for longer sequences without simple regularities, more different cue combinations occur and therefore more subsequence registers must be formed to predict transitions accurately. Both FC (Fig. 5-5) and internal BG units (not shown) displayed a superficially random appearing pattern of binary transitions. However, the formation of interim state modules proceeds according to the orderly mechanism outlined in Figs. 5-1 and 5-2. For example, in the simulation shown in Fig. 5-5, the particular sequence to be facilitated is A-B-D-A-B-A-C-D-B. During the SRT task, a second-order register (e.g. RAB) is gradually established due to (i) the repeated overlapping of a cue signal (e.g. B) and the activity of the first-order register for the
Figure 5-4: Experimental and simulated procedural learning during SRT (serial reaction time) task using sequences of different lengths. a: Experimental data showing mean differences in response times (ms) between block 1 (random) and the subsequent blocks (2-5 repeating, 6 random) in 30 normal volunteers according to the length of the repeating sequence (circles, 8 items; triangles, 10 items; squares, 12 items). This figure is adapted with permission from [150]. b: Simulation results using the MIMOAS model.
Cue signals and cortical (Pre-SMA) register activities

Figure 5-5: Examples of some register activities during SRT task procedural learning. The top four rows show the time history of four cue signals (A, B, C, and D), and other rows show the activities of registers. A first-order register (RA, RB, RC, or RD) is activated immediately following the corresponding cue signal, and sustains its activity until another cue signal of a different type is given. A second-order register (RAB, RAC, or RCD) is activated only when the previous two cue signals are of a specific combination, while a third-order register (RABD) is specific for the previous three cues.
preceding cue signal (e.g. RA) and (ii) phasic increase of dopamine release during the motion (consistent with [167]). Similarly, a third-order register (e.g. RABD) will gradually be activated in response to repeated occurrence of three-cue sequence ABD. Using these interim registers, the BG can learn to predict transitions. In the example shown in Fig. 5-5, target D is followed by two different targets, either $D \rightarrow A$ or $D \rightarrow B$. In this case, the eventual availability of the second-order registers RBD and RCD (not shown here) will enable the BG to recognize the two different antecedent contexts within which D occurs and thereby correctly predict the targets that follow D.

5.2.2 Derangement in Procedural Learning

Pascual-Leone et al. noted the patients with PD showed less efficient procedural learning, while those with CB degeneration had either very erratic or effectively no learning. PD was simulated by reducing the winner-take-all selectivity in the striatum, reducing the direct pathway activity, and increasing indirect pathway activity as suggested by the modulatory actions of dopamine. CB degeneration was simulated by low-pass filtering the response of pre-SMA. No attempt was made to reproduce ataxia in the lower motor system. Fig. 5-6 shows the performance of the model with these lesions. Compared with the experimental data (Fig. 5-6a), the simulations using the MIMOAS model (Fig. 5-6b) were able to reproduce the weakened learning in PD and the absence of significant improvement in CB degeneration.

5.2.3 Regeneration of a Movement Sequence

Fig. 5-7 shows the position trace of the motor response to an externally guided repeating sequence of single-joint movements to four targets in different locations. This task is quite similar to the experiments of Pascual-Leone et al. [150], but is simpler in order to demonstrate a further capacity. Initially, the response is controlled entirely by a conscious system that uses the step-like cues to generate appropriate step-like intended trajectories. As expected, over the course of several repetitions, the BG circuitry learns to facilitate the movement sequence predictively. The improvement in performance is indicated by progressively decreasing latency of response. In this
Figure 5-6: Experimental and simulated procedural learning during SRT task in health, Parkinson’s disease (PD), and cerebellar (CB) degeneration. a: Experimental data showing mean differences in response times (ms) between block 1 (random) and the subsequent blocks (2-5 repeating, 6 random) in 30 normal volunteers (triangles), 20 patients with PD (squares), and 15 patients with CB degeneration (diamonds). This picture is adapted with permission from [150]. b: Simulation results using the MIMOAS model.
case, however, after every practice trial, the external sequence guidance is replaced by simply an initial cue (a “go” signal) and a tonic non-sequential input to pre-SMA as in Fig. 5-3. The BG are then challenged to reproduce the sequential behavior autonomously, if possible. This mechanism presumably corresponds to an individual’s decision to allow a particular movement sequence to be guided by automatic “muscle memory.” The entire movement can still be consciously aborted at any time by withdrawal of the tonic activation signal. Examination of the BG output units (not shown) indicates that the BG first disable all movement diffusely via the indirect pathway and then focally enable selected target modules via the direct pathway (following Mink [123]). With repetition, the internally driven sequence becomes more complete, corresponding to the establishment of “rote” memory. Eventually, a full sequence can be driven internally following the initial cue. Presumably, in healthy individuals the initial external cue itself can also be replaced by an internal decision, so that the entire motor program becomes internalized.

It is also evident from the internally driven movement sequence that although the sequence has been reproduced, some of the timing information has not been retained. Each successive movement is triggered by the completion of the preceding movement without additional delay. Thus, there is loss of variation in inter-movement timing. The MIMOAS model assumes that cadence control is not contributed by the basal ganglia (but by the cerebellum).

5.2.4 Learning General Vector Sequences

The previous examples corresponded to learning a sequence of 4-dimensional vectors having only a single active unit, i.e. \([1, 0, 0, 0]\), \([0, 1, 0, 0]\), etc. The MIMOAS model also learns sequences where any number of the units may be active simultaneously. Fig. 5-8 shows the progression of general vector learning. Because this task does not include simulated movement, the learning characteristics are, however, not identical to those of the previous examples.
Figure 5-7: Learning a multi-step movement. Each of nine sets of four sub-plots shows the performance of a modeled single-joint motor system (as in Fig. 5-3) during repeated practice of a sequence of point-to-point movements. During each trial, the system first tracks a target position cue (dotted lines in the upper two subplots in each set), and afterwards tries to repeat the multi-step movement following a single “go” signal (dotted line in the lower left subplot). The two left subplots show the position profiles of the multi-movement, while the right two subplots show simulated register activities in pre-SMA. With learning, state registers develop that respond to the external cue, the sensed movement position, and the movement completion condition \( e(t) = 0 \). As in Fig. 5-4, successful learning is associated with reduced reaction times. Slightly later the capacity develops for automatic reproduction of the whole multi-step movement in response to the first movement cue alone.
Figure 5-8: Activities of cortical units during learning of a vector sequence. The first array at top left shows the target sequence of seven four element binary vectors with each element represented by an FC unit. The remaining arrays show the sequential activities of the four cortical units in different stages of the learning as indicated by iteration number. The activity of a cortical unit is averaged over 100 ms and this value is indicated by grayscale, dark for inactive (= 0), white for fully active (= 1).
5.3 Discussion

The MIMOAS model reproduces a number of features of cognitive behavior referable to the BG function. Above all, the model provides a realistic representation of the dynamics of simple procedural learning. The model considers such procedural learning prototypical function of the basal ganglia. This fits well with the presumed role of the BG and FC areas in sequence generation and control [13, 131, 157]. In particular, the dependence of learning rate on the sequence length for a fixed number of targets and the deterioration of learning rate with simulated PD stay closely with the findings of Pascual-Leone et al. [150]. The marked impairment of procedural learning by cerebellar disease noted by these investigators was also reproduced based on the assumption of its support role in FC-BG switching. Recently, it has been noted that focal BG lesions were not correlated with reduction in implicit learning, while the volume of pre-SMA and cerebellum were positively related to implicit learning in these patients [43]. The latter findings are consistent with the MIMOAS model. The former findings may be due to the BG input-output remapping which is able to overcome focal BG lesions. Note that the focal BG lesions would not be possible in PD because PD affects all BG circuits.

In the following, we further examine the sensitivity of procedural learning efficiency in terms of formation and usage of registers or working memories.

5.3.1 Sensitivity of Procedural Learning Efficiency to Sequence Complexity

We propose that the reduction of reaction time during the procedural learning in the SRT test mainly depends on the formation of registers or internal state representations based on internal and external contexts. Once sufficient registers are available, the prediction of the next cue signal can be made by a single context-to-output mapping through the BG. Based on this proposal, we make prediction on how the learning efficiency depends on the complexity of the sequence used in the SRT test. Specifically, denoting $RT(n)$ the expected response time of a subject in the $n$-th trial of the SRT test, we assert that $RT(n)$ can be determined and thus predicted by the parameters $p_n(l)$ and $r_n(l)$, which characterize the complexity of the sequence and will be introduced
in the following.

To simplify analysis, here we only consider the SRT tests that are created by repeating a fixed sequence of trials; we do not consider random trials. Let \( N \) be the total number of considered trials, and denote \( s_1, \ldots, s_N \) the symbols of cues presented in the whole sequence. Let \( M \) be the total number of trials in the fixed sequence that repeats itself. Obviously, \( s_{kM+m} = s_m \) for any \( k = 0,1,\ldots \) and \( m = 1,\ldots,M \). For example, given the fixed sequence of \( A-B-C-C-B-A-B-C-A \), we have \( M = 9 \) and \( s_1 = A, s_2 = B, s_3 = C, s_4 = C, \ldots \), and moreover, \( s_{9k+m} = s_m \) for any \( k = 0,1,\ldots \) and \( i = 1,\ldots,9 \).

The parameter \( p_n^{(l)} \) is defined as the probability for the symbol appearing as the \( n \)-th cue in the sequence to follow the combination of \( l \) symbols that appear as the \( l \) consecutive cues preceding the \( n \)-th cue:

\[
p_n^{(l)} = \text{Prob}(\text{current cue is } s_n | \text{previous } l \text{ consecutive cues are } s_{n-l}, s_{n-l+1}, \ldots, s_{n-1}),
\]

where \( l \geq 1 \) and \( l < n \leq N \). For example, in the sequence created by repeating \( A-B-C-C-B-A-B-C-A \), consider the case of \( n = 17 \): Since \( M \) equals 9, we have \( s_{17} = s_8 = C \) and \( s_{16} = s_7 = B \). Note that in the whole sequence the cue signal \( B \) is followed by two different cues, either \( B \to A \) or \( B \to C \). The frequency of \( B \to C \) appearing in the sequence is twice as much as that of \( B \to A \), so

\[
\begin{align*}
\text{Prob(\text{current cue is } C | \text{previous cue is } B)} &= 2/3, \\
\text{Prob(\text{current cue is } A | \text{previous cue is } B)} &= 1/3.
\end{align*}
\]

According to (5.1), \( p_{17}^{(1)} \) equals 2/3. Now taking into account the cue preceding \( s_{16} \), we consider the combination of \( s_{15} \) and \( s_{16} \), i.e. \( AB \). Cue \( C \) uniquely follows cues \( AB \) in the sequence, so

\[
\text{Prob(\text{current cue is } C | \text{previous two consecutive cues are } AB)} = 1.
\]

Therefore, \( p_{17}^{(2)} \) equals 1.

The parameter \( r_n^{(l)} \) is defined as the total repeat number of cue combination \( s_{n-l}, \ldots, s_{n-1} \) in the sequence preceding the \( n \)-th trial. Referring back to the above example, \( r_6^{(1)} \) equals 2, since the symbol \( B \)—the cue of 5-th trial—appears twice in
Figure 5-9: Hypothetical reduction of response time in a SRT test with respect to the formation of registers of different orders. a: Hypothetical curves for the function of $\text{RT}^{(l)}(r, p)$ with different values of $l$ and $p$: $l = 1, p = 1/2$ and $l = 2, p = 1$, respectively. The definition of $\text{RT}^{(l)}(r, p)$ may be found in the text. b: Contribution from different registers in reducing the response time $\text{RT}(kM + m)$, where $kM + m$ ($k = 0, 1, ...$ and $m = 1, ..., M$) represent the indices of a specific cue in the fixed sequence of length $M$ when this sequence repeats itself in the trials.

Following the proposal mentioned at the beginning of this subsection, the reduction of reaction time $\text{RT}(n)$ depends on how well the cue signal can be predicted from the internal registers created during the procedural learning. We now take a close look at the contribution from registers of different orders in reducing the response time $\text{RT}(n)$. Recall that Figs. 5-1, 5-2, and 5-5 have demonstrated the formation of registers of different orders. Denote $\text{RT}^{(l)}(n)$ the hypothetical response time of the subject in the $n$-th trial of the SRT test, provided that the FC-BG system only uses the $l$-th order registers to predict the cue signal. Such hypothetical response time
RT^{(l)}(n) is assumed to be a function of \( r_n^{(l)} \) and \( p_n^{(l)} \), i.e.,

\[ RT^{(l)}(n) = RT^{(l)}(r_n^{(l)}, p_n^{(l)}). \]  

(5.2)

Furthermore, we propose that RT(n) is determined by

\[ RT(n) = \min \{ RT^{(1)}(r_1^{(1)}, p_1^{(1)}), RT^{(2)}(r_1^{(2)}, p_1^{(2)}), \ldots \}. \]  

(5.3)

See Fig. 5-9 for an example.

Based on the proposed role of the BG in procedural learning and register formation, we assert that the function RT^{(l)}(r, p), where \( r \) is a nonnegative integer and \( p \) belongs to \([0, 1]\), should have the following properties:

- RT^{(l)}(r, p) is a monotone nonincreasing function of \( r \). A larger \( r \) implies a larger number of repeats that a specific combination of \( l \) cues has appeared in the trials tested, and thus it is more likely for a register representing that cue combination to be formulated. Therefore, RT^{(l)}(r, p) tends to decrease (or at least not increase) due to the prediction of the coming cue signal using this \( l \)-th order register.

- RT^{(l)}(r, p) is a monotone nonincreasing function of \( p \). This follows that in the presence of a specific combination of \( l \) cues, if the probability is higher for a cue signal to come after that combination, then the subject’s response to the coming cue signal should be more rapid.

- \( \lim_{r \to \infty} RT^{(l)}(r, 1) \) takes the same value for all \( l \geq 1 \). That is, RT^{(l)}(r, 1) with different values of \( l \) will converge to the same plateau of response time as \( r \) increases. Here, \( p = 1 \) implies that in the presence of a specific combination of \( l \) cues, the probability of next cue signal has reached the maximum probability 1; in other words, the subject is 100% sure about what is going to show next. The response in this situation is equivalent to a response in a simple reaction time test [158].

- RT^{(l)}(r, p) is no greater than RT^{(k)}(r, p) if \( l < k \). This follows that it generally takes longer for the FC-BG system to form a register of higher order,
Figure 5-10: State transitions during a SRT test with A-B-C-C-B-A-B-C-A as the repeating sequence. As the test proceeds (a-c), gradually more registers are created and available for the FC-BG system to predict next cue signal. 

a: State transitions when only first-order registers are available. 
b: State transitions when two additional second-order registers RAB and RBC are available. Since the combination of cues AB is uniquely followed by C, the availability of RAB enables the system to predict the next cue (i.e., C) with no ambiguity. 
c: The minimum set of registers required for the system to predict every transition correctly.
because the formation of a register of higher order depends on the availability of corresponding register(s) of lower order(s) (see Fig. 5-2).

For example, consider a SRT test that uses \( A-B-C-C-B-A-B-C-A \) as the repeating sequence. With the above argument in mind, we may predict that during the procedural learning the response to the cue of \( C \) which follows cue combination \( AB \) should be more rapid than the response to either the cue of \( C \) or \( A \) which follows cue combination \( BC \). Note that \( AB \) is uniquely followed by \( C \), while \( BC \) is followed by either \( C \) or \( A \). So the correct prediction of \( C \) at the third or eighth place in the sequence of \( A-B-C-C-B-A-B-C-A \) only requires a second-order register, i.e. \( R_{AB} \) (see Fig. 5-10), but the correct prediction of \( C \) at the fourth place or \( A \) at the ninth place in the repeating sequence requires a fourth-order register, i.e. \( R_{AABC} \) for \( C \) or \( R_{BABC} \) for \( A \) (see Fig. 5-10). Further experimental investigation on this issue is presented in [116].
Chapter 6

Role of Basal Ganglia in Ballistic and Cruise Movement Generation

This chapter investigates the role of the basal ganglia (BG) in lower-level movement control under the framework of the MIMOAS model. The BG motor control function is demonstrated with simple simulations of single joint movements. The major simulated task is the production of ballistic and more constant velocity cruise movements under normal physiological conditions and in simulated Parkinson’s disease (PD). These are considered examples of frontocortico (FC)-BG involvement in lower-level motor control. For cruise movement production, it is assumed that a deliberative cerebral system is first able to specify a series of relatively evenly spaced intermediate positions to transition from one to the other at somewhat regular cadence. It can be shown that with time, the BG can learn this as a special case of the task of learning a movement sequence shown in Chapter 5. This chapter also examines the sensitivity of model behavior to external disturbances and loads, added internal noise, low-pass filtering and phase lags. These variations represent respectively the sensitivity to peripheral input, high-frequency stochastic nature of neuronal firing, and certain internal lesions.

6.1 Sensorimotor Cortical Servo Control of Single Joint Motion

We first consider the generation of ballistic movement. Ballistic movements with bell-shaped velocity profiles have been considered as typical of many natural reaching
movements. They have been studied by a number of models of the supraspinal motor control system (e.g. [21, 59, 114]). These models accept either smooth sigmoidal intended position profiles or step-like reference commands as inputs.

In this chapter, we propose to use a simplified long-loop transcortical control system for the generation of ballistic movement of single joint. Our system is demonstrated in Fig. 5-3 (see Chapter 5), and is driven by the MIMOAS model described in Chapter 2. The plant is a single joint controlled by a pair of antagonistic viscoelastic muscles. It is modeled as a simple second-order spring-damping system having transfer function:

\[
\frac{k_m}{hs^2 + b_m s + k}
\]

with parameters \( h = 0.1, b_m = 1, \) and \( k_m = 10 \) representing the inertia, viscosity, and stiffness, respectively. During externally specified voluntary point-to-point movements, activation of motor cortex arises from a proprioceptive targeting error signal \( e(t) \) formed in parietal area 5 (PPC). Similar to several proposals [21, 59, 114], some motor cortical units are considered to integrate the error signal to produce a raw motor command step. These motor cortical units function like nonlinear integrators as suggested in [114]. Further cerebellar processing [21, 30, 114] presumably completes the command formation, the details of which are not included here. However, as a byproduct, the targeting error may be partially predictive based on internal corollary efference copy discharge. Such a mechanism can contribute significantly to the stability of long loop feedback control [114] and is consistent with known projections of cerebellum to parietal area 5 [44, 87] and area 4.

Generation of constant velocity movements requires an additional mechanism. Navas and Stark [134] argued that there is no direct speed control in the human arm movement system. Instead, speed is controlled by intermittent production of small steps in position. This notion is consistent with a number of studies of intermittency in human movement [119, 175, 190], although many mechanisms may contribute to this phenomenon. Although some aspects of apparent segmentation may be related to visual input, some may clearly exist in its absence. It has been demonstrated recently [110] that the BG may be involved in the generation of the sequence of small position steps that underlie constant velocity or “cruise” (or “ramp”) movements.
that are internally generated without visual tracking. The MIMOAS model modifies and extends that work by utilizing the frontal cortex registers (Fig. 5-3) described in Chapter 5 to implement the delivery of position step sequences to area 4.

The model also uses the tracking error signal $e(t)$ as the input $Z$ to the STN in Fig. 2-4 within lower (or motor control) channels of the basal ganglia. In this case, the BG motor loop becomes automatically engaged when movement is attempted and is not activated when there is no attempt. This is a minor assumption that could be omitted. However, it is consistent with the observation of kinematic and passive sensory information in GPi activity [66, 187].

Finally, the limb control servo includes separate integrators for agonist and antagonist channels. This feature is a single degree-of-freedom representation of the distribution of cortical control across “population vectors” [52], and is particularly important for reproducing tremor and abnormalities in muscular tone that attend several BG related disorders.

6.2 Simulations

6.2.1 Automated Cruise Movements

Fig. 6-1b shows the position and speed profiles of the arm under deliberate execution of a series of short point-to-point movements. This can be viewed as a special case of a single joint movement sequence to progressively farther targets. For such a sequence of short movements, the cadence is intended to be regular. Under this framework, the movement can be learned to be produced by the BG in response to an initial cue in conjunction with a low level of tonic background activity as described above. This produces a cruise-like movement that has reduced maximum speed and a flatter speed profile. Still, because submovements are fundamentally discrete, there may remain some subtle fluctuation in speed. A series of staggered pulses of activity is predicted for pre-SMA and the BG if control is divided between several registers as in Fig. 5-5. If control is mediated by fewer registers, each might instead display (asynchronous) on-off toggling similar to that shown in vivo in globus pallidus [129].
Figure 6-1: Simulated ballistic movement and cruise movements in health (a,b) and Parkinson’s disease (PD) (c,d). Command signals in (a)-(d) are defined and illustrated in Fig. 5-3. In both (c) and (d), a 4 Hz rest tremor occurs. Compared with the normal movements, movements in PD show both slowness and co-activation of agonist and antagonist pathways. Also, pre-SMA register activities attenuate before completion of movement yielding hypometria. This is particularly prominent in cruise movements. In both position and velocity profiles the units are arbitrary but consistent.
6.2.2 Derangement of Movement in Disease

Simulated cruise and ballistic movements were made with simulated reductions in dopamine level (Fig. 6-1c, d). The first effect was the slowing of both types of movement and the development of a 4 Hz tremor at rest before and after the movements. During attempted very slow cruise movements, arrest (akinesia) was noted before the end of movement. At the same time, progressive coactivation of agonist and antagonist muscles is observed which corresponds to joint rigidity. Simulated antagonism of acetylcholinergic pathways by increasing lateral inhibition in the striatum (not shown) reversed the tremor, reduced the coactivation somewhat but did not correct the slowness.

6.2.3 Sensitivity Analysis

The responsiveness of the BG units to externally imposed disturbances was first examined. This provided a further check on physiological plausibility of the BG internal circuit model. It was found (not shown) that all units of globus pallidus (GP) responded to simulated passive manipulation of the joint. Most striatal units, however, were not activated by passive motion. The relationship between the intensity of neuronal responses in GP and peak elbow movement speed was also verified to be monotonic but quite sublinear [187].

Next the single joint movement tasks were repeated with high frequency (band) Gaussian noise added to all BG units. The results were essentially unaffected by the presence of noise that significantly distorted the signal at high frequencies (see Figs. 6-2 and 6-3. The simulation results also revealed that the cortico-BG-thalamocortical actions are more robust against the high frequency noise in simple one-step motion than in multi-step motion. For some levels of noise added to the BG units (e.g. the ratio between maximum amplitude of noise and maximum amplitude of neural signals was 0.5), the simulated motor system could successfully make the one-step motion but failed to complete the multi-step motion. This is consistent with the fact that multi-step motion requires more BG programming and more complex interaction between the BG and other components of the motor system.

Effective slowing about 60 ms in trans BG transit time was produced by adding,
Figure 6-2: Sensitivity test on a multi-step motion with high frequency (band) Gaussian noise added to all BG units. The ratio between maximum noise amplitude and maximum amplitude of BG internal signals is 0.15. 

a: Position profile of the multi-step motion. 
b: Velocity profile. 
c: External cue signal and command signal. The command signal (presumably in area 4) is integrated from the output of registers in pre-SMA. 
d: The upper five rows show simulated register activities in pre-SMA, and bottom row shows the absolute value of the error signal $e(t)$. 
e: Activity of 16 striatal units involved in this task of motion.
Figure 6-3: Sensitivity test on a simple ballistic (one-step) motion with high frequency (band) Gaussian noise added to all BG units. The ratio between maximum noise amplitude and maximum amplitude of BG internal signals is 0.5. **a:** Position profile. **b:** Velocity profile. **c:** Simulated activity of cortical units in M1 and SMA. **d-f:** Activity of BG units in STN, striatum (STR) direct pathway, STR indirect pathway, GPe, and GPI, respectively.
Figure 6-4: Sensitivity test on a multi-step motion with a pure delay of 15 ms added in each stage of neuronal signal conduction in the cortico-BG-thalamocortical loop. This produces a 60 ms slowing in trans BG transit time. Captions for (a)-(e) are similar to those of Fig. 6-2.
Figure 6-5: Sensitivity test on a multi-step motion with neuronal membrane time constant increased 25% of the original value (10 ms) for each BG unit. Captions for (a)-(e) are similar to those of Fig. 6-2.
Figure 6-6: Sensitivity test on a multi-step motion with inertia of plant increased to 500% of the original value (see (6.1)). Captions for (a)-(e) are similar to those of Fig. 6-2.
respectively, a delay of 15 ms to each of BG units. This was shown to have only a slightly destabilizing effect on the motor response (see Fig. 6-4). The effect of delaying to 60 ms either only the direct or indirect pathway alone was also examined (not shown here). In neither case was the motor performance affected strongly.

Finally, the performance of the BG-involved motor system was examined for the multi-step movement generation with (i) increased BG neuron membrane time constants (see Fig. 6-5) and (ii) increased inertia of plant (see Fig. 6-6), respectively. The simulation results also demonstrated the robustness of the BG actions.

6.3 Discussion

This section presents some discussion on the simulation results. Most of the following content is taken from [115] (Massaquoi and Mao, 2005).

6.3.1 Speed and Postural Control and Dyscontrol

The ability of the model to produce cruise movements helps to connect the roles of the BG in behavioral execution and motor and posture control. Cruise or “ramp” movements have been particularly associated with intermittency of EMG, in unit activity in the GPe [124], and with small fluctuations in speed about a relative plateau [117, 190]. The example presented here reproduces such speed profiles as well as intermittencies in unit activity of the BG (not shown) qualitatively, although it does not attempt to account for all intermittencies in human movement. The possibility that cruise movements may be generated by a series of internally programmed short point-to-point movements is consistent with the failure to detect pure velocity control experimentally [134]. Such proposal is also consistent with the preferential role of the BG and SMA in internally generated movements. This mechanism asserts an importance of the BG for cruise movements as was conjectured by Delong and Strick [36], but does not restrict their action to these movements. Indeed, model signals in the BG may be more active in ballistic movements as is observed experimentally [66, 124].

The proposed role of the BG-mediated switching in lower motor control is also consistent with several clinical features of PD. Bradykinesia results from inadequate en-
abling of agonist channels due to hypoactivity of the direct pathways as has been sug-
gested by many. This is caused by hypofacilitation of glutamatergic striatal synapses by decreased D1 agonism. The model explains akinesia, or “getting stuck,” and hypo-
metria by a general failure of the BG to recognize or act adequately in response to nonzero tracking error signals $e(t)$. As is consistent with clinical observation, this in particular is likely to occur when the error signals are small, such as just before reaching a target, or when initiating a cruise or small movement. In the latter situation, the model indicates that akinesia may also arise from lesions of frontal cortex [42, 123].

Rigidity and tremor are found to result from failures to properly disable muscle activity in relation to the intended end of movement. Two alternative mechanisms may be operative. In PD, lack of striatal dopamine diminishes the rapidity and precision of striatal neuronal selection (see Chapter 4). This consequently causes slower and less definitive switching between agonist and antagonist activations, and thus results in involuntary agonist-antagonist coactivation that manifests as rigidity. There is also an associated phase-lag that results in an instability in the transcerebellar internal feedback loop (Fig. 5-3) that appears as the classical 4-6 Hz rest tremor superimposed on the increased background muscle tone. The model tremor does not depend upon intrinsic instability in GPSTN interaction [55, 136, 159] (although it does not argue directly against such internal oscillation). Therefore, tremor disappearance during moderate speed and rapid motion can be explained by the fact that in this case the $e(t)$ is sufficiently large that it does not change sign due to phase lag. On the other hand, with very slow movements and at rest, sign change occurs and the tremor displays the characteristic alternating EMG pattern [85] (roughly $u_{ag}$ and $u_{ant}$ in Fig. 6-1). Appropriately, the model predicts that both tremor and rigidity can be treated by dopamine administration, and that the tremor in particular is sensitive to anticholinergic therapy. The different response properties owe to the slightly different mechanisms by which these agents affect striatal neuronal selectivity (see Chapter 4). Due to its strongly central mechanism, Parkinsonian tremor is characteristically little affected by peripheral loading [85], as was verified here (not shown). The basis for rigidity in most other forms of parkinsonism (e.g. strio-nigral degeneration, progressive supranuclear palsy or multi-infarct state [85]) is predicted to be hypoactivity of
the indirect pathways or of the output from GPi to thalamus that ordinarily disable antagonist pathways when agonist pathways are active unless muscular coactivation is commanded explicitly. Such a defect presumably occurs because of neuronal degeneration or injury in the striatum or globus pallidus. Consistent with the model and clinical observation, the rigidity in these disorders is not dopamine responsive and is most often not accompanied by tremor.

Modeling of other important BG disorders is beyond the scope of this thesis. However, it may be noted that a focal failure of the BG to recognize or act correctly upon the $e(t) = 0$ condition at the intended termination points of some movements, is likely to result in inappropriately persistent motor activity with mal-positioning of the extremity. Preliminary simulations indicate that this is may be able to account for certain types of dystonia.

6.3.2 Internal Neuronal Signal Characteristics

In the MIMOAS model, BG units function effectively in a binary fashion, so saturation of unit output should have little effect on system operation. This was confirmed in sensitivity testing (not shown). Accordingly, in many BG units, quantitative correlations with kinematics or dynamics have been found often to be weak or nonlinear [16], although BG unit activities are often found to be coactive with motion in a particular body part. Unit activity, especially in the striatum, is often better correlated with cues and phase of the motion within a behavior [84]. This was also confirmed in sensitivity analysis. On the other hand, some model units, especially those in the globus pallidus, derive significant input from the tracking error signal $e(t)$ via the STN. Therefore these units include considerable kinematic information as has been observed in this region [187]. Moreover, model GP units respond to passive manipulation of an extremity as has been noted in vivo [66]. Many of these features are inconsistent with the model of Suri et al. [182] that does not employ tracking error signals, or binary representations of internal state.

The simulated signals in pre-SMA, SMA and area 5 were grossly similar to unit activity recorded in these areas in vivo. Pre-SMA and BG units have been noted to be active during specific phases of multi-step movements sometimes with sharp onsets
and often with sharp offsets [95, 130, 185]. This is consistent with the switching function of the basal ganglia. During a simulated sequential movement task, frontal units (Fig. 5-7) (and therefore the striatal units to which they project) can show offsets that are time linked to the onsets of external cues as observed by [95]. Because of the quasi-binary operation of the BG under normal conditions, once FC module activity level becomes substantial, the precise time course of further activity increase is largely irrelevant in the tasks simulated here. In particular, the rise may be somewhat more gradual which would potentially cause somewhat phasic-appearing caudate signals as observed by Kermadi and Joseph [95] (Fig. 5-3) without affecting performance. We bear in mind the possibility that greater sensitivity to rise-rate may occur when sequences must be executed very rapidly or according to an internally specified cadence. Area 5 signals are often velocity-like [88], consistent with position error. However, no examples of area 5 signals in cruise movements were found in the literature. Thus, the model signals in this case represent specific predictions.
Chapter 7

More Discussion on MIMOAS Model

Based on the results of simulations and experiments presented in Chapter 5 and Chapter 6, we extend our discussion on the MIMOAS model concerning the robustness and binary formalism, modeling assertions and suggested neuroanatomical connections, and FC-BC interactions. Most content of this chapter is taken from [115] (Massaquoi and Mao, 2005).

7.1 Robustness and Binary Formalism

Sensitivity analysis in Section 6.2.3 confirmed that the overall behavior of the MIMOAS model is quite insensitive to high frequency noise and to some extent time lags and changes of neuron membrane time constants within the basal ganglia. Such robust behavior is similar to that of digital computers and other discrete state/discrete time devices, where switching operations can be reliably implemented. The robustness also minimizes our attention to the millisecond-to-millisecond details of signal processing in the basal ganglia. These details, however, are considered to be more important for spinal, cerebral cortical, and cerebellar circuitry. Thus, an important feature of the MIMOAS model is its possible clarification of an important interface between discrete time and continuous time control subsystems within the central nervous system. The model suggests that in the presence of adequate continuous time support by other systems, normal system-level cortico-BG function may be analyzed primarily in terms of discrete-time logical operations as formalized by Eq. (2.8).
It should also be noted that Eq. (2.8) asserts that if direct and indirect pathways are activated simultaneously at the striatum, direct pathway action will dominate. It is not clear, however, that direct and indirect pathways ever compete. The simple winner-take-all mechanism (Assumption 2.1 and Chapter 4) predicts that with learning, only one striatal projection unit within a local striatal network (usually a striosome/matrix complex) becomes active in response to a given context. Since the winning striatal unit projects exclusively to either GPi or GPe, the direct and indirect pathways really do not compete. On the other hand, it is conceivable that a striosome/matrix complex does not include all units that ultimately project to a given unit in GPi. In this case, certain striatal units belonging to the direct pathway and indirect pathway could “win” simultaneously and the logical asymmetry of Eq. (2.8) with respect to competition between these pathways would then become relevant. Further experimental studies will be required to evaluate this possibility.

7.2 Modeling Assertions and Suggested Neuroanatomical Connections

Since the model’s central assumptions and proposed FC motor system anatomy have not been established experimentally, they may be viewed as predictions.

The core assumption/prediction of the model is that the BG map cortical activation patterns into either direct pathway-mediated enabling, or indirect pathway-mediated disabling, of FC target systems. The strength of this mechanism is its strong consistency with previous conceptions of BG function. In comparison, specific proposals for how BG neurocircuitry operates such mechanism have incrementally novel contributions. A major simplifying assertion is that BG circuitry functions in a significantly nonlinear “on versus off” manner (Section 2.2.2) as discussed above. While the model does not insist that the BG can not operate in a continuously graded fashion, it argues that such a view (e.g. [182]) is not necessary to understand how the structures operate under physiological conditions even when it concerns grossly continuous movement.

The proposal that BG circuitry is organized as parallel, substantially indepen-
dent channels (Section 2.2.3) is consistent with the emerging picture of BG modular operation at a macroscopic and microscopic level [93]. This proposal is also shared by several other investigators [17, 49, 141]. Although it has not been sufficiently established that the BG circuits are fully functionally independent, it is an attractive hypothesis. First, it greatly simplifies both the architecture and dynamics of the internal circuits of the basal ganglia, and therefore greatly increases BG operational robustness to focal injury or dysfunction. This is consistent with the observation that individuals may suffer from regionally focal, or even task-specific dystonia while their other motor function and behavior are completely unaffected. A second attractive feature is the increased computational power that is afforded by the parallel architecture. With the effect of selecting a single winning, or enabled, thalamocortical output module among q alternatives, as in the model of Berns and Sejnowski [13], BG circuits allow for q possible output control patterns: $X = [1, 0, 0, ..., 0], [0, 1, 0, ..., 0], ..., [0, 0, ..., 0, 1]$. By contrast, when any number of output modules may be enabled simultaneously, e.g. $X = [0, 1, 0, ..., 1, 0]$, then $2^q$ control patterns are possible. This underlies capacity for full vector sequencing shown in Fig. 5-8. The extraordinary difference in processing power advocates strongly for the MIMOAS-type architecture.

The MIMOAS model also asserts that the BG continuously gate reverberating thalamocortical activity (Section 2.1.2). The importance of this feature is that the BG can both enable cortical activity and continuously suppress activity that would be driven by other systems. The latter capability is lost with the assumption of inherently bistable frontal cortical circuits (e.g. [49]). In the MIMOAS model, gated cerebellum-supported thalamocortical modules are difficult to drive, therefore monostable around zero, when disabled by the BG (see [115]). They become bistable around fully on or fully off only when continuously enabled by the BG. Thus, it is argued that the BG have the effect of making flip-flop registers either available or unavailable, rather than simply toggling them. This view is consistent with the marked hypoactivity (not simply less frequent switching) in FC circuits observed in BG disorders [42]. Further experimental work will be required to determine the validity of this contention.

The model asserts that cortically represented contexts become associated with striatal neurons projecting to either the direct or indirect pathways, but not to both
(Assumption 2.8). As indicated above, this assumption is consistent with the known
differences in dopamine D1/D2 receptor ratio in the two pathways, together with the
learning mechanism proposed (Chapter 3). Assumption 2.8 is, however, not adopted
by a number of BG models, including [9].

Assumption 2.5 suggests that under normal conditions, the BG either enable or
disable frontal circuits, but do not drive them directly. At the moment, this assum-
tion is an independent hypothesis with regard to how motor control is partitioned
within the central nervous system. It is not critical to the MIMOAS model.

Learning in the MIMOAS model relies heavily upon the convergence of certain sig-
nals at the striatum as summarized in Assumption 2.10 and Chapter 3. The proposed
mechanism is simple and the known neuroanatomical connections are quite consis-
tent with this hypothesis. However, available evidence is not yet sufficiently specific
to consider this mechanism established. Experimental validation of the proposed
learning rule is therefore very important for the MIMOAS model.

Fig. 5-3 constitutes a gross formulation of possible information flow in the control
of visually and internally triggered simple movements. There is neuroanatomical evi-
dence for all of the connections indicated. However, further refinement and validation
will be necessary. The overall concept of the figure is that pre-SMA and dorsolateral
premotor area appear to link prefrontal cortical area 46 with the lower motor system,
and that pre-SMA receives input from the “complex” BG loop and interacts with the
lateral cerebellum. This can be considered reasonably well established [80, 81, 107]
and is consistent with a more “cognitive” or programming function for these areas.
Yet, although pre-SMA is known to project to SMA, this pathway is comparatively
“modest” relative to other targets of pre-SMA [107], e.g. premotor area 6. Also, SMA
projects not only to area 4, but to area 5 as well, a connection that is not shown ex-
plicitly. Moreover, premotor area 6, SMA and area 4 all project directly to spinal
cord leading away from the notion that area 4 is a final common pathway [38].

For the moment, Fig. 5-3 assumes that the modest projection from pre-SMA to
SMA involves the subset of SMA units that do not receive proprioceptive input [144].
It is possible that this path is specifically for producing cruise movements because the
programmed sequence of command steps need not depend upon external information.
The rest of the path to SMA/area 4 involves typical units in these areas that receive proprioceptive influence. Other possible paths and further details of the motor output mechanism were considered beyond the current study. The most important omission is a more detailed analysis of the direct interaction between SMA, BG, and cerebellum. These anatomical connections are significant, and therefore the interaction plays an important role in low-level motor control. It is conjectured here that these pathways are likely most important in movement speed control. Along the thinking of Escola et al. [42], it is plausible that focal lesions or specific lack of BG support of SMA without compromise of pre-SMA could cause bradykinesia without akinesia.

7.3 Frontocortical-Basal Ganglionic Interaction

At both programming and lower motor control levels, the role of the BG in the examples presented in Chapters 5 and 6 is to learn to enable reproduction of a specific pattern of FC activation in response to a cue or context. This has the effect of reducing the amount of cortical activity or attention required for precise frontal activation, and thus rendering actions more automatic and potentially drivable concurrently with other behaviors. For behaviors that are initially directed by deliberative systems, the BG enables them to be memorized as a procedure, or “macro” in the computer science sense.

Because the BG mechanism appears to implement subcortical mappings from one state vector to another, this mechanism can define a dynamic system that codes sequences implicitly (Fig. 5-7, like several other proposals of BG function [13, 63, 64]). This contrasts with mechanisms such as Competitive Queuing based upon maintenance of tonic priming signals [20] that embody sequences explicitly (see below). Dynamic generation of sequences is comparatively efficient whenever there exist underlying rules or regularities. The proposed FC-BG system takes advantage of these possibilities by creating auxiliary state representations, in this case indicators of predictive subsequences, that are needed to realize such rules. For such reason, the model implies that the rate of sequence learning depends upon the lengths of the subsequences required to determine correct actions, rather than overall sequence length.
The dynamic selection of a next action on the basis of recent subsequences can be viewed as a simple form of action rule switching: e.g. "if $B$ was just struck, next will be either $A$ or $D$ depending upon what preceded $B$." As such, this process is consistent with observed functions of pre-SMA [160].

During procedural learning (e.g. Figs. 5-4 and 5-6) and learning a sequence of actions (e.g. Fig. 5-8), the MIMOAS model gradually establishes FC registers that encode contextual state information. Meanwhile, the model also establishes corticostriatal connections that allow decoding of the contextual state information to trigger actions or further state vector changes. Figs. 5-1, 5-2, and 5-3 illustrate how the BG are involved in both the establishment of working memory units and the transfer of such information to a motor output system. Figs. 5-1, 5-2, 5-5, and 5-8 show how register activity can be propagated to other registers in response to a sequence of stimulus events. These patterns of simulated register activities are seemingly consistent with those recorded in frontal cortical neurons during a sequence of movements [137, 185] and reproduced by Beiser and Houk [9]. However, in the MIMOAS model the encoding process involves reinforcement learning, and thus differs from that described by Beiser and Houk [9], where serial sensory events are encoded by spatially distributed sets of BG-facilitated corticothalamic units having pre-established, nontrivial corticostriatal connections. In the MIMOAS model, sensory events are not immediately or automatically encoded in FC. Rather, working memory registers are formed specifically when the system attempts to act (thereby causing phasic dopamine release [167]) in response to certain cues. We consider that phasic dopamine output might diminish as tasks become well-learned (e.g. [168]) and performance achieves expectation, and therefore it is possible that no further registers become formed. This would afford selectivity and economy of circuit usage, but at the expense of learning time. The model suggests that this accounts for much of the time required to learn new procedures and for why learning time depends upon sequence complexity. Some additional time may be required for the development of proper decoding. But this is found to be comparatively modest and less dependent on sequence complexity: Once the appropriate internal registers are available, each motion can be facilitated by a single context-to-output mapping through the BG. Thus, the ability to predictively enable
a sequence occurs rapidly upon formation of sufficient internal state representations based on internal and external contexts. The capacity to autoregressively regenerate a sequence depends upon further strengthening of internal context-to-output mappings. This may also occur quickly, but the lag is nontrivial (Fig. 5-7). Note that in the model autoregressively regenerated sequences lack proper cadence. This finding is consistent with a cerebellar role in “internal clocking” [92] and the postulated supportive role for the BG (Assumption 2.5).

The model assumes that all behaviors are initially driven by systems external to the BG. Therefore, macroscopically, BG learning should be considered supervised, with dopamine level grossly related to some internal assessment of BG performance by some mechanism that is not modeled here. This learning is analogous to classical conditioning. Initially, some non-BG system provides commands, or unconditioned “stimuli,” that elicit (unconditioned) action responses. Subsequently, contexts constitute conditioned stimuli that eventually elicit these actions as conditioned responses. By contrast, on a finer scale, internal representations in the striatum are proposed to develop via unsupervised reinforcement and demotion in winner-take-all mechanisms regulated by dopamine. The latter have the effect of allowing internal striatal representations to be highly flexible as long as appropriate sets of units (i.e. appropriately within either direct pathway or indirect pathway) come to be activated in relation to a given cortical context. Such flexibility could enable the system to remap around certain focal lesions as conjectured above. Work by Houk, Adams, and Barto [77] has raised the possibility that the BG allow reinforcement of behaviors that anticipate rather than directly generate reward. A variation of this mechanism has been proposed by O’Reilly and Frank [141] that also pushes the teaching effect of a reward earlier in time. These allow learning of associations between stimuli and actions when initially there are only coincidental rather than causal relationships. The current MI-MOAS model does not address this issue, i.e. the time credit assignment problem of reinforcement. Conceivably, incorporation of such mechanisms could further expand the model’s functionality to handle situations in which proper output is rewarded when it happens to occur spontaneously (or improper output is punished when it occurs). This should enable the model to address the operant type conditioning
of contingent decision tasks such as 1-2-AX treated by the model of O’Reilly and Frank [141].

The MIMOAS model may be related to the work by Brown, Bullock, and Grossberg on BG participation in saccade control [17]. The latter model is based on competitive queuing, and appears to pertain most specifically to circumstances where there exists a pre-formed cognitive representation of an intended behavioral sequence. In that model, the BG are argued to facilitate efficient program read-out. This proposal has been recently shown to be associated with distinct neuronal assemblies in prefrontal cortex [5, 20], as predicted by Grossberg, Bullock, Rhodes et al. [58, 157].

Prior to movement, each neuronal assembly maintains a different level of tonic activation, corresponding to a different component of the intended upcoming sequence. It seems likely that this situation corresponds to a state of full declarative or explicit knowledge of the upcoming task; in other words, there exists some consciousness of both the set of actions to be performed and their intended sequence. In the learning task of [150], however, procedural knowledge as indicated by shortened reaction time typically occurs well in advance of consciousness of the sequence. Correspondingly, the MIMOAS model depends only on real time specification of the actions to be performed based on interim subconscious state information.

Finally, the lateral cerebellum has been increasingly implicated in many cognitive functions [165, 166]. This would also be anticipated from the extensive reciprocal connections between the lateral cerebellum and the FC. The mechanism of cerebellar support of cognition has not been established and is not the focus of the MIMOAS model. However, our simulations have supported the conjecture that cerebellar filtering may help to sharpen and sustain the responses of FC modules and thereby critically facilitate their FC-BG interaction. Therefore, the MIMOAS model presents a generally similar but in places significantly different view of BG, FC, and cerebellar interaction from that presented by Houk and Wise [78]. Both models stress the importance of the BG in recognizing contextual information from cortex and triggering switches in the states of FC modules. However, the MIMOAS model emphasizes continuous BG support of both FC registers and lower motor output. On the other hand, the MIMOAS model does not attribute any gating or binary programming role
to the cerebellum. Rather, the cerebellum contributes to context-dependent continuous time, continuously valued precision filtering of signals that supports BG and FC interaction among other processes. Cerebellar systems are prime candidates for this role as these have been associated with timing functions [92] and cerebellar lesions result in loss of prosody [164]. The simple cerebellar coprocessing currently employed in the MIMOAS model currently addresses signal sharpening and maintenance (Section 2.1.2), but not timing. Further studies will be required to evaluate these alternatives.
Chapter 8

Conclusions

As primitive neural structures buried deep in the brain, the basal ganglia may be “smarter” than we used to think: Studies have shown that the BG are involved in a variety of brain functions, ranging from lower-level movement and posture control to higher-level cognitive decision making. Although the BG have produced a long time of strong research interests and there have been a tremendous amount of related studies, the need for a new neurobiological theory or model of BG function is still enormous. Many excellent computational models have been proposed to understand the BG operation, but they focused on selected aspects of BG function and have not sought to reconcile the operation of both lower-level motor control and higher-level cognitive processing.

This thesis is aimed at providing an integrated view of the BG function. The main contribution of the thesis is a multi-input multi-output adaptive switching (MIMOAS) model of the basal ganglia (see Chapter 2). In the model, BG circuitry effectively implements a large set of parallel noncompetitive logical OR and NOR circuits that can be driven by specific patterns of cortical activity. These afford selective gating of target thalamocortical neurons. This process can be considered as a general mapping between binary context and response vectors. The mapping is shown to be learnable via a reinforcement process that is consistent with actions commonly proposed for nigro-striatal dopaminergic pathways. Such BG learning contributes to the establishment of programmed actions as well as behavioral rules (such as “stop at red” and “go at green”) that become automatic enough for the brain to follow without much thought.
The thesis also investigates some theoretical issues concerning computation, learning, and dynamics of the BG under the framework of the MIMOAS model. It is proved that the proposed BG architecture ensures the basal ganglia to perform any transformation that associates cortical contexts with corresponding thalamocortical actions (see Chapter 3). Under learning rules that are consistent with the known lasting effect of dopamine and hypothetical synaptic normalization in the striatum, the BG behavior is shown to converge with expected output in response to specific cortical input (see Chapter 3).

The dynamics of the striatal network are investigated with an emphasis on the winner-take-all competition among neurons that are reciprocally connected with lateral inhibition (see Chapter 4). The general networks of this type are analyzed in terms of number and stability of the equilibria, including a condition derived for the existence of a globally asymptotically stable equilibrium. This condition depends upon the product of the lateral inhibitory strength and the maximal steepness of the neuronal activation functions. It is less conservative than that for uniqueness of equilibrium in more general Hopfield networks and holds for recurrent lateral inhibitory networks with arbitrary nondecreasing continuous activation functions. For recurrent networks with identical neuronal activation functions and symmetric lateral inhibition, it is shown that an equilibrium must exist that is order preserving with respect to the network input. Given this property, conditions are established that guarantee the dominance of activity of a single neuron at an order preserving equilibrium. This winner-take-all effect results from augmented separation of supra-threshold activities of competing neurons. For networks with discontinuous threshold activation functions, conditions for winner-take-all behavior, and a richer characterization of the equilibrium states can be derived using Filippov’s analysis. Under conditions that give rise to multiple equilibria, transitions between network equilibrium states (and thus outputs) resulting from shifts in inputs may show hysteresis which potentially affords these networks considerable robustness to input noise. This consequently explains the significance of dopamine and acetylcholine in the modulation of selectivity of competing corticostriatal activities.

The MIMOAS model attempts to provide a unified explanation for the role of
the BG and their interaction with the frontal cortex in higher-level cognitive decision making (see Chapter 5) and lower-level motor control (see Chapter 6). When the cortical context includes internal representations of system output, an autoregressive mechanism obtains that yields automatic play-out of implicitly programmed behavior. The model reproduces the experimentally observed sensitivity of procedural learning efficiency to sequence complexity and to the integrity of BG circuits. It is argued that this dependence is related to the ability of the system to form representations of higher-order context states within the frontal cortex and to switch these states with rapidity and precision. In the situation where the gated thalamocortical pathways effectively regulate input to motor cortex, the mechanism affords the ability to generate cruise movements. Simulated loss of dopaminergic activity in BG circuits yields the principal clinical features of Parkinson’s disease. Thus, the MIMOAS model subsumes within a broader framework, prior conceptions of the roles of the BG in controlling posture and movement speed, in acquiring and generating sequences of single actions, and in the selective sustenance of activation in working memory regions of frontal cortex. In so doing, it provides specific predictions regarding the activity of BG circuits in health and disease.

The computational BG model proposed in this thesis explores our understanding on the neural architecture and functions of the basal ganglia. It is with hope that further study based on the thesis would provide us with insights into the development of advanced clinical techniques for the diagnosis and treatment of neurological diseases involving dysfunction of the BG (such as the Parkinson’s disease, dystonia, Huntington’s disease, and schizophrenia). Furthermore, the study of biological control strategies of the BG would possibly bring attractive ideas to engineering applications like robotics and intelligent embedded systems.
Appendix A

Properties of Filippov Solutions

Appendix A presents the definition as well as properties of Filippov solutions to differential equations with discontinuous right hand side.

Definition A.1 (Filippov [46]) Consider a differential equation

\[
\frac{dx}{dt} = f(x, t), \quad (A.1)
\]

where \( f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \) is measurable and essentially locally bounded. A vector function \( x(\cdot) \) is called a solution of (A.1) on \( [t_0, t_1] \) if \( x(\cdot) \) is absolutely continuous on \( [t_0, t_1] \) and for almost all \( t \in [t_0, t_1] \)

\[
\frac{dx}{dt} \in \bigcap_{\delta > 0} \bigcap_{\mu N = 0} \text{conv} f(B(x, \delta) - N, t), \quad (A.2)
\]

where \( \mu \) is the Lebesgue measure, \( \bigcap_{\mu N = 0} \) denotes the intersection over all sets \( N \) of Lebesgue measure zero, \( B(x, \delta) \) represents a \( \delta \)-neighborhood of the point \( x \) in the space, \( \text{conv} E \) denotes the closure of convex hull of a set \( E \) (convex hull of \( E \) is the smallest convex set containing \( E \) ), and \( f(E) \) is the set of values that the vector function \( f(E) \) takes on the set \( E \).

So as for the absolutely continuous vector function \( x(t) \) to be a Filippov solution of (A.1), it is necessary and sufficient that for almost all \( t \) and for every vector \( e \) the following inequality hold [46]:

\[
\frac{dx}{dt} \cdot e \leq M_x \{ f(x, t) \cdot e \}, \quad (A.3)
\]
where $M_x\{\phi(x,t)\}$ is defined by $\lim_{\delta \to 0} \max_{x' \in B(x,\delta)} \{\phi(x',t)\}$, and $\max_{x' \in B} \{\phi(x',t)\}$ denotes the essential upper bound of $\phi(\cdot)$ on set $B$ if we neglect the values of the function $\phi(\cdot)$ on sets of measure zero: $\max_{x' \in B} \phi(x',t) = \inf_{\mu^N=0} \sup_{x' \in B-N} \phi(x',t)$. Especially, if denote $f(\cdot) = [f_1(\cdot), ..., f_n(\cdot)]^T$ and set $e = [0, ..., 0, \pm 1, 0, ..., 0]^T$ with the $i$-th element of $e$ being $\pm 1$ and other elements being $0$, then a Filippov solution $x(t)$ should satisfy

$$\frac{dx_i}{dt} \leq M_x\{f_i(x,t)\}, \quad (A.4)$$
$$-\frac{dx_i}{dt} \leq M_x\{-f_i(x,t)\}. \quad (A.5)$$

The below condition is required by the Existence Theorem and Comparison Theorem.

**Condition A.1** In an open or closed region $Q$ of the space $x, t$, the function $f(x,t)$ of (A.1) is defined almost everywhere in $Q$, is measurable, and for any bounded closed domain $D \subset Q$ there exists a summable function $A(t)$ (with well-defined and finite integral of $|A(t)|$ on $D$) such that almost everywhere in $D$ we have $|f(x,t)| \leq A(t)$.

**Theorem A.1 (Existence Theorem [46])** Let the right hand side of (A.1) be measurable in a region $Q$ and satisfies Condition A.1. Then for arbitrary initial conditions $x(t_0) = a$, where $(a,t_0) \in Q$, a solution of (A.1) exists satisfying these initial conditions and defined on the interval $[t_0 - d, t_0 + d]$, where $d$ is such that the $(n+1)$-dimensional “cylinder”

$$|t - t_0| \leq d, \quad \|x - a\| \leq \| \int_{t_0}^{t_0 \pm d} A(t) dt \| \quad (A.6)$$

is situated entirely inside $Q$.

Now consider a single (scalar) equation

$$\frac{dx}{dt} = f(x,t) \quad (A.7)$$

in a domain $Q$ of the $(x,t)$-plane. According to [46], if (A.7) satisfies Condition A.1, then among all solutions of this equation passing through an arbitrary fixed point
There is an upper solution \( \bar{x}(t) \) such that any solution \( x(t) \) passing through this point satisfies the inequality \( x(t) \leq \bar{x}(t) \).

**Theorem A.2 (Comparison Theorem [46])** If (A.7) and

\[
\frac{dx}{dt} = F(x,t)
\]

satisfy Condition A.1 in a domain \( Q \) and \( F(x,t) \geq f(x,t) \) almost everywhere in this domain, then any solution \( x(t) \) of (A.7) passing through the point \( (x_0,t_0) \) is not greater than the upper solution \( \bar{X}(t) \) of (A.8) through the same point when \( t \geq t_0 \).
# Appendix B

## Main Abbreviations Used in Text

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG</td>
<td>basal ganglia or basal ganglionic</td>
</tr>
<tr>
<td>CB</td>
<td>cerebellum or cerebellar</td>
</tr>
<tr>
<td>FC</td>
<td>frontal cortex or fontocortical</td>
</tr>
<tr>
<td>GABA</td>
<td>$\gamma$-aminobutyric acid</td>
</tr>
<tr>
<td>GP</td>
<td>globus pallidus</td>
</tr>
<tr>
<td>GPe</td>
<td>external segment of globus pallidus</td>
</tr>
<tr>
<td>GPi</td>
<td>internal segment of globus pallidus</td>
</tr>
<tr>
<td>LTD</td>
<td>long term depression</td>
</tr>
<tr>
<td>LTP</td>
<td>long term potentiation</td>
</tr>
<tr>
<td>MIMOAS</td>
<td>multi-input multi-output adaptive switching (model)</td>
</tr>
<tr>
<td>MPTP</td>
<td>1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine</td>
</tr>
<tr>
<td>MI</td>
<td>primary motor area</td>
</tr>
<tr>
<td>NMDA</td>
<td>N-methyl-D-aspartate</td>
</tr>
<tr>
<td>PD</td>
<td>Parkinson’s disease</td>
</tr>
<tr>
<td>PMdr</td>
<td>dorsal rostral prefrontal cortex</td>
</tr>
<tr>
<td>SMA</td>
<td>supplementary motor area</td>
</tr>
<tr>
<td>SN</td>
<td>substantia nigra</td>
</tr>
<tr>
<td>SNc</td>
<td>substantia nigra pars compacta</td>
</tr>
<tr>
<td>SNr</td>
<td>substantia nigra pars reticulata</td>
</tr>
<tr>
<td>SRT</td>
<td>serial reaction time (task)</td>
</tr>
<tr>
<td>STN</td>
<td>subthalamic nucleus</td>
</tr>
<tr>
<td>STR</td>
<td>striatum</td>
</tr>
<tr>
<td>VLo</td>
<td>ventrolateral nucleus (of thalamus)</td>
</tr>
<tr>
<td>VPLo</td>
<td>oral part of the ventral posterior lateral nucleus (of thalamus)</td>
</tr>
<tr>
<td>WTA</td>
<td>winner-take-all</td>
</tr>
</tbody>
</table>
Bibliography


[77] J. C. Houk, J. L. Adams, and A. G. Barto. A model of how the basal ganglia
generate and use neural signals that predict reinforcement. In J. C. Houk, J. L.
Davis, and D. G. Beiser, editors, Models of Information Processing in the Basal

[78] J. C. Houk and S. P. Wise. Distributed modular architectures linking basal
ganglia, cerebellum and cerebral cortex: their role in planning and controlling

Homeostatic synaptic plasticity can explain post-traumatic epileptogenesis in

lamocortical projections to the presupplementary motor area (pre-SMA) in the

and corticosubthalamic input zones from the presupplementary motor area in
the macaque monkey: comparison with the input zones from the supplementary
err in 2000.

bidge, MA, 1993.

[83] M. Ito. Cerebellar microcomplexes. In J. D. Schmahmann, editor, The Cere-

and pallidum of primates related to the execution of externally cued reaching

[85] J. Jankovic and E. Tolosa. Parkinson’s Disease and Movement Disorders.

hybrid long-loop control of upright balance. Biological Cybernetics, 91(3):188–

inputs to corticocortical and corticofugal neurons in areas 5 and 7 in the cat.

5 neuronal activity encodes movement kinematics, not movement dynamics.

[89] A. Kaneko. Physiological and morphological identification of horizontal, bipolar

[90] F. Kasanetz, R. Riquelme, and M. G. Murer. Disruption of the two-state mem-
brane potential of striatal neurons during cortical desynchronisation in anaes-


