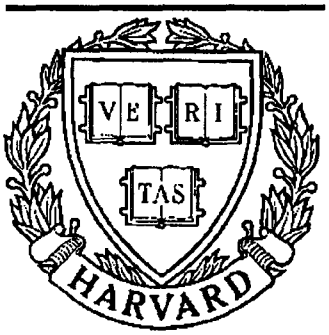
THESIS REPORT
Ph.D.

SYSTEMS
RESEARCH
CENTER



Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

Detection and Classification of Neural Signals and Identification of Neural Networks

by X. Yang
Advisor: S.A. Shamma

Ph.D. 89-2
Formerly TR 89-62

ABSTRACT

Title of Dissertation: DETECTION AND CLASSIFICATION OF NEURAL SIGNALS AND IDENTIFICATION OF NEURAL NETWORKS

Xiaowei Yang, Doctor of Philosophy, 1989

Dissertation directed by: Shihab A. Shamma, Assistant Professor, Electrical Engineering Department

This thesis aims to develop the theoretical and experimental means to study the nature of the neural networks of the nervous system. The most important parameters in a neural network are its synaptic connectivities (connection weights). Once the unknown connectivities in the nervous system are discovered, appropriate neural network models can be designed and used to mimic their action.

To study the functional connectivity, reliable recording and identification of the simultaneous activities of a group of neurons is essential. In the first part of this thesis, a system for neural spike detection and classification is presented, which does not require *a priori* assumptions about spike shape or timing. The system consists of two subsystems. The learning subsystem, comprising a Haar transform detection scheme, a feature learning phase and a template learning phase, extracts templates for each separable spike class. The real-time detection and classification subsystem identifies spikes in the noisy neural trace and sorts them into classes, according to the templates and the statistics of the background noise. Three fast algorithms are proposed for

the real-time sorting subsystem, and comparisons are made among different schemes. Performance of the system is illustrated by using it to classify spikes in segments of neural activity recorded extracellularly from monkey motor cortex and from guinea pig and ferret auditory cortices. The system is implemented without human supervision and therefore is suitable for real-time multichannel recording.

In the second part, analytical and experimental methods are provided for estimating synaptic connectivities from simultaneous recordings of multiple neurons (after separation). The results are based on detailed, yet flexible neuron models in which spike trains are modeled as general doubly stochastic point processes. The expressions derived can be used with nonstationary or stationary records, and can be readily extended from pairwise to multineuron estimates. Furthermore, we show analytically how the estimates are improved as more neurons are sampled, and derive the appropriate normalizations to eliminate stimulus-related correlations. Finally, we illustrate the use and interpretation of the analytical expressions on simulated spike trains and neural networks, and give explicit confidence measures on the estimates.

DETECTION AND CLASSIFICATION OF NEURAL SIGNALS AND IDENTIFICATION OF NEURAL NETWORKS

by

Xiaowei Yang

Dissertation submitted to the Faculty of the Graduate School
of The University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1989

Advisory Committee:

Professor Shihab A. Shamma, Chairman/Advisor

Professor Nariman Farvardin

Professor William S. Levine

Professor Prakash Narayan

Professor Herbert Levitan

DEDICATION

To Diming, Yuan and Jessica

ACKNOWLEDGMENTS

I wish to take this opportunity to express my sincere thanks to the people who made this study possible.

In particular, I would like to express the deepest gratitude to my advisor, Dr. Shihab A. Shamma, for his consistent encouragement, support, inspiration and guidance throughout the course of production of this dissertation. I thank him for the many hours of discussion and for his continuing effort to cultivate my ability to write clearly. It is my privilege to be his student, friend, and colleague. I would like to express my appreciation to Dr. James W. Fleshman for his contribution in the guinea pig and ferret experiments and for his constant help and valuable comments and suggestions. I would like to express my gratitude to Dr. Nariman Farvardin, Dr. William S. Levine, Dr. Prakash Narayan and Dr. Herbert Levitan for serving on my advisory committee, reading this dissertation and providing numerous helpful comments. I thank Drs. Narayan, Farvardin, and Levine for their frequent help, instructions, and suggestions and for course lectures that I have enjoyed so much. I would like to acknowledge all my fellow students at the Neural Systems Laboratory of the Systems Research Center for their help and for their effort to make the laboratory such an excellent research environment. Thanks are also due to Dr. E.M. Schmidt of the Laboratory of Neural Control, NINCDS, National Institutes of Health, for kindly providing the monkey motor cortex data.

A special debt of gratitude is owed to my parents, parents-in-law, and my sister for giving me the ambition to do this and for helping to take care of my children. My greatest debt is owed to my wife, Diming, for her abundant love and for never ending faith in me.

This work was performed at the Neural Systems Laboratory, and at the Communications and Signal Processing Laboratory of the Systems Research Center, University of Maryland. Some computational facilities were also provided by the Electrical Engineering Department, University of Maryland. I gratefully acknowledge partial support by grants from the Whitaker Foundation, the National Science Foundation, and the Naval Research Laboratory.

TABLE OF CONTENTS

List of Tables vii

List of Figures viii

1 Introduction 1

1.1 Machine Intelligence and Neurophysiology 1

1.2 Detection and Classification of Neuronal Signals 3

1.3 Connectivity Identification in Neural Networks 5

2 A Neural Spike Sorting System 8

2.1 Introduction 8

2.2 System Description 10

2.3 Detection by Haar Transformation 11

2.3.1 Neural Spike Detection 12

2.3.2 A Recursive Algorithm of Haar Transform 17

2.4 Feature Extraction and Classification 19

2.5 Real-Time Sorting 28

2.5.1 Multithreshold Scheme 28

2.5.2 Multiwindow Scheme 32

2.5.3 Soft-Decision Scheme 34

2.5.4 Comparisons Among the Schemes 37

2.6 Testing Examples 39

2.7 Summary 41

3 Identification of Connectivity in Neural Networks **50**

3.1 Introduction 50

3.2 The Neuron Model 53

3.3 Analytical Results 57

3.3.1 Further Relationships 61

3.3.2 Discussion of Result 1 63

3.3.3 Discussion of Result 2 64

3.3.4 Discussion of Result 3 65

3.3.5 Discussion of Result 4 71

3.4 Experimental Considerations 74

3.4.1 Using the Scatter Plot to Determine Neuronal Connectivities 76

3.4.2 Establishing Confidence Measures on the Estimates 77

3.5 Simulations and Discussion 80

4 Concluding Remarks **89**

Appendix 93

Bibliography 104

LIST OF TABLES

2.1 Performance Comparisons for Detection 42

2.2 Performance Comparisons for Classification 43

LIST OF FIGURES

2.1 Schematic Diagram of the System 44
2.2 Illustration of Haar Transform Detection 45
2.3 Spikes Detected by the Haar Transform 46
2.4 Neural Spikes Recorded by a Silicon-Based Multielectrode from Guinea Pig Auditory Cortex 47
2.5 Classes of Neural Spikes from Monkey Motor Cortex Identified by Multithreshold Sorting 48
2.6 One Second of Data Containing Neural Spikes from Ferret Auditory Cortex Separated by Soft-Decision Sorting 49

3.1 A Dynamical Nonlinear Neuron Model 83
3.2 A Pair of Neurons Stimulated by a Common Input Source 84
3.3 The Common Input Correlation vs. the Synaptic Correlation 85
3.4 Simulations for Pairwise Correlations 86
3.5 Interaction Among Three Neurons 87
3.6 Comparison of the Preferred with the Difference Normalizations .. 88

BIBLIOGRAPHY

CHAPTER
ONE
INTRODUCTION

1.1 Machine Intelligence and Neurophysiology

The brain is a powerful, versatile, and robust structure whose performance far exceeds that of any machine and algorithm currently in existence in such tasks as pattern recognition. It is estimated that the brain contains 10^{11} neurons that are organized in complex unknown networks. It is not yet understood how this system allows us to store, represent, retrieve and manipulate data such as speech, images, smells, sensations, and thoughts.

While neurophysiologists are revealing the mechanisms of perception, attention, thinking and other functions, engineers and physicists are attempting to mimic the parallel, distributed architecture of the mammalian nervous system, and hence build machines that can duplicate the performance of the brain. In order to do so, extensive research is underway to understand how the brain works and thus to formulate theories about how computations actually occur in nature. Entirely new computational paradigms are evolving that are based on simple models that are biologically influenced [14], [15], [18] and [19].

These systems have been used to solve difficult optimization problems and to implement associative memories [41], [21]. Other applications such as speech recognition, image data compression, adaptive pattern recognition, motion detection, as well as VLSI and simulation implementations are developing rapidly in the hope of achieving human-like performance.

To study the nervous system, one must observe neural activities. Suppose that we are given a record of neural activities of a group of neurons which is extracellularly recorded from a neural tissue (e.g. auditory cortex, visual cortex, etc.). The question arises as to what information can be obtained from it. There are many aspects to this question. Collections of neurons that might be influencing each other or are affected by common input sources of stimulation are often referred to as neural networks. Understanding the functioning of neural networks is one of the principal aims of theoretical neurophysiology. And engineering applications can be developed following the new discoveries in neurophysiology.

Our goal is to find the neurophysiologically realistic structure of these simultaneously recorded neurons. In other words, we want to discover how neurons are connected and organized in the given stimulus environment. Once these connectivities are discovered, appropriate neural network models can be designed which mimic the robustness, versatility, and sensitive responses of the mammalian nervous system.

To approach this goal, the first thing one must do is to separate activities recorded from several neuron. Involved in the separation task are the detection of neural signal from background noise and the classification of different signals. After the separation, one obtains spike trains from the original multiunit

recording, each train representing the activity of a single neuron. Then the connectivities between the neurons can be estimated through analysis of these spike trains.

This thesis consists of two stages. The first is to provide an automated system for the detection and classification of neural spikes so that the multi-neuron recordings will be separated into individual spike trains. The second is to develop analytical and experimental means for identification of neural network connectivities based on these spike trains so as to derive logical wiring diagrams in the network systems.

1.2 Detection and Classification of Neural Signals

In studying the functional connectivity of neural systems, reliable observation of the simultaneous activity of a group of neurons is essential. An extracellular electrode often records such electrical activity from several adjacent neurons. To analyze the contribution of each individual unit, one needs to distinguish the signals of each unit from the rest. In principle, signals from different neurons can be classified by their characteristic spike shapes. However, these shapes are often unpredictable functions of the neuron type, the electrode construction and placement, and the electrical characteristics of the intervening tissue [10]. In addition, multiunit recordings are always contaminated by noise, which comes both from external sources and from the weaker neural signals of more distant units.

There are several techniques available for the classification of multiunit neural signals [1], [4], [5], [9], [10], [12], [25], [26], [28], [29], [33], [34], [37]–[39]. Some techniques require prior information about spike shapes and epochs [4], [5], [9], [10], and [12], which is often not available; other methods employ time consuming computations [1], [9], [12], [25], [28], [29], [33], [34], [37] –[39], making real-time implementation impossible without specialized hardware; and others involve human supervision [4], [5], [26]. With the increasing availability of multi-channel extracellular microelectrode arrays [22] and the potential large number of simultaneous recordings, data processing capability and capacity will become very important, and on-line neural spike-separating techniques must strike a balance between performance and speed.

The fundamental motivation behind this work is the desire to overcome some of limitations presented by the newly developed extracellular microelectrode arrays. In most such electrodes, recording sites are fabricated on the same substrate. Consequently, the quality of the recordings obtained from a given channel can not be improved by a simple adjustment of the electrode position (as is the case with single electrodes) since this will necessarily affect the position (hence the recording quality) on all other channels. Therefore, the best strategy is to utilize these recordings the best instead of constantly attempting to adjust them. Another serious problem concerns the large number of recording sites and the presence of multiple spike shapes in every channel. Using traditional manual spike detection methods such as threshold and window discriminators to determine which channels carry useful information and to separate out the different spikes on each channel is tedious, extremely time consuming, and unlikely to succeed if adjustments have to be made during a recording session as is often the case.

Therefore, an efficient and convenient on-line multispike separation system will be the first stage to be developed which is totally automated from detection to classification, without presupposing any knowledge of spike shapes or of interspike intervals. Examples of multiunit extracellular recording from auditory cortex and primate motor cortex will be used to test the system.

After neural activities are separated for each neuron, the neural connectivity analyses are the next stage.

1.3 Connectivity Identification in Neural Networks

The central nervous system (which comprises the brain and the spinal cord), can be viewed as a highly complex communication system which receives, encodes, processes, transmits and retrieves a gigantic amount of information. Within a network, neurons contact each other via synapses. It is believed that information processing and transmission in the nervous system is carried by sequences of nerve-impulses called spikes generated by neurons. In other words, a neuron receives information from other neurons through synapses and sends its own spike train to others. The information-carrying parameters of a spike train are the time intervals between spikes, or the impulse rates. Given the spike trains of simultaneously recorded neurons, the question is what synaptic connectivity information can be obtained from them.

The synaptic connections between neurons have usually been estimated from inter-spike train correlations such as the cross-interval histogram, the cross correlogram and the joint peri-stimulus time (PST) scatter diagram all of which are stimulus-dependent. To destroy the stimulus effects, a shuffling method has also been used. The qualitative interpretations for these commonly used methodologies have been described in the literature [11], [13]. These histograms are statistical measures on the given spike trains which reflect the underlying neural network structure. By investigating these histograms, one may support a hypothesis as to whether the two neurons under study interact or not. One may also qualitatively infer that the two neurons have an excitatory or an inhibitory connection, or a common input. However, the choice between a direct connection and a common input is sometimes difficult [11]. Moreover, the time-averaging cross-interval and the time-averaging cross-correlation measures are not suitable for nonstationary neuronal firings. Therefore, a method of time-dependent histogram has been proposed to analyze nonstationary spike trains [40].

As a part of this work, the mathematical analysis will be established on a given neural net model based on stochastic processes where the dynamics and nonlinearity of each neuron is explicitly represented. Various structures in the network will be simulated so that the investigations can be made in different aspects. We will model spike trains as the realizations of general doubly stochastic point processes. The objectives in this stage are (1) to develop analytical and experimental methods to estimate synaptic connectivities from simultaneous recordings of multiple neurons, (2) to express synaptic connectivity in terms of probability densities of joint neuronal firings and individual neuronal firings, (3) to extend the method from pair-wise to multiunit

correlations, and (4) to choose the appropriate normalization method used in the correlation histograms, (5) to indicate the invalidity of time-averaging correlation histograms in nonstationary neuronal firings.

CHAPTER
TWO

A NEURAL SPIKE SORTING SYSTEM

2.1 Introduction

An important aspect in study of neurobiological system is the reliable observation of groups of neurons. When an extracellular electrode is used to record electrical neural activities from several adjacent neurons, the spike waveforms attributable to each single nerve cell can be distinguished. By detecting each of the spikes and classifying each according to its shape, the record of activity of each of the single neural units can be recreated from the original multiunit recordings. One can analyze each contributing unit independently to obtain the equivalent several simultaneous single unit recordings. If the units are considered as an interactive group, correlations between the units can be studied to discover complex information encodings not apparent in individual unit activity. The correlation analyses are means of revealing physiology of the neural network being observed.

As stated in section 1.2, it is desirable to use multichannel microelectrode array to record large groups of neurons. The overriding goal of the spike sorting

algorithm to be used with these multielectrode arrays is *not* so much to detect the smallest spikes in the midst of noisy traces, but rather to isolate the most reliable spikes with no or minimal human intervention.

We propose an efficient and convenient on-line multispike separation system which is totally automated from detection to classification, without presupposing any knowledge of spike shapes or of interspike intervals. The system is a software realization described in section 2.2, and it consists of two parts: a learning module which extracts templates for each class of spikes in the neural trace, and a real-time classification module which detects the spikes and sorts them into classes. At the beginning of the learning module, no knowledge of spike templates is available. Hence a Haar transformation is first performed to locate the occurrences of spikes, based on the fact that there is similarity between Haar transform bases and spike shapes. The Haar transform detection scheme is proposed in section 2.3.

The next step in classification problem is to extract features of entities so that the classification can be done based on these well-selected features within minimum errors. Section 2.4 describes a general approach to this problem. After features are ready, we shall briefly review that a matched filter can perform the optimal classification in the additive Gaussian case. However, the matched filter requires so many computations that it can hardly apply to on-line multichannel situation. Hence one needs more economic algorithms which have performance as nearly good as the matched filter has.

We propose three algorithms which meet our requirements. The multi-threshold detection scheme is presented in subsection 2.5.1. The multiwindow threshold discriminator is described in subsection 2.5.2, and the analysis of

the soft-decision discriminator is detailed in subsection 2.5.3. Any of these algorithm can be used in the real-time sorting module. Comparisons for the different schemes is given in subsection 2.5.4. In section 2.6, performance of the system is illustrated by using it to classify spikes in segments of neural activity recorded from monkey motor cortex and from guinea pig and ferret auditory cortexes. The system is implemented without human supervision and therefore is suitable for real-time multichannel recording.

2.2 System Description

The system is divided into two parts as indicated in the block diagram of Fig. 2.1. The first part is a learning subsystem which extracts templates of spikes for every class. This subsystem includes a feature learning phase and a template learning phase. The second part is a real-time detection and classification subsystem which detects spikes in the noisy trace and sorts them out into classes, based on the templates that the learning subsystem provides and the statistics of the background noise.

Assume that a segment of sampled data containing spikes from several neurons is stored in a memory buffer. The learning subsystem begins with the detection of spikes in the noise, using the discrete Haar transformation (DHT). In the feature learning phase, specific features — the peak-to-peak amplitude and the peak-to-peak time interval — are measured for each spike detected and are used to construct the feature histogram. The histogram serves to determine

the number of separable unit classes and typical feature values of each class. In the template learning phase, the same DHT detection scheme is applied to detect spikes which are sorted into classes by comparing their features with the typical features. As a result, typical templates for each class are formed by averaging all the classified spikes.

The real-time detection and classification subsystem is supported by either a multithreshold sorting scheme, or a multiwindow sorting scheme, or a soft-decision sorting scheme. Once the optimal threshold parameters for all classes are derived, based on the templates and the background noise, the real-time on-line processing begins.

2.3 Detection by Haar Transformation

In this section, we present a detection method for neural signals. The neural signals we are dealing with are action potentials produced by cell body. Because the action potentials are nerve electrical impulses which occur abruptly and elapse quickly, they are often called spikes. All spikes generated by the same neuron have essentially the same shape. The information is carried by the time intervals between spikes and the firing rate of spikes. Theoretically, a spike train is modeled as a series of impulses with random intervals called a realization of a stochastic point process. The task of spike detection is to convert noisy spike trains into the abstract realizations of point processes.

First, the Haar transform detection (HTD) is proposed in subsection 2.3.1. Because of the spike-like characteristics of the transform bases, the Haar transform detects spikes without knowledge of spike templates. Although it is very powerful for detecting neural spikes, the HTD has a computational requirements that do not allow a current mini-computer to perform on-line in real-time, even with a fast Haar transform algorithm. Therefore, a recursive algorithm for performing on-line discrete Haar transform is developed in subsection 2.3.2.

2.3.1 Neural Spike Detection

The discrete Haar transform detection scheme plays a key role in the learning subsystem. Similar to the Fourier transform, the Haar transform is an orthogonal transformation with spike-like bases. The Haar function [16], originally proposed in 1910 by Alfred Haar, a famous Hungarian mathematician, is defined in [0, 1) as

$$har(r,\ m,\ t)\ =\ \begin{cases} 2^{\frac{r}{2}}, & \frac{m-1}{2^r}\ \leq\ t\ <\ \frac{m-1/2}{2^r} \\ -2^{\frac{r}{2}}, & \frac{m-1/2}{2^r}\ \leq\ t\ <\ \frac{m}{2^r} \\ 0, & otherwise \end{cases} \tag{2.1}$$

and

$$har(0,\ 0,\ t)\ =\ 1,\quad t\ \in\ [\,0,\ 1\,)$$

where $r = 0,\ 1,\ \cdots,\ log_2 N$, and $m = 1,\ 2,\ \cdots,\ 2^r$.

It is shown that the Haar function $\{har(n,m,t)\}$ is orthogonal for any N which is a power of two. If $\{har(n,m,t)\}$ is sampled at rate N, then an $N \times N$ orthogonal matrix is obtained [2]

$$\mathbf{W} = \begin{pmatrix} w_0(0) & w_0(1) & \cdots & w_0(N-1) \\ w_1(0) & w_1(1) & \cdots & w_1(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ w_{N-1}(0) & w_{N-1}(1) & \cdots & w_{N-1}(N-1) \end{pmatrix} \tag{2.2}$$

where each element is a sample of the Haar function

$$w_k(i) = w_{2^m+n-1}(i) = \begin{cases} 0, & i = 0, \cdots, \frac{N(n-1)}{2^m} - 1 \\ 2^{\frac{m}{2}}, & i = \frac{N(n-1)}{2^m}, \cdots, \frac{N(n-1/2)}{2^m} - 1 \\ -2^{\frac{m}{2}}, & i = \frac{N(n-1/2)}{2^m}, \cdots, \frac{Nn}{2^m} - 1 \\ 0, & i = \frac{Nn}{2^m}, \cdots, N-1 \end{cases} \tag{2.3}$$

and

$$w_0(i) = 1, \quad i = 0, 1, \cdots, N-1$$

for $k = 1, 2, \cdots, N-1$, where the ranges of m and n are $0 \leq m \leq log_2 N - 1$ and $1 \leq n \leq 2^m$, respectively. Every row of the transform matrix is a basis of the Haar transform. Notice that the inverse Haar matrix is

$$W^{-1} = \frac{1}{N} W^T \tag{2.4}$$

where T denotes the transpose.

Let us denote the sampled time trace as $\mathbf{x} = [x(0),\ x(1),\ \cdots,\ x(N-1)]^T$ and the discrete Haar transformed sequence as $\mathbf{y} = [y_0,\ y_1,\ \cdots,\ y_{N-1}]^T$, they

are related by $\mathbf{y} = \mathbf{W}\mathbf{x}$. In transforming the original neural trace, a large component appears in the Haar-transformed domain if the basis for that component is similar in width and phase to the spike in the time trace. By thresholding these outstanding components, spikes are detected in the transformed domain. To determine the corresponding time of the spike occurrences, we do the following. First, a filtering procedure in the transform domain is performed (which will be described by the following second example). Then the filtered sequence is transformed back to the original time domain, resulting in noise-free step-like spikes at their original time in the trace. Finally, time of occurrence is defined by the zero-crossing of the step-like spike. The following simple cases shows how the Haar transform locates a spike.

Example 1: An ideal spike.

An ideal spike is an artificial spike with a bipolar rectangular shape which is similar to a Haar transform basis. If an ideal spike occurs at the beginning of a trace,

$$x(i) = \begin{cases} a, & i = 0, 1 \\ -a, & i = 2, 3 \\ 0, & otherwise \end{cases}$$

there is a pulse corresponding to the spike in the transform domain,

$$y_k = \begin{cases} a_N\ , & k = N/4 \\ 0\ , & otherwise \end{cases}$$

where $a_N = a\sqrt{N}$.

Example 2: A more realistic spike.

Figure 2.2 illustrates this example. An artificial spike has the shape

$$x(i) = \begin{cases} a\ sin[2\pi(i-3)/5]\ , & i = 4,\ 5,\ 6,\ 7 \\ 0\ , & otherwise. \end{cases}$$

The resulting transformed sequence becomes

$$y_k = \begin{cases} 1.54\ a_N\ , & k = N/4+1 \\ 0.36\ a_N\ , & k = N/2+2 \\ -0.36\ a_N\ , & k = N/2+3 \\ 0\ , & otherwise \end{cases}$$

where, again, $a_N = a\sqrt{N}$. It is seen that the largest component ($1.54a_N$) occurs in the transform domain corresponding to the basis having the same width as the spike has. There are "harmonic" components (0.36 a_N, and -0.36 a_N) in high portion of the transform domain due to the imperfectness of the spike. This example corresponds to the case where no noise is present. Experimentally, there is always an additive noise process in the background of neural recordings. The noise may be removed by a "threshold filtering" method, using a properly selected threshold in the transform domain. To perform the threshold filtering, set every component below the threshold equal to zero and leave components above the threshold unchanged. Then the threshold filtered sequence is transformed back to the original time domain, resulting in noise-free, step-like spikes at their original positions in the trace. Thus time of occurrence is defined by the zero-crossing of the step-like spikes.

Let us continue to use Example 2 to illustrate how the procedure works. The threshold t in the transform domain is set to a value, say, $|t| = 0.5a_N$, so that the filtered transformed sequence $\tilde{\mathbf{y}} = [\tilde{y}_0,\ \tilde{y}_1,\ \cdots,\ \tilde{y}_{N-1}]^T$ is expressed as

$$\tilde{y}_k = \begin{cases} 1.54\ a_N\ ,\ \ k\ =\ N/4+1 \\ \\ 0\ ,\ \ otherwise \end{cases}$$

which is transformed back to the time domain by $\tilde{\mathbf{x}} = \mathbf{W}^{-1}\tilde{\mathbf{y}}$, yielding the time sequence $\tilde{\mathbf{x}} = [\tilde{x}(0),\ \tilde{x}(1),\ \cdots,\ \tilde{x}(N-1)]^T$ as

$$\tilde{x}(i) = \begin{cases} 0.77\ a\ ,\ \ i\ =\ 4,\ 5 \\ \\ -0.77\ a\ ,\ \ i\ =\ 6,\ 7 \\ \\ 0\ ,\ \ otherwise. \end{cases}$$

It is seen that the reconstructed spike appears at the original time with a step-like shape (ideal in this case). We can also image that if there were noise, the noise amplitude in the transform domain below threshold would be filtered out. One likes the threshold to be large when the noise level is high, and vice versa. In fact the threshold for the filter can be selected according to the noise variance. Assume that the noise sequence $\{n(i)\}_{i=0}^{N-1}$ is independent and identically distributed (iid) with common Gaussian distribution $\mathcal{N}(0,\ \sigma^2)$. It can easily be shown that if $\{\ y_k\ \}$ is the transformation of the noise only, then the components of $\{y_k\}_{k=0}^{N-1}$ are identically distributed with common Gaussian distribution $\mathcal{N}(0,\ N\sigma^2)$. Since $P_r(|y_k| \geq 3.0\sigma\sqrt{N}) < 0.0028$, and $P_r(|y_k| \geq 4.1\sigma\sqrt{N}) = 0.000042$, a nonlinear transform domain threshold is set to be $t = C(\sigma)\sigma\sqrt{N}$, where $C(\sigma)$ is a nondecreasing function varying within range [3.0, 4.1]. Thus the probability that $|y_k|$ exceeds the threshold t is very

small. The noise will, therefore, be filtered out by setting every component below the threshold to be zero.

Figure 2.3 demonstrates the HTD procedure. A segment of neural data from monkey motor cortex extracellular recording in Fig. 2.3(a) was transformed by the discrete Haar transform. In the transform domain (Fig. 2.3(b)) there are two large components in the low "frequency" portion and two in the high "frequency" portion, indicating two spikes in the original trace. After filtering out noise in the transform domain, the spikes were reconstructed without the high frequency components and noise, as shown in Fig. 2.3(c).

The Haar transform is a powerful tool for detecting neural spikes without knowledge of spike templates. The DHT at this stage serves to detect spikes with the goal of accumulating enough of them to construct the histogram and templates for the totally automated spike classification system.

2.3.2 A Recursive Algorithm of Haar Transform

To perform an N point discrete Haar transform (DHT), one needs $N(log_2 N - 1)$ multiplications and $N log_2 N$ additions. Even for the fast algorithm by Andrews and Caspari [3], one still needs N multiplications and $2(N-1)$ additions. These expensive computational costs prohibit a real-time implementation. To overcome this problem, we propose a recursive algorithm for the Haar transform.

Suppose that a spike is represented by an M-point template. We look through an M-point-wide window to observe signals when the spike trace keeps moving to the left (or the window is sliding to the right). If an M-point DHT

is performed, a large component at the 2-nd transform basis is expected. If a $2M$-point DHT is performed, then we expect a large component at the 4-th transform basis first, and it will appear at the 3-rd basis after M shifts. If a $4M$-point DHT is performed, then we expect a large component at the 8-th transform basis first, then it will appears at the 7-th basis, the 6-th basis, and finally the 5-th basis, after every M shifts.

Let $\mathbf{x}$ denote an N-point segment of the spike trace and $\mathbf{y}$ the corresponding transformed data. The components of $\mathbf{y}$ can be expressed as

$$y_k = \sum_{i=0}^{N-1} x(i) w_k(i) \ , \quad k = 0, 1, \cdots, N-1. \tag{2.5}$$

Let $\mathbf{u}$ be the shifted version of $\mathbf{x}$, i.e.,

$$u(i) = x(i+1), \quad \forall\, i \tag{2.6}$$

and $\mathbf{v}$ be the transformed data of $\mathbf{u}$, then

$$v_k = \sum_{i=0}^{N-1} u(i) w_k(i)$$

$$= y_k + (\sqrt{2})^m [2u(\frac{(n-1/2)N}{2^m} - 1) - u(\frac{(n-1)N}{2^m} - 1) - u(\frac{nN}{2^m} - 1)],$$

$$k = 1, 2, \cdots, N-1, \tag{2.7}$$

$$v_0 = y_0 + u(N-1) - u(-1)$$

where m, n and k are related by $k = 2^m + n - 1$, $1 \leq n \leq 2^m$, and $0 \leq m \leq \log_2 N - 1$. This is the recursive form of the Haar transformation. At each shift, we compute the current Haar transform data v_k based on the previous Haar transform data y_k and up-to-date information

$$(\sqrt{2})^m [2u(\frac{(n-1/2)N}{2^m} - 1) - u(\frac{(n-1)N}{2^m} - 1) - u(\frac{nN}{2^m} - 1)].$$

For every v_k ($k \neq 0, 1$), one needs one multiplication and four additions; for v_1, four additions only. But it is not necessary to calculate v_k for every k. Only v_1 is needed, for example, if an M-point Haar transform is performed. Or only v_2 and v_3 are useful for a $2M$-point transform, etc. In this way, one can avoid a lot of computations and obtain a real-time performance.

In contrast to the long sequence Haar transform, the recursive algorithm detects one spike at a time. This may be another advantage of the algorithm besides the computational savings.

2.4 Feature Extraction and Classification

Signals from different neurons can be classified by characteristic spike shapes, which are unpredictable functions of the neuron type, electrode construction, placement, and the electrical characteristics of the intervening tissue [10]. In order to classify spikes, one may either use the complete templates, or instead base it on selected unique features of those spikes. In principle, spikes belonging to the same class possess relatively similar features.

The features are carefully selected so that they can best represent spikes of a class. After enough spikes are collected, an M-dimensional feature histogram $h(i_1, i_2, \cdots, i_M)$ is generated with index i_k as the ordered bin number for the k-th feature. If the resolution of the bins is high enough, the histogram will approximately reflect the joint probability density function (pdf) $f(\xi_1, \xi_2, \cdots, \xi_M)$ of the M features. Let us denote $\xi = (\xi_1, \xi_2, \cdots, \xi_M)$ as the feature vector and

$f(\xi) = f(\xi_1, \xi_2, \cdots, \xi_M)$ as the joint pdf of the features. By the motivation of the maximum likelihood estimation, we pick the particular value of the vector ξ at which $f(\xi)$ has a peak as the typical features for a class. We consider only those peaks that are separated by a minimum distance ϵ. For example, suppose that $f(\xi)$ has peaks at ξ' and ξ''. We say that peaks at ξ' and at ξ'' represent different units, if

$$\frac{||\xi' - \xi''||}{\min(||\xi'||, ||\xi''||)} > \epsilon \tag{2.8}$$

where $\epsilon > 0$ is a specific value. Peaks within the minimum distance are considered as the same class with the features at which $f(\xi)$ has the highest peak.

An important question which arises here is how accurate are the typical feature values determined in this way, assuming there are no classification errors? Suppose that there are S classes with *a priori* probabilities P_i, $i = 1, 2, \cdots, S$, and that the histogram is made from n observations of feature ξ , of which approximately $n_i = nP_i$ observations belong to class i. The feature has a pdf

$$f(\xi) = \sum_{i=1}^{S} P_i f_i(\xi) \tag{2.9}$$

where $f_i(\xi)$, is the pdf of ξ under class i. In the Gaussian case, let ξ have a Gaussian distribution with unknown mean μ_i and variance σ_i^2 under class i. Thus $f_i(\xi)$ has its peak at $\xi = \mu_i$. Therefore, according to the feature selection criterion, μ_i is chosen as the typical feature value for class i. The maximum likelihood estimator of μ_i is the sample mean of class i, and it can be written as

$$\hat{\mu}_i = \frac{1}{n} \sum_{\xi_k \epsilon C_i} \xi_k \, . \tag{2.10}$$

If one uses the confidence interval estimation argument, it is easy to see that

$$P_r(\hat{\mu}_i - \frac{\sigma}{\sqrt{n_i}}t_\alpha \leq \mu_i \leq \hat{\mu}_i + \frac{\sigma}{\sqrt{n_i}}t_\alpha) = 1 - \alpha \,. \tag{2.11}$$

In words, one has $(1-\alpha)100$ percent confidence that the true value μ_i is in the interval $[\ \hat{\mu}_i - \frac{\sigma}{\sqrt{n_i}}t_\alpha,\ \ \hat{\mu}_i + \frac{\sigma}{\sqrt{n_i}}t_\alpha\ ]$. The length of the interval is $2\sigma t_\alpha/\sqrt{n_i}$, where t_α is chosen so that $P_r(|N(0,1)| \leq t_\alpha) = 1-\alpha$, where $N(0,1)$ is a normalized Gaussian random variable. Clearly, $\hat{\mu}_i$ is a consistent estimator of μ_i . The final step is to show that $\hat{\mu}_i$ is the value at which $f_i(\xi)$ has a peak with probability 1. This is true because $E(\hat{\mu}_i) = \mu_i$ and $Var(\hat{\mu}_i) = \sigma_i^2/n_i$. The strong law of large numbers says that $\hat{\mu}_i$ converges to μ_i almost surely, and it is apparent that $P_r(\lim_{n_i\to\infty} Var(\hat{\mu}_i) = 0) = 1$; i.e., $\hat{\mu}_i$ will approach μ_i with no variation so that $f_i(\hat{\mu}_i)$ is definitely the peak as n_i goes to infinity. Now, it is in confirmation of Gaussian cases that one can make the typical feature values as accurate as one wants providing that the sample size n is sufficiently large.

In the case of insufficient sample size, local averaging is used to smooth the histogram. The resulting histogram is revised by

$$\bar{h}(i_1, i_2, \cdots, i_M,) = \frac{1}{W} \sum_{(\mathbf{j})\in N_b[(\mathbf{i}),\, r]} w_{\mathbf{i}-\mathbf{j}} h(j_1, j_2, \cdots, j_M) \tag{2.12}$$

where

$$W = \sum_{(\mathbf{j})\in N_b[(\mathbf{0}),\, r]} w_{\mathbf{j}}$$

and $N_b[(\mathbf{i}),\ r]$ is defined as the r-neighborhood of $(\mathbf{i})$ (the ball centered at $\mathbf{i} = (i_1, i_2, \cdots, i_M)$ with radius r).

To classify spikes, one may consider either the complete templates, or the selected unique features of these templates as signals. In general, it is a multiple hypothesis testing problem and one likes to make the assumption that the observation data are additively composed by signal and noise. That is, the observed time series $\mathbf{x} = \{x(1), x(2), \cdots, x(M)\}$ under hypothesis $\mathcal{H}_i$ is represented as

$$x(m) = s_i(m) + n(m), \quad m = 1, 2, \cdots, M \tag{2.13}$$

where $\{s_i(m)\}$ is the template (or partial template) of class i, and $\{n(m)\}$ is the contaminating noise.

The classification problem is stated as follows. S classes are assumed to have a prior probabilities P_i, $i = 1, 2, \cdots, S$, and the noise class has a prior probability P_0. Using the risk function concept [36], we associate a risk r_{ij} with choosing class i when the correct classification is class j. The average risk is given by

$$R = \sum_{i=0}^{S} \sum_{j=0}^{S} r_{ij} P_j P_r(D_i|\mathcal{H}_j) \tag{2.14}$$

where $P_r(D_i|\mathcal{H}_j)$ is the probability of choosing class i when the true hypothesis is $\mathcal{H}_j$. The feature space A is partitioned into exhaustive and mutually exclusive subset A_i, $i = 0, 1, \cdots, S$, so that spikes with feature $\mathbf{x}$ are classified to be in class i if $\mathbf{x} \in A_i$. Noting that $A_i = A - \bigcup_{j \neq i} A_j$ and $\int_A f_j(\mathbf{x}) d\mathbf{x} = 1$, $j = 0, 1, \cdots, S$, the average risk can be expressed as

$$R = \sum_{i=0}^{S} \sum_{j=0}^{S} r_{ij} P_j \int_{A_i} f_j(\mathbf{x}) d\mathbf{x}$$

$$= \sum_{i=0}^{S} r_{ii} P_i + \sum_{i=0}^{S} \int_{A_i} \sum_{j \neq i}^{S} (r_{ij} - r_{jj}) P_j f_j(\mathbf{x}) d\mathbf{x}$$

$$= \sum_{i=0}^{S} r_{ii} P_i + \sum_{i=0}^{S} \int_{A_i} \lambda_i(\mathbf{x}) d\mathbf{x} \tag{2.15}$$

where $\lambda_i(\mathbf{x}) = \sum_{j \neq i}^{S} (r_{ij} - r_{jj}) P_j f_j(\mathbf{x})$ ($i = 0, 1, \cdots, S$) are the decision variables. In order to minimize the risk R, we like to minimize the integrals $\int_{A_i} \lambda_i(\mathbf{x}) d\mathbf{x}$ for each class. It can be done by the following. The observed value $\mathbf{x}$ is claimed to be in class i, if $i = arg \min_j \{\lambda_j(\mathbf{x})\}$. Therefore, the optimal partition of A is

$$A_i^* = \{\mathbf{x} : \lambda_i(\mathbf{x}) \leq \lambda_j(\mathbf{x}), \forall j\}, \quad i = 0, 1, \cdots, S. \tag{2.16}$$

It can be shown that under minimum error probability criterion, the matched filter performs the optimal classification if the noise is white Gaussian. The risk R is equal to the probability of error P_e if the risk r_{ij} is set to be

$$r_{ij} = \begin{cases} 1, & i \neq j \\ 0, & i = j. \end{cases} \tag{2.17}$$

In this case we have

$$P_e = R = \sum_{i=0}^{S} \int_{A_i} \lambda_i(\mathbf{x}) d\mathbf{x} \tag{2.18}$$

where $\lambda_i(\mathbf{x})$ is reduced to

$$\lambda_i(\mathbf{x}) = \sum_{j \neq i}^{S} P_j f_j(\mathbf{x}) = \sum_{j=1}^{S} P_j f_j(\mathbf{x}) - P_i f_i(\mathbf{x}), \quad i = 0, 1, \cdots, S. \tag{2.19}$$

Notice that the first term is invariant for i, hence $\lambda_i(\mathbf{x}) \leq \lambda_j(\mathbf{x})$ is equivalent to $P_i f_i(\mathbf{x}) \geq P_j f_j(\mathbf{x})$. Define the decision variable $d_i(\mathbf{x})$ as

$$d_i(\mathbf{x}) = \sigma^2 \ln(P_i \frac{f_i(\mathbf{x})}{f_0(\mathbf{x})}), \tag{2.20}$$

the optimal partition can be expressed in terms of the decision variable:

$$A_i^* = \{\mathbf{x} : d_i(\mathbf{x}) \geq d_j(\mathbf{x}),\ \forall\, j\},\ \ i = 0, 1, \cdots, S. \tag{2.21}$$

Under the assumption that the noise is white Gaussian distributed with zero-mean and variance σ^2, the decision variable can be simplified as

$$d_i(\mathbf{x}) = \sigma^2 \ln P_i + \frac{1}{2}\sum_{m=1}^{M}(2s_i(m)x(m) - s_i(m)^2) = \sigma^2 \ln P_i - \frac{1}{2}||\mathbf{s_i}||^2 + <\mathbf{s_i}, \mathbf{x}>. \tag{2.22}$$

Therefore, the decision variable is nothing but the output of a matched filter

$$d_i(\mathbf{x}) = P_i^o + <\mathbf{s_i}, \mathbf{x}>,\ \ i = 0, 1, \cdots, S \tag{2.23}$$

where

$$P_i^o = \sigma^2 \ln P_i - \frac{1}{2}||\mathbf{s_i}||^2,\ \ i = 0, 1, \cdots, S. \tag{2.24}$$

For performance measurement, we now compute the probability of error for the matched filter. Since

$$P_e = 1 - \sum_{i=0}^{S} P_i P_c^*(i)\,, \tag{2.25}$$

let us compute the probability of the optimal correct decision $P_c^*(j)$ for each hypothesis $\mathcal{H}_j$. This is the probability that $d_j(\mathbf{x})$ exceeds all other decision variables $d_i(\mathbf{x})$ under hypothesis H_j, and it can be expressed as

$$P_c^*(j) = P_r(d_i(\mathbf{x}) \leq d_j(\mathbf{x}),\ \forall\, i | H_j) \tag{2.26}$$

$$= E_j[P_r(d_i(\mathbf{x}) \leq d_j(\mathbf{x}),\ \forall\, i | d_j(\mathbf{x}), H_j)]. \tag{2.27}$$

We see that $d_0(\mathbf{x}) = P_0^o = \sigma^2 ln P_0$ is a constant. And under hypothesis H_j, the expectation and the variance of $d_i(\mathbf{x})$ are written as

$$\mu_{ij} \;=\; E_j[d_i(\mathbf{x})] \;=\; P_i^o + < \mathbf{s_i}, \mathbf{s_j} >, \;\; i = 1, 2, \cdots, S \tag{2.28}$$

and

$$\sigma_{ij}^2 \;=\; Var_j[d_i(\mathbf{x})] \;=\; \sum_{m=1}^{M} \sum_{n=1}^{M} Cov_j(x(m), x(n)) \;=\; \sigma^2 ||\mathbf{s_i}||^2, \;\; i = 1, 2, \cdots, S \tag{2.29}$$

respectively. By the same token, the covariance of $d_k(\mathbf{x})$ and $d_l(\mathbf{x})$ is

$$c_{kl} \;=\; Cov_j(d_k, d_l) \;=\; \mathbf{s_k^T} Cov_j(\mathbf{x}) \mathbf{s_l} \;=\; \sigma^2 < \mathbf{s_k}, \mathbf{s_l} > . \tag{2.30}$$

Hence under hypothesis H_j, the decision variables obey the Gaussian distribution

$$\begin{pmatrix} d_1(\mathbf{x}) \\ d_2(\mathbf{x}) \\ \vdots \\ d_S(\mathbf{x}) \end{pmatrix} \;\sim\; \mathcal{N}(\mu_\mathbf{j}, \mathbf{\Sigma_d}) \tag{2.31}$$

where the mean value is

$$\mu_\mathbf{j} \;=\; \begin{pmatrix} \mu_{1j} \\ \mu_{2j} \\ \vdots \\ \mu_{Sj} \end{pmatrix} \;, \;\; j = 0, 1, \cdots, S \tag{2.32}$$

and the covariance matrix is

$$\mathbf{\Sigma_d} \;=\; \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1S} \\ c_{21} & c_{22} & \cdots & c_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ c_{S1} & c_{S2} & \cdots & c_{SS} \end{pmatrix} \;, \;\; j = 0, 1, \cdots, S \tag{2.33}$$

with $c_{ii} = \sigma_i^2$, $i = 1, 2, \cdots, S$. The probabilities of the optimal correct decisions can be expressed further with these statistical quantities

$$P_c^*(0) = \frac{1}{(2\pi)^{S/2}|\Sigma_d|^{1/2}} \int_{-\infty}^{d_0} \cdots \int_{-\infty}^{d_0} \exp\left[-\frac{1}{2}(\mathbf{t}-\mu_\mathbf{0})^\mathbf{T}\mathbf{\Sigma_d^{-1}}(\mathbf{t}-\mu_\mathbf{0})\right] dt_1 \cdots dt_S, \tag{2.34}$$

$$P_c^*(j) =$$

$$\int_{-\infty}^{\infty} P_r(d_0 \le \tau,\ d_1 \le \tau,\ \cdots,\ d_{j-1} \le \tau,\ d_j = \tau,\ d_{j+1} \le \tau,\ \cdots,\ d_S \le \tau\ |H_j) d\tau$$

$$= \frac{1}{(2\pi)^{S/2}|\Sigma_d|^{1/2}} \int_{d_0}^{\infty} \int_{-\infty}^{\tau} \cdots \int_{-\infty}^{\tau} \exp\left[-\frac{1}{2}(\mathbf{t}-\mu_\mathbf{j})^\mathbf{T}\mathbf{\Sigma_d^{-1}}(\mathbf{t}-\mu_\mathbf{j})\right]|_{t_j=\tau}$$

$$\cdot\, dt_1 \cdots dt_{j-1} dt_{j+1} \cdots dt_S d\tau\ ,\quad j = 1, 2, \cdots, S. \tag{2.35}$$

Consider a special case where all templates are mutually orthogonal, resulting in the covariance matrix Σ_d to be diagonal. The probabilities of the optimal correct decisions $P_c^*(i)$ can be reduced further to

$$P_c^*(0) = \prod_{i=1}^{S} \Phi\left(\frac{d_0 - \mu_{i0}}{\sigma_i}\right) \tag{2.36}$$

and

$$P_c^*(j) = \int_{d_0}^{\infty} \left[\prod_{i=1, i\neq j}^{S} \Phi\left(\frac{\tau - \mu_{ij}}{\sigma_i}\right)\right] \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2\sigma_j^2}(\tau - \mu_{jj})^2\right) d\tau,$$

$$j = 1, 2, \cdots, S. \tag{2.37}$$

If $S = 1$, the classification is reduced to the detection, and the above integrals have a closed analytical form

$$P_c^*(0) = \Phi\left(\frac{d_0 - \mu_{10}}{\sigma_1}\right) \tag{2.38}$$

and

$$P_c^*(1) = \int_{d_0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(\tau - \mu_{11})^2\right) d\tau = 1 - \Phi\left(\frac{d_0 - \mu_{11}}{\sigma_1}\right). \quad (2.39)$$

Because $\mu_{10} = \sigma^2 ln P_1 - \frac{1}{2}\|s_1\|^2$, $\mu_{11} = \sigma^2 ln P_1 + \frac{1}{2}\|s_1\|^2$ and $d_0 = \sigma^2 ln P_0$, the probabilities of the optimal correct decisions for equal prior probabilities ($P_0 = P_1$) become

$$P_c^*(1) = 1 - \Phi\left(-\frac{\gamma^*}{2}\right) = \Phi\left(\frac{\gamma^*}{2}\right) = P_c^*(0) \quad (2.40)$$

and the total probability of error is

$$P_e^* = 1 - \Phi\left(\frac{\gamma^*}{2}\right) \quad (2.41)$$

where $\gamma^* = \|s_1\|/\sigma$ is signal-to noise ratio.

It is known that under minimum error probability criterion, a matched filter performs the optimal classification if the noise is white Gaussian. Except with specialized hardware, implementing a matched filter is costly because of its computational complexity (multiplication and addition operations are proportional to template points). Faster implementation speeds are possible, however, if performance is compromised by using the following real-time sorting schemes.

2.5 Real-Time Sorting

Three different schemes are proposed for real-time sorting. The theoretical comparisons will be made at the end of the section.

2.5.1 Multithreshold Scheme

The multithreshold method is a fast real-time on-line spike detection scheme. Unlike the single-threshold detection that is commonly used, the multithreshold technique provides more reliable detection, especially when the signal-to-noise ratio is low. To see the effectiveness of the scheme, a comparison is made with the single-threshold scheme by an example later in this section.

The method can be described as a hypothesis test problem. The underlying assumption is that the observed data trace is an additive combination of spikes and noise. Suppose that the spike is deterministic neural signal and has shape $\{s(t) : t = 1, 2, \cdots, M\}$. Denote the observations and the noise as $x(t)$, and $n(t)$, respectively. The hypotheses are stated as follows.

$$\mathcal{H}_0 : \quad x(t) = n(t), \qquad t = 1, 2, \cdots, M$$

$$\mathcal{H}_1 : \quad x(t) = s(t) + n(t), \quad t = 1, 2, \cdots, M$$

The null hypothesis, $\mathcal{H}_0$, is that there is no spike in the observation, and the alternative hypothesis, $\mathcal{H}_1$, means that there is a spike in the observation. The decision as to whether there is a spike is made by M comparisons with the following rule. If $x(t) > \eta_t$, for *all* $t = 1, 2, \cdots, M$, then we accept $\mathcal{H}_1$, otherwise reject $\mathcal{H}_1$, where η_t 's are M independent thresholds. Our

aim is to optimize the thresholds according to a set of performance objectives. There are two types of error in a statistical hypothesis test. The false alarm is the first type of incorrect decision, rejecting the hypothesis $\mathcal{H}_0$ when that hypothesis is true. The second type of error, accepting $\mathcal{H}_0$ when $\mathcal{H}_0$ is false, is called the missing detection. For convenience, it is assumed that the noise is white Gaussian distributed with zero-mean and variance σ^2. This assumption does not lose generality because (i) the noise consists of many weak signals of distant units so that it is Gaussian distributed by the central limit theorem; (ii) the colored noise can be transformed into white by passing it through a whitening filter. In practice, therefore, we can pass $x(t)$ through a whitening filter of which the output $\tilde{x}(t)$ is composed by $\tilde{n}(t)$, the white version of the noise, and $\tilde{s}(t)$, a transformed version of the spike. Under this assumption, the probability of the false alarm P_F can be expressed as

$$P_F \;=\; P_r(D_1|\mathcal{H}_0) \;=\; P_r(x(1) > \eta_1,\; x(2) > \eta_2,\; \cdots,\; x(M) > \eta_M|\mathcal{H}_0). \quad (2.42)$$

Because $n(t)$ is white, and $x(t)$'s are mutually independent and with identical Gaussian distribution $\mathcal{N}(0,\; \sigma^2)$. Therefore, P_F can be further expressed in terms of the error function $\Phi(y) \;=\; \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ as

$$P_F \;=\; \prod_{t=1}^{M} P_r(x(t) > \eta_t|\mathcal{H}_0) \;=\; \prod_{t=1}^{M} \Phi(-\frac{\eta_t}{\sigma}). \quad (2.43)$$

Similarly, the probability of the missing P_M can be written as

$$P_M \;=\; P_r(D_0|\mathcal{H}_1) \;=\; P_r(x(t) < \eta_t,\; \textit{for some } t\; |\mathcal{H}_1). \quad (2.44)$$

For the mathematical manipulation, we write it in terms of the probability of

the complement of the missing

$$P_M = 1 - P_r(D_1|\mathcal{H}_1) = 1 - P_r(x(1) > \eta_1,\ x(2) > \eta_2,\ \cdots,\ x(M) > \eta_M|\mathcal{H}_1). \tag{2.45}$$

Therefore, in terms of the error function, P_M may be written as

$$P_M = 1 - \prod_{t=1}^{M} \Phi(\frac{s(t) - \eta_t}{\sigma}). \tag{2.46}$$

It is not difficult to see that if one wishes P_F to be small, then one chooses thresholds η_t to be large, thus increasing P_M. Conversely, by choosing η_t to be small, P_M decreases while P_F increases. This implies that minimizing both P_F and P_M is a conflicting objective. One may reach a compromise by constructing an objective function J,

$$J = \theta P_F + P_M \ , \quad for\ some\ \theta > 0. \tag{2.47}$$

The goal is now to minimize J.

The necessary condition for achieving a minimum of J is to set

$$\frac{\partial J}{\partial \eta_k} = 0 \ , \quad k = 1,\ 2,\ \cdots,\ M \tag{2.48}$$

which results in M simultaneous equations

$$\theta \prod_{l \neq k}^{M} \Phi(-\frac{\eta}{\sigma_l}) = e^{(2\eta_k - s(k))s(k)/2\sigma^2} \prod_{l \neq k}^{M} \Phi(\frac{s(l) - \eta_l}{\sigma}) \ , \quad k = 1,\ 2,\ \cdots,\ M. \tag{2.49}$$

The possible optimal set of thresholds η_k 's is obtained by solving those equations.

Theorem 2.1. There exists a unique minimum of the objective function J if a spike is detected, and if a false alarm is more costly than a missing detection.

A proof of the theorem is given in the Appendix. The existence and uniqueness of the minimum of J guarantees the risk-free solution of the M simultaneous equations.

The significance of θ may be explained as follows. If the false alarm is as costly as the missing detection, one sets $\theta = 1$; $\theta > 1$ means that the false alarm is more costly than the missing detection, and $\theta < 1$ otherwise. If the *a priori* probability $P(\mathcal{H}_0)$ is known, then the total error probability is $P_e = P(\mathcal{H}_0)P_F + P(\mathcal{H}_1)P_M$ and the objective function becomes $J = P_e/P(\mathcal{H}_1)$ with $\theta = P(\mathcal{H}_0)/P(\mathcal{H}_1)$. Thus minimizing J is equivalent to the minimum error probability criterion. Obviously, the higher the signal to noise ratio $s(t)/\sigma$, the smaller the error probabilities P_F and P_M.

Only partial information about the template of spikes $\{s(t)\}_{t=1}^{M}$ is needed to generate the optimal multithresholds. From (2.42) and (2.46) we know that the threshold η_t's should be between $s(t)$ and σ for each t, which indicates that one only chooses such points in the spike template $\{s(t)\}_{t=1}^{M}$ that $s(t)$ is well above the noise level σ. The multithresholds are optimized by using the values of these well chosen template points and the value of σ. The real-time on-line implementation is simple: Compare the observation datum at every instant t with the corresponding threshold η_t; we announce that there is a spike if all thresholds are exceeded by the data.

2.5.2 Multiwindow Scheme

The multiwindow method is another fast real-time on-line spike classification scheme since no arithmetic operations other than comparisons are involved. Here, for each test, an upper and a lower threshold is chosen to construct a window. As in the previous case, the method can also be described as a multiple hypothesis test problem. The underlying assumption about the noise is the same as in the multithreshold scheme. The decision as to which class a spike belongs is made by M comparisons with the following rule.

There are M amplitude windows for each class. The m-th window for class i is $[l_{im}, u_{im})$. If all selected components in the observed time series data, $\{x(m)\}_{m=1}^{M}$, pass through their windows for class i, we claim that a spike in class i occurs. This decision rule may be expressed as

$$D = \mathcal{H}_i\,, \quad if \ \mathbf{x} \in A_i \tag{2.50}$$

where

$$A_i = \{\mathbf{x} : x(m) \in [l_{im}, u_{im}),\ m = 1, 2, \cdots, M\} \tag{2.51}$$

is another exhaustive and mutually exclusive partition of A.

As before, we can compute the probabilities of correct decision for each class,

$$P_c(j) = \int_{A_j} f_j(\mathbf{x})d\mathbf{x} = \prod_{m=1}^{M} [\Phi(\frac{u_{im} - s_i(m)}{\sigma}) - \Phi(\frac{l_{im} - s_i(m)}{\sigma})]\,,$$

$$j = 1, 2, \cdots, S \tag{2.52}$$

and

$$P_c(0) = \int_{A_0} f_0(\mathbf{x})d\mathbf{x} = 1 - \int_{\cup_{i=1}^{S} A_i} f_0(\mathbf{x})d\mathbf{x} = 1 - \sum_{i=1}^{S}\prod_{m=1}^{M}[\Phi(\frac{u_{im}}{\sigma}) - \Phi(\frac{l_{im}}{\sigma})] \tag{2.53}$$

where $\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$.

Our aim is to optimize the window thresholds according to some criterion. Since the best performer is the matched filter, the probabilities of correct classification for each class, $P_c^*(j), j = 0, 1, \cdots, S$, are the ultimate performance that can be achieved. Therefore, one chooses window thresholds such that the $P_c(j)$'s are arbitrarily close to the corresponding $P_c^*(j)$'s. Suppose that the criterion is chosen to be the weighted squared error, with weight P_j, the *a priori* probability. Then we want to minimize

$$\sum_{j=0}^{S} P_j[\int_{A_j} f_j(\mathbf{x})d\mathbf{x} - P_c^*(j)]^2 \tag{2.54}$$

with respect to l_{im}'s and u_{im}'s, subject to constraints

$$A_i \neq \phi, \quad i = 1, 2, \cdots, S \tag{2.55}$$

and

$$A_i \bigcap A_j = \phi, \quad for \ i \neq j, \ \ i, \ j = 1, 2, \cdots, S. \tag{2.56}$$

By the Lagrange multiplier method, it is equivalent to minimize the objective function J

$$J = \sum_{j=0}^{S} P_j[\int_{A_j} f_j(\mathbf{x})d\mathbf{x} - P_c^*(j)]^2 + a\sum_{j=0}^{S}\sum_{m=1}^{M}(u_{jm} - l_{jm} - \alpha_{jm}^2)^2$$

$$+ b\sum_{i=1}^{S}\sum_{j=0}^{S}\sum_{m=1}^{M}(l_{im} - u_{jm} - \beta_{ij}^2)^2 I_{ijm} \tag{2.57}$$

with respect to l_{im}, u_{im}, α_{im} and β_{ij}, $i,j = 1,2,\cdots,S$; $m = 1,2,\cdots,M$, where

$$I_{ijm} = \begin{cases} 1\,, & m = m_{ij} \ \ and \ \ s_i(m) \geq s_j(m) \\ 0\,, & otherwise \end{cases} \tag{2.58}$$

where m_{ij} is the index for the component in which the templates between class i and class j differ the most.

The optimization of the multiwindow thresholds has an interesting geometric interpretation. Imagine that the M-dimensional Euclidean space $R^m = A$ consists of $S+1$ exhaustive and mutually exclusive subspaces A_i^*, $i = 0,1,\cdots,S$ with arbitrary shapes. The multiwindow threshold partitions the same space with $S+1$ super-rectangles A_i. Moreover, there is a one-to-one correspondence between A_i^* and A_i. The optimization of the multiwindow thresholds tends to make each super rectangle A_i closely approach its corresponding optimally partitioned subspace A_i^*. It is apparent therefore that the multiwindow threshold sorting scheme would perform as well as the matched filter if the templates $\mathbf{s_i}$'s are such that every optimal subspace A_i^* is rectangular.

2.5.3 Soft-Decision Scheme

Now borrowing from communication technology, we turn to another economical detection and classification scheme called soft-decision sorting. The sums of the selected points of templates, rather than individuals of these points as in previous schemes, are taken as the decision variables that will be examined in order to make a decision. As in previous cases, S classes of spikes are to be sorted. The decision rule is described as

$$D = \mathcal{H}_i, \ \ if \ d(\mathbf{x}) \in [\eta_{i-1}, \eta_i), \ \ i = 0,1,\cdots,S \tag{2.59}$$

where $d(\mathbf{x})$ is the decision variable. The decision is made in favor of the spike class corresponding to the decision variable within the partition of the decision space for that class. Using the same underlying assumptions about the noise, the decision variable is Gaussian distributed with mean ν and variance $M\sigma^2$, leading to $S+1$ hypotheses.

Under $\mathcal{H}_i$:

$$d(\mathbf{x}) \;=\; \sum_{m=1}^{M} w(m)x(m) \;\sim\; \mathcal{N}(\nu_i, M\sigma^2), \;\; i = 0, 1, \cdots, S \tag{2.60}$$

where $w(m) = sign(s(m))$, and $\nu_i = \sum_{m=1}^{M} w(m)s_i(m)$ is defined as the signal intensity for class i. Having these statistics, one can compute the probabilities of a correct decision ($j = i$) and of an error ($j \neq i$)

$$\begin{aligned} P_r(D_j|\mathcal{H}_i) \;&=\; P_r(d(\mathbf{x}) \;\in\; [\eta_{j-1},\, \eta_j)|\mathcal{H}_i) \\ &= \int_{\eta_{j-1}}^{\eta_j} \phi_i(t)dt \;=\; \Phi(\frac{\eta_j - \nu_i}{\sqrt{M}\sigma}) - \Phi(\frac{\eta_{j-1} - \nu_i}{\sqrt{M}\sigma}) \end{aligned} \tag{2.61}$$

with $\eta_{-1} = -\infty$, and $\eta_S = \infty$. Therefore, the total probability of error can be written in terms of $P_r(D_j|\mathcal{H}_i)$ as

$$\begin{aligned} P_e \;&=\; \sum_{i=0}^{S}\sum_{j\neq i}^{S} P_j P_r(D_i|\mathcal{H}_j) \;=\; 1 - \sum_{i=0}^{S} P_i P_r(D_i|\mathcal{H}_i) \\ &=\; 1 - \sum_{i=0}^{S} P_i[\Phi(\frac{\eta_i - \nu_i}{\sqrt{M}\sigma}) - \Phi(\frac{\eta_{i-1} - \nu_i}{\sqrt{M}\sigma})]. \end{aligned} \tag{2.62}$$

To optimize the scheme, the necessary condition for achieving a minimum of P_e is to set

$$\frac{\partial P_e}{\partial \eta_k} \;=\; 0 \;, \quad k \;=\; 0,\, 1,\, \cdots,\, S-1. \tag{2.63}$$

This results in S simultaneous equations

$$P_k\phi(\frac{\eta_k-\nu_k}{\sqrt{M}\sigma}) \;=\; P_{k+1}\phi(\frac{\nu_{k+1}-\eta_k}{\sqrt{M}\sigma}) \quad k \;=\; 0,\, 1,\, \cdots,\, S-1. \tag{2.64}$$

Notice that $\phi(t) = \frac{d}{dt}\Phi(t)$ and the even property of $\phi(t)$ has been used. For equal prior probabilities, $P_k = P_{k+1}$, then we have

$$\eta_k \;=\; \frac{\nu_k+\nu_{k+1}}{2}. \tag{2.65}$$

Without loss of generality, one can rank the classes so that

$$\nu_1 \le \nu_2 \le \nu_3 \cdots \le \nu_S,$$

hence

$$\eta_0 \le \nu_1 \le \eta_1 \le \nu_2 \le \eta_2 \cdots \le \nu_{S-1} \le \eta_{S-1} \le \nu_S.$$

In the equal prior probability case, equation (2.65) is valid for $k \;=\; 0,\, 1,\, \cdots,\, S-1$, thus the minimal error probability is calculated to be

$$\min P_e \;=\; 1+\frac{S-1}{S+1}-\frac{2}{S+1}\sum_{k=0}^{S-1}\Phi(\frac{\nu_{k+1}-\nu_k}{2\sigma\sqrt{M}}) \tag{2.66}$$

We can see in the extreme cases that

$$\min P_e \longrightarrow 0 \quad as\ \sigma \to 0 \tag{2.67}$$

and

$$\min P_e \longrightarrow \frac{S}{S+1} \quad as\ \sigma \to \infty. \tag{2.68}$$

If there is only one class, $S \;=\; 1$, it becomes a detection problem, $\min P_e$ is reduced to

$$\min P_e \;=\; 1-\Phi(\frac{\nu_1}{2\sigma\sqrt{M}}) \;=\; 1-\Phi(\frac{\gamma}{2}) \tag{2.69}$$

where $\gamma \doteq \frac{1}{\sqrt{M}} \sum_{m=1}^{M} s_1(m)/\sigma$ is the signal-to-noise ratio. We can also show that the error probability of the matched filter is

$$P_e^* = 1 - \Phi(\frac{\gamma^*}{2}) \tag{2.70}$$

where $\gamma^* = \sqrt{\sum_{m=1}^{M} s_1^2(m)}/\sigma$ is the signal-to-noise ratio for the matched filter in the same situation. It is important to note that

$$\gamma \leq \gamma^* \tag{2.71}$$

and equality holds iff the signal is time invariant, i.e., $s_1(1) = s_1(2) = \cdots = s_1(M)$, which implies that the matched filter is always superior to the soft-decision except for time invariant signals, in which case the soft-decision becomes a matched filter too.

We can view the soft-decision scheme as a special case of the Haar transformation. This is because $d(\mathbf{x})$ has the form which is equivalent to the second component in the transformed domain of an M-point Haar transformation if the entire templates are used to form the decision variable. In other words, with knowledge of spike templates of S classes, the Haar transform can be used in the classification — a soft-decision classifier.

2.5.4 Comparisons Among the Schemes

In order to evaluate the performance of these different schemes, the following comparisons were made. Consider an artificial spike with template $s_c(\tau) = 8\sin(2\pi k\tau)$ contaminated by white Gaussian noise with zero-mean and variance $\sigma^2 = 3.5^2$. The data are sampled at frequency 12 kHz. Let $M = 3$, i.e., only

three points are taken from the template, $s(1) = 8.0000$, $s(2) = -8.0000$, and $s(3) = 6.9282$.

For the detection case, the comparisons of false detection P_F, missing detection P_M, and the computational complexity are given in Table 2.1. For example, with $\theta = 1$, the optimal threshold values of the multithreshold scheme are calculated to be $\eta_1 = 1.470$, $\eta_2 = -0.215$, and $\eta_3 = 1.450$. This results in the error probabilities $P_F \leq 0.0599$ and $P_M \leq 0.0802$. For single-threshold detection with $s(1) = 8.0$, $\sigma = 3.5$ and the optimal threshold $\eta_1 = s(1)/2 = 4.0$, the error probabilities are $P_F = P_M = 0.1265$. It can also be seen from Table 2.1 that the soft-decision scheme performs almost as optimally as the matched filter while saving M multiplication operations. This increases implementation speed considerably.

To classify one spike, the matched filter requires MS multiplications, $(M-1)S$ additions and S comparisons, and the soft-decision operates $(M-1)S$ additions and S comparisons while the multiwindow sorting needs only $2MS$ comparisons.

Table 2.2 illustrates the classification comparisons, calculated using data from an epoch of extracellular recording from monkey motor cortex. The noise level was estimated to be $\sigma = 2.8604$, and three classes were extracted with three most significant points (in terms of signal-to-noise ratio) in each template: $\mathbf{s_1}$ = (24.1917, -14.7333, 19.8833) for class 1, $\mathbf{s_2}$ = (12.6188, -8.8000, 12.0250) for class 2, $\mathbf{s_3}$ = (6.6200, -4.2800, 6.3600) for class 3. Also estimated was a prior probability for every class to be $P_1 = 0.1$, $P_2 = 0.1$, $P_3 = 0.2$, and that for noise class to be $P_0 = 0.6$.

The matched filter is the best scheme for classification in the additive white Gaussian case. Its computational complexity, however, leads to slower processing speed. The alternative classification schemes are inferior in performance but superior in speed to the matched filter. In particular cases such as shown in Table 2.2, however, they can perform as well as the matched filter.

2.6 Testing Examples

Several epochs of extracellular recording from guinea pig and ferret auditory cortexes were used to test the system. The neural signals were recorded via distributed microelectrode sensor arrays developed in collaboration with the Microelectronics Facility of the Naval Research Laboratories (NRL). The data were stored on tape and later sampled at 10 kHz. This guarantees that a spike was represented by sufficient samples, necessary for choosing M points in the templates properly to get maximal signal-to-noise ratio.

The first 60,000 samples (corresponding to 6 s of data) were used to generate templates for each class. The whole algorithm was written in a high level programming language and run on a Masscomp 5500 computer. The computer required about 1 – 5 min (depending on the length of learning data) to complete the extraction of templates. It took few seconds to determine the optimal thresholds for the soft-decision scheme; or about 0.5 min to finish the calculation of the optimal multithresholds for multithreshold scheme. The mean value and the variance of noise were then estimated. After that, the real-time sorting

subsystem was invoked to detect and classify spikes simultaneously.

A display routine was implemented as part of the real-time sorting subsystem in which different classes of spikes are separated and highlighted in different colors on the neural activity trace as shown in Figs. 2.4 – 2.6. Figure 2.4 shows an epoch of extracellular recording from guinea pig auditory cortex using one channel of a silicon-based 40-channel microelectrode. Fig. 2.4(a) shows 200 ms of original data. Fig. 2.4(b) shows the original data contaminated by band-limited noise with a bandwidth of 2.5 kHz. Clearly, it is impossible to detect both spikes in this trace by a single-threshold device. Fig. 2.4(c) demonstrates the detection of the spikes in noisy trace 2.4(b) by the multithreshold scheme. In Fig. 2.4(d), a much noisier trace is formed by adding 60 Hz and 90 Hz sinusoids to the trace in Fig. 2.4(b) and both spikes are now totally buried in the waves. We see in Fig. 2.4(e) that both spikes in 2.4(d) can still be detected by the powerful soft-decision scheme. Figure 2.5 shows spikes from monkey motor cortex extracellular recordings, kindly provided by Dr. E. M. Schmidt. Classes were determined in the learning subsystem and separated in the real-time sorting subsystem by the multithreshold scheme, we are only interested in the first three most significant classes for this data. Figure 2.6 shows 1 second of data extracellularly recorded from ferret auditory cortex with corresponding sorted four classes performed by the soft-decision scheme, again, the first four most significant classes are interesting in this case.

A comparison of the performance of the automatic system with that of experienced human observers indicates that most spikes are detected and that the discrepancies between classifications done by human observers and the recognition system are small. In particular, for the first four most significant classes of

the ferret data the rate of missing detection was 2 % and that of false detection was 1 % and the discrepancies between classifications were 4 %.

2.7 Summary

A software-based neural spike sorting system has been developed to implement detection and classification of signals from multiple neurons without human supervision. After the spike templates for each class are estimated from original data and the optimal multithresholds are selected based on the templates and the noise level, the real-time subsystem starts spike sorting. Testing examples show the potential of the system. Further improvements under consideration include expanding the dimension of the feature space in the template learning subsystem and developing classifiers using recursive algorithm for Haar transform detection and classification.

Schemes	P_F	P_M	Multiplications	Additions	Comparisons
Single-threshold	.1265	.1265	0	0	1
Multithreshold †	.0599	.0802	0	0	M
Soft-decision	.0599	.0140	0	$M-1$	1
Matched filter	.0599	.0139	M	$M-1$	1

Table 2.1: Performance Comparisons for Detection with Noise Level $\sigma = 3.5$ and M (=3) Points from the Template $\mathbf{s}$ = (8.0000, -8.0000, 6.9282). † For $\theta = 1$.

Schemes	$P_c(0)$	$P_c(1)$	$P_c(2)$	$P_c(3)$	$\min P_e$	Speed
Multiwindow	.9318	.9527	.8318	.7778	.1069	$2MS$ comparisons
Soft-decision	.9802	.9948	.9171	.8907	.0426	$(M-1)S$ additions and S comparisons
Matched filter	.9802	.9967	.9175	.8909	.0423	MS multiplications, $(M-1)S$ additions and S comparisons

Table 2.2: Performance Comparisons for Classification with Noise Level $\sigma = 2.8604$. S (=3) units are separated with each having M (=3) points from the templates $\mathbf{s_1}$ = (24.1917, -14.7333, 19.8833), $\mathbf{s_2}$ = (12.6188, -8.8000, 12.0250), $\mathbf{s_3}$ = (6.6200, -4.2800, 6.3600) with prior probabilities $P_0 = 0.6$, $P_1 = 0.1$, $P_2 = 0.1$, $P_3 = 0.2$.

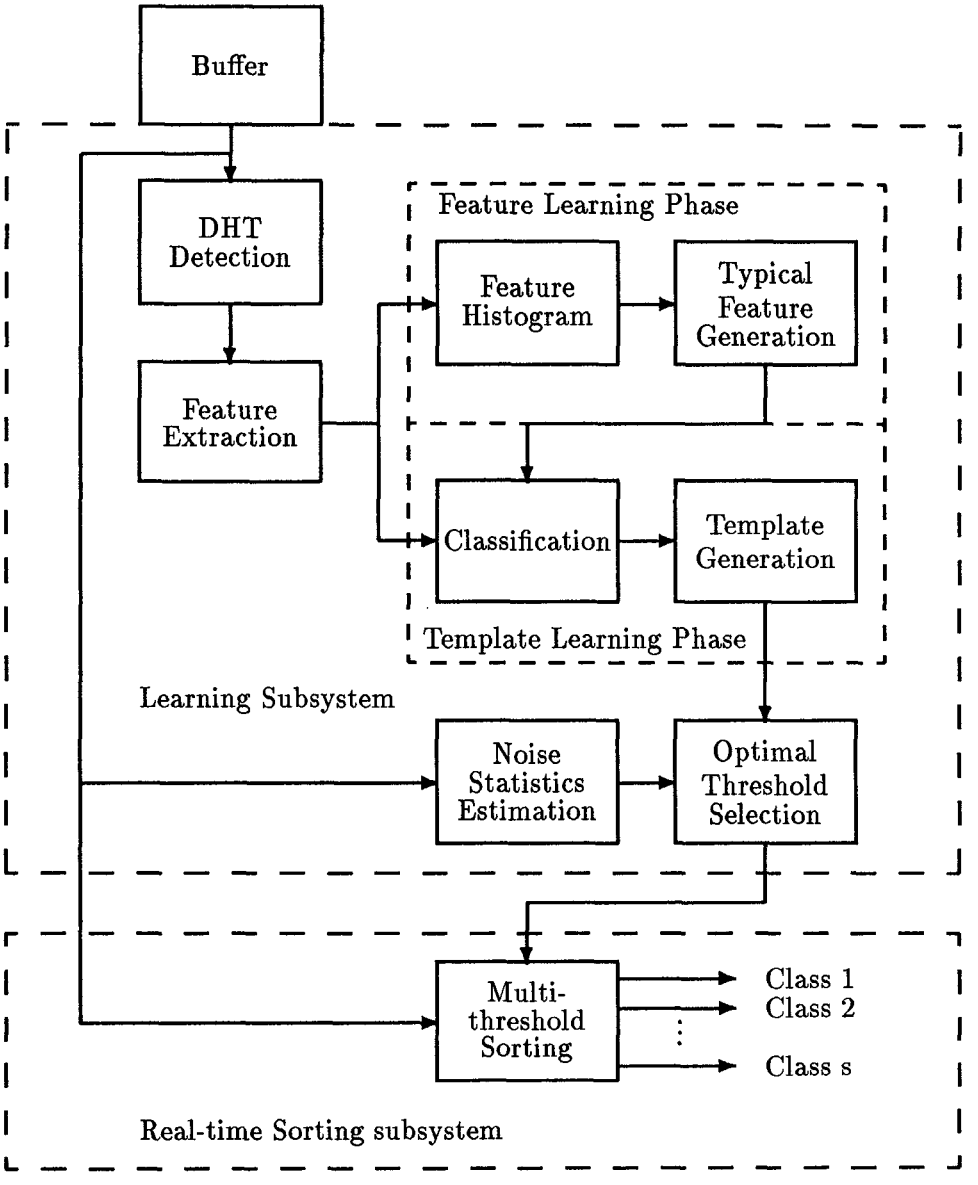


Figure 2.1: Schematic Diagram of the System.

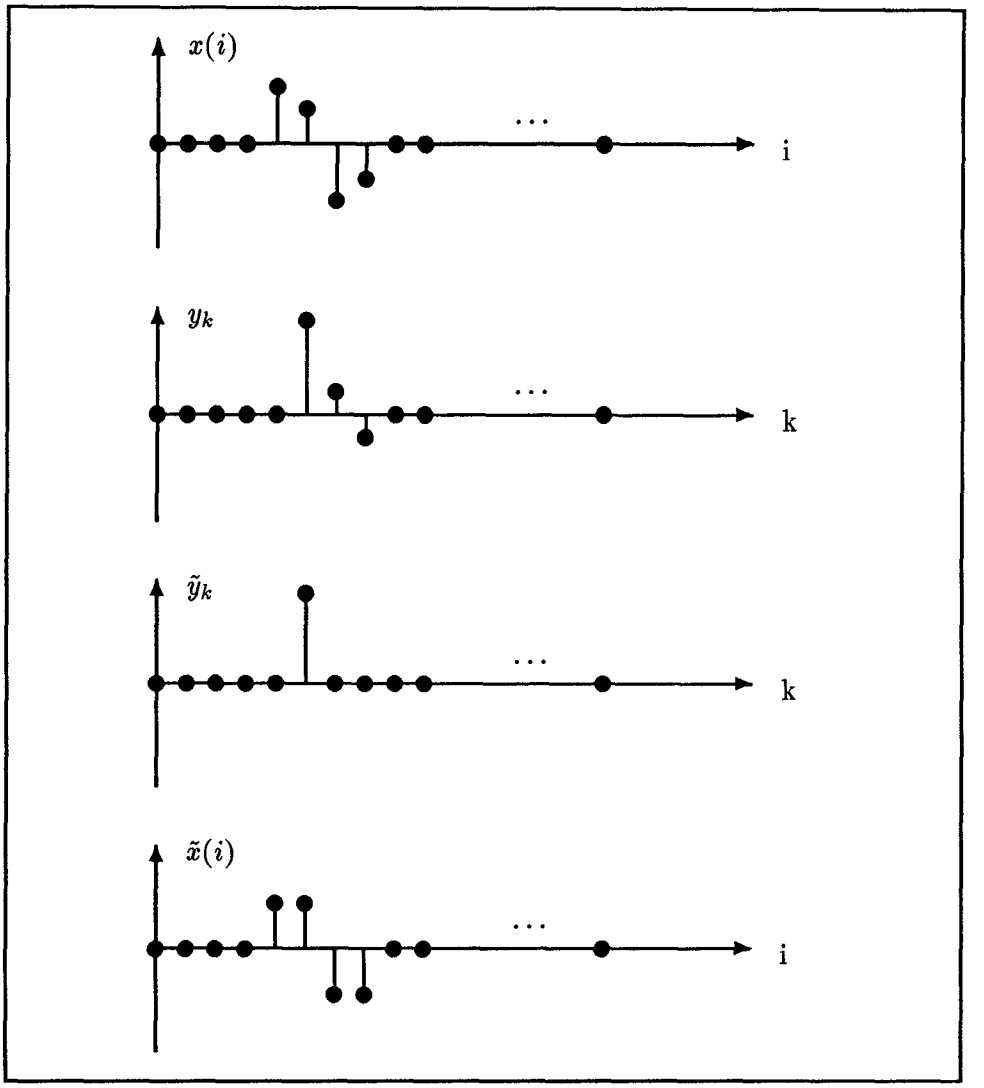


Figure 2.2: Illustration of Haar transform detection by using an artificial spike, where $x(i) = a\sin(2\pi(i-3)/5)$, $i = 4,5,6,7$; $x(i) = 0$, otherwise.

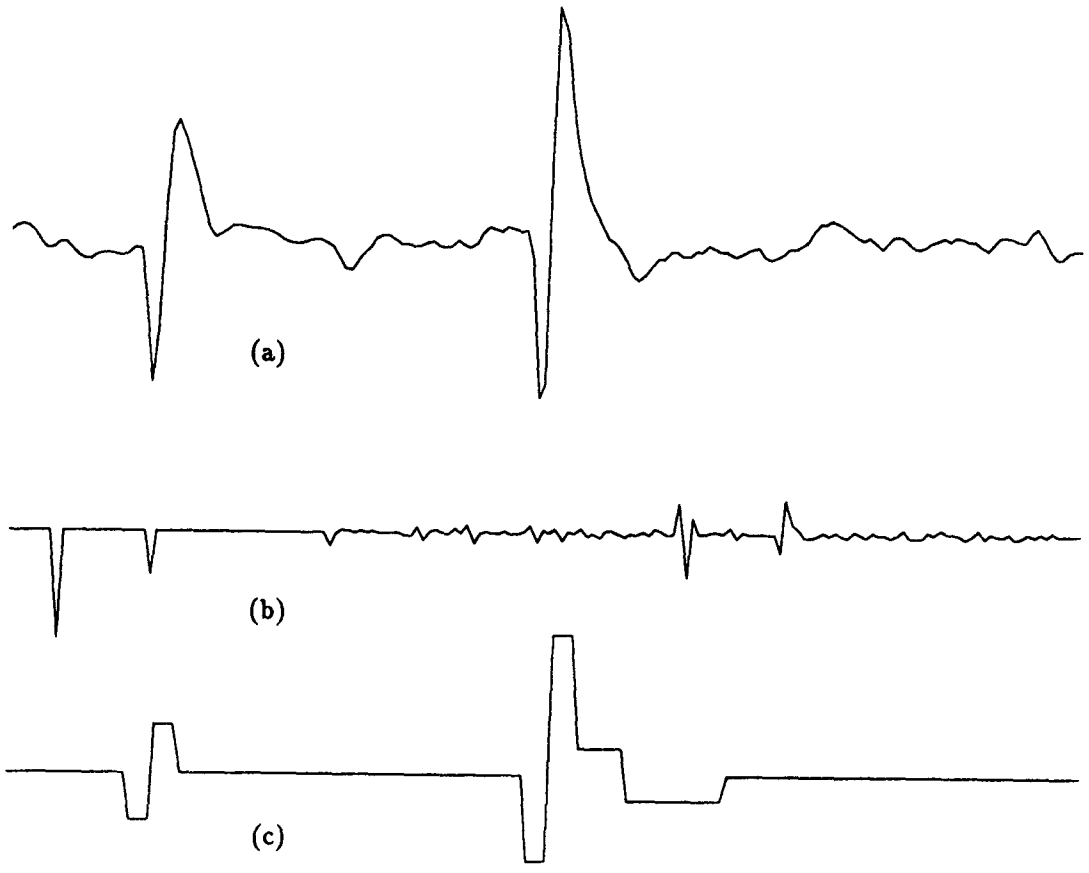


Figure 2.3: Spikes detected by the Haar transform:

(a) The original trace.

(b) The transform domain equivalence; large components indicate presence of spikes.

(c) Reconstructed spikes in the time domain with noise filtered out. Spike times are determined by zero-crossing detection.

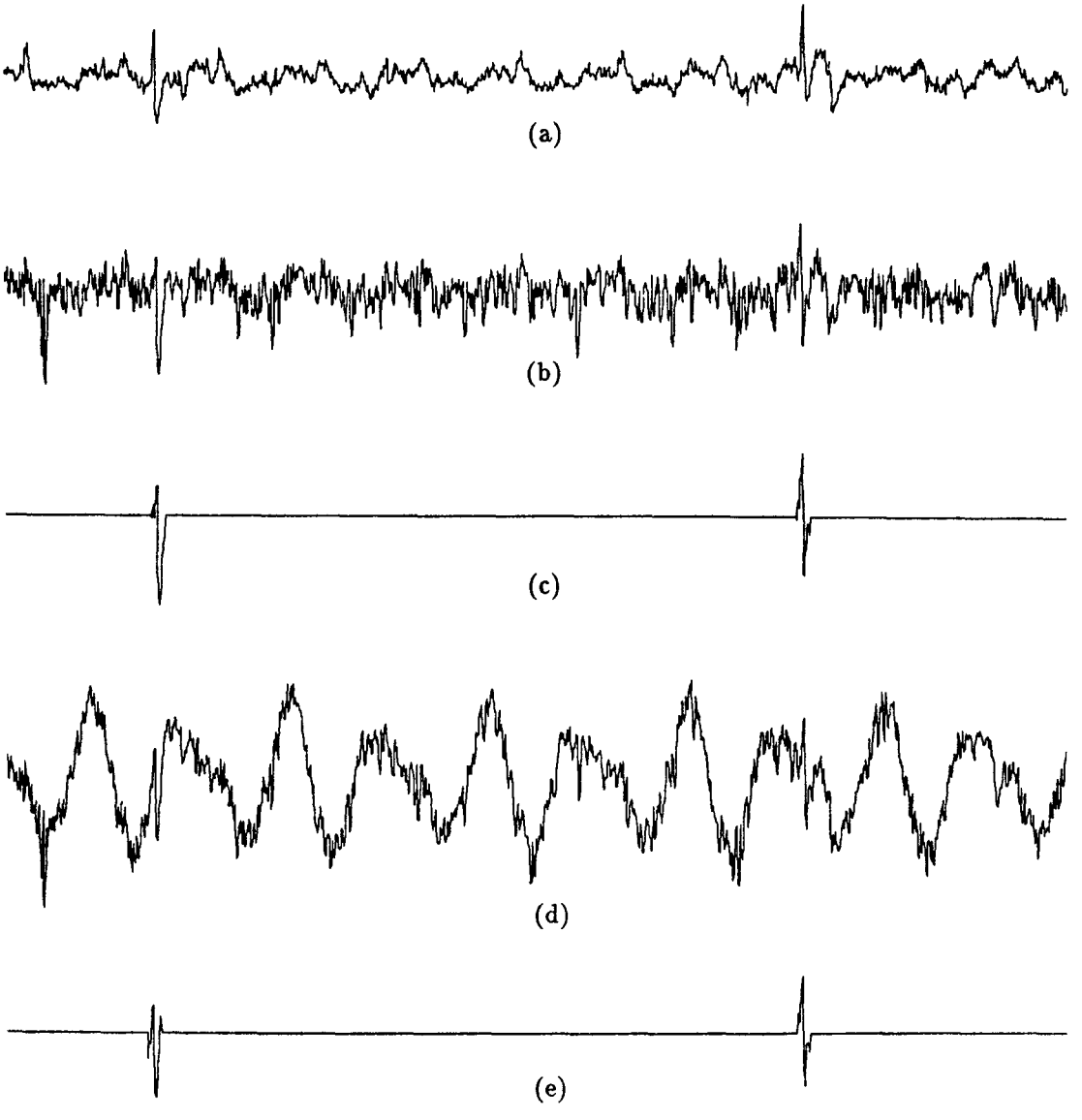


Figure 2.4: Neural spikes recorded by a silicon-based multielectrode from guinea pig auditory cortex:

(a) 200 ms of original data.

(b) The original data contaminated by band-limited noise with a bandwidth of 2.5 kHz.

(c) Spikes in noisy trace (b) detected by the multithreshold scheme.

(d) Much noisier trace formed by adding 60 Hz and 90 Hz sinusoids to the trace in (b); both spikes are totally buried in the waves.

(e) Both spikes in d are detected by the soft-decision scheme.

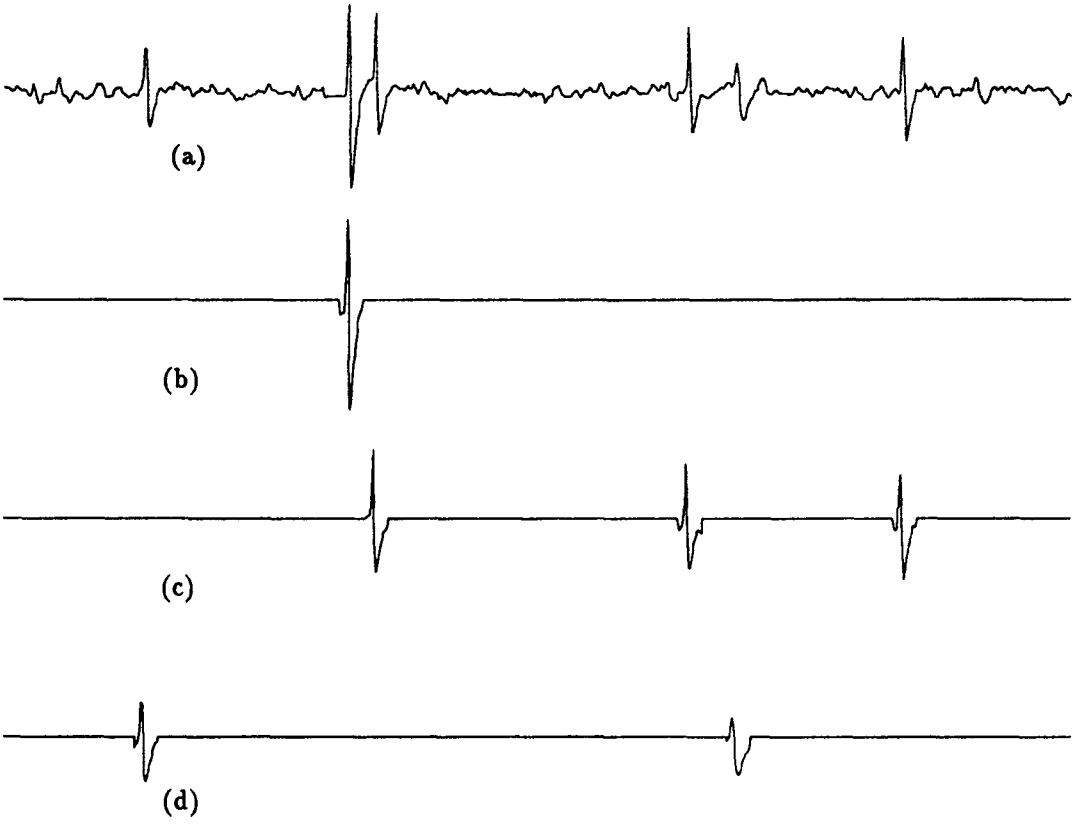


Figure 2.5: Classes of neural spikes from monkey motor cortex identified by multi-threshold sorting; Only the first three most significant units are plotted:

(a) 100 ms of original data.

(b) Unit 1.

(c) Unit 2.

(d) Unit 3.

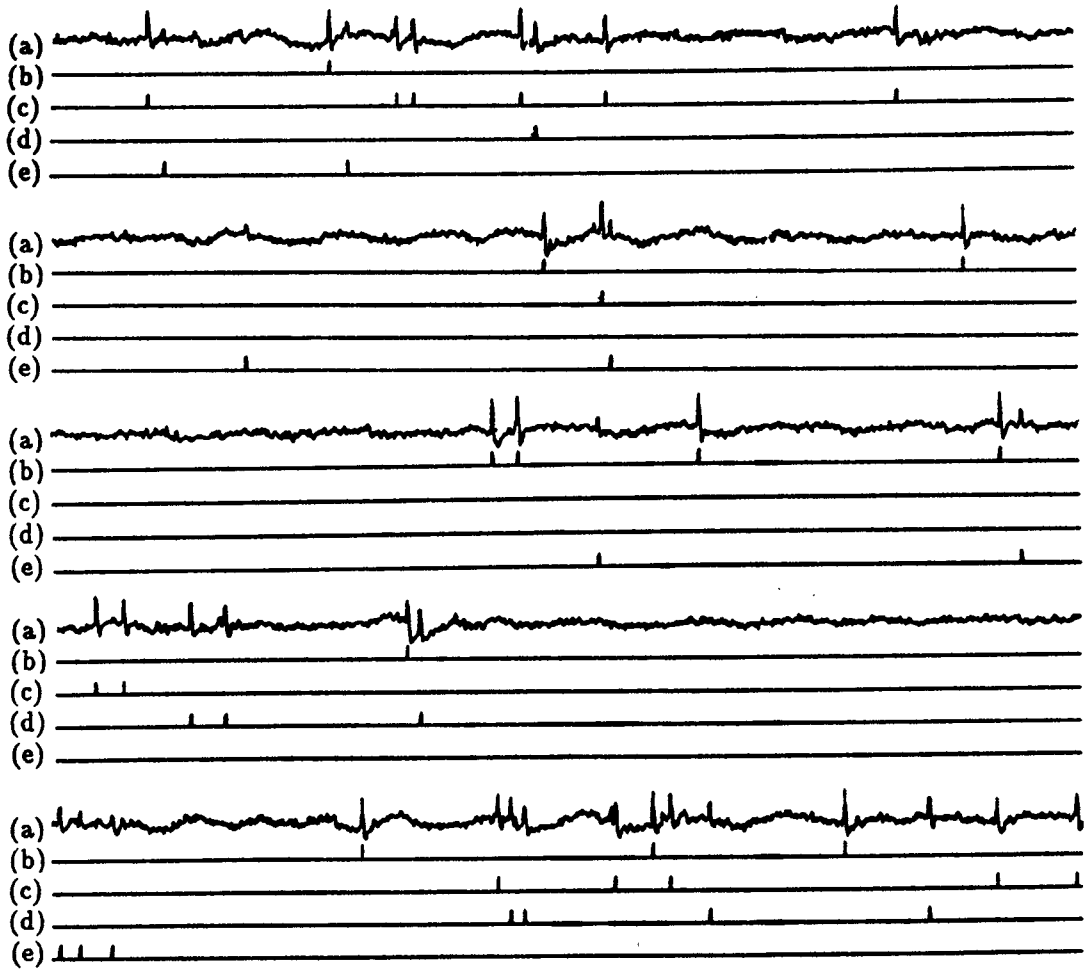


Figure 2.6: One second of data containing neural spikes from ferret auditory cortex separated by soft-decision sorting, chopped into five frames with 200 ms of data each; Only the first four most significant classes are plotted:

(a) 1 second of original data.

(b) Unit 1.

(c) Unit 2.

(d) Unit 3.

(e) Unit 4.

CHAPTER
THREE

IDENTIFICATION OF CONNECTIVITY IN NEURAL NETWORKS

3.1 Introduction

Most functions of the mammalian nervous system are performed by networks of highly interconnected neurons. In the experimental study of these networks, extracellular recordings are often employed to sample the patterns of action potentials simultaneously generated by several neurons [1], [13], and [42]. The correlations among the recorded firings of the different cells are then used as measures of the type and strength of their interconnections. Many such measures have been proposed to accomplish the latter task; they include the cross-interval histograms, the cross-correlation histograms, the cross-covariance histogram, and the joint peri stimulus time (PST) histogram (the scatter diagram) [11], [13]. In all cases, the histograms provide statistical measures in support of various hypotheses such as whether the two (or more) neurons under study directly influence each other or simply share common inputs, and whether the influences are excitatory or inhibitory.

There are three basic difficulties with these methods that we tackle in this chapter. The first concerns the lack of flexible general analytical treatments that outline the relations between the synaptic connectivities and the correlation measures that are used to estimate them. Thus, while various features in the above mentioned histograms may reflect qualitatively the underlying connections, several parameters and conditions can render these measures inadequate. Examples of such difficulties are the differing integrating dynamics of different cell types, and the potentially severe errors due to stimulus-induced (rather than synaptic) correlations. Attempts to overcome these problems, as in the use of the *shuffling method* to reduce stimulus effects, are shown here to be largely inadequate.

The second basic shortcoming of the above correlation methods stems from the nonstationarity of the neural records. In constructing cross-interval and cross-correlation histograms, counts are usually obtained not only by averaging over different stimulus presentation but also by averaging over the time duration of each presentation period. This makes these two estimates inadequate when working with nonstationary records and, instead, measures based on time-dependent histograms such as the joint PST scatter diagram should be used for the analysis [17], [40].

Finally, it is unclear in many existing methods how to extend the analysis to more than two neurons, and how to evaluate the degree to which a pairwise estimate is improved when the records from many other neurons are included. This is a particularly important criterion as progress in multiunit recording technologies which promises to increase significantly the number of records of simultaneously active neurons.

To summarize, the objectives of this chapter are (1) to provide rigorous analytical and experimental methods to estimate synaptic connectivities from simultaneous recordings of multiple neurons that are based on accurate and flexible neuron models; (2) to express synaptic connectivity in terms of probability densities of joint neuronal firings and individual neuronal firings that can be used with nonstationary (or stationary) records; (3) to extend these methods from pairwise to multiunit correlations.

The chapter is organized as follows. In the next section (3.2), a stochastic nonlinear neuron model is proposed, and the spike train generated by the model is expressed by a doublely stochastic process. This model will serve as the fundamental tool upon which the analytical results are based. In section 3.3, quantitative analyses of neuronal connectivities are carried out through the model. These include derivations of the relations between the synaptic connectivity and the firing probability densities, and extending the pairwise correlations to the multineuron case. In section 3.4, the results are summarized and discussed in the context of practical implementations and considerations of the accuracy of the estimates. Finally, the analytical results are simulated and discussed in section 3.5. The proofs of lemmas and theorems are given in the Appendix.

All the analytical treatments are contained within sections 3.2 and 3.3. For the reader interested only in using the final expressions, section 3.4 outlines the results and is sufficient as a guide for their experimental applications.

3.2 The Neuron Model

The basic unit of the nervous system which receives and transmits neural signals is the neuron. The interactions of neurons in a network occur in most cases through synaptic connections between them. Most synapses are found between the axon terminals of a presynaptic neuron and the soma or dentritic tree of a postsynaptic neuron. Since there can be many synapses between any two neurons, it is impractical in modeling the neural network to account for individual synapses; rather, it is more fruitful both for experimental investigation and mathematical description to consider the *total effective influence* of one cell on another.

A synaptic connection from a presynaptic neuron B to postsynaptic neuron A is said to be excitatory (inhibitory) if the firing rate of neuron A increases (decreases) when neuron B fires. For the purposes of the model, we assume that the postsynaptic potentials due to many presynaptic inputs are continuously integrated to produce a change in postsynaptic *membrane potential.* A neuron fires an action potential when its membrane potential exceeds a *threshold* level. After each action potential, there is a period during which the probability of firing is reduced. This period is divided into two intervals: The first is the *absolute refractory period*, in which the neuron cannot fire again; the second is the *relative refractory period* where the neuron may generate a spike only when the stimulus is fairly strong.

Since the action potentials of a given neuron are similar in shape, we assume that the transmitted information is carried only through the temporal patterns of the spike trains, and hence we use a sequence of impulses to abstract a train of action potentials. Because the instantaneous firings of a neuron are not

deterministic, a stochastic point process is adopted to model the firings [31], [32].

All the stochastic processes and random variables to be discussed are defined on some probability space $(\Omega, \mathcal{F}, P)$. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\{\mathcal{F}_t : t \geq 0\}$ be a non-decreasing family of sub σ-fields of $\mathcal{F}$ (i.e., $\mathcal{F}_s \subset \mathcal{F}_t$, for every $s < t$). The family $\{\mathcal{F}_t\}$ is called a *history*, and $\mathcal{F}_t$ represents the information collected during $[0, t]$. Let $\{V_t : t \geq 0\}$ be a stochastic process (representing the semi-membrane potential process) defined on $(\Omega, \mathcal{F}, P)$. The family of the sub σ-fields generated by $\{V_t : t \geq 0\}$, $\mathcal{H}_t = \sigma\{V_s : 0 \leq s \leq t\}$, is called the history of $\{V_t : t \geq 0\}$ if $\mathcal{H}_t \subset \mathcal{F}_t$, for $t \geq 0$. And in this case, $\{V_t : t \geq 0\}$ is said to be *adapted* to $\{\mathcal{F}_t\}$. Let $\mathcal{R}^+$ be the σ-field generated by set $[0, \infty)$. A function f from $(\Omega, \mathcal{H}_t)$ into $([0, \infty), \mathcal{R}^+)$ is measurable if for every $S \in \mathcal{R}^+$, $f^{-1}(S) \in \mathcal{H}_t$.

Consider that neuron A is influenced by a family of neurons B_i, $i = 1, 2, \cdots, n$. The model we use is depicted in Fig. 3.1; it is similar in many respects to that studied by Knox [20] and by Van Den Boogaard et al. [7]. A sequence of impulses from neuron B_i is transformed into a membrane potential in neuron A. If the integrated membrane potential exceeds a threshold value $\theta(t)$, an impulse (spike) is generated while the membrane potential discharges to a resting level v_o. $h_i(t, s)$ is the impulse response (not necessarily time-invariant) which describes the total temporal influence of neuron B_i on neuron A from past up to present, including the conduction and transmission delay. A synaptic connection is said to be excitatory if $h(t, s) \geq 0$ for all t, all s in the real line $\mathbf{R}$; it is said to be inhibitory if $h(t, s) \leq 0$ for all t, all s in $\mathbf{R}$.

The somatic (membrane) potential (W_t^A) of neuron A is represented by a linear spatial-temporal superposition of all input action potentials of neurons $B_1, B_2, \cdots, B_n$ (including self-inhibition and/or self-excitation), and an unknown random potential U_t which represents the influence of all other unobservable neurons and biophysical factors. A sigmoid function g is used to map the somatic potential as follows:

$$W_t^A = g(U_t + \sum_{i=1}^{n} \int_0^t h_i(t,\tau) dN_{B_i}(\tau)) = g(U_t + \sum_{i=1}^{n} \sum_{k=1}^{N_{B_i}(t)} h_i(t, T_k^{B_i})) \qquad (3.1)$$

where $\{T_k^{B_i} : k = 1, 2, \cdots\}$ are the epoch times of spike train from neuron B_i, and $\{N_{B_i}(t) : t \geq 0\}$ is the associated counting process, i.e., the number of spikes arriving from neuron B_i in the interval $(0, t]$.

For mathematical simplicity, let us assume that the nonlinearity g has the form of $g(x) = \alpha e^x$, $\alpha > 0$, i.e., that neuron A is operating around threshold and is thus not strongly driven. Suppose further, without loss of generality, that we are interested in finding the connectivity between two neurons A and B_1. In the following discussion, we write $B = B_1$ and $h(t,s) = h_1(t,s)$ for simplicity. Then we write

$$W_t^A = \frac{1}{\alpha} g(U_t + \sum_{i=2}^{n} \sum_{k=1}^{N_{B_i}(t)} h_i(t, T_k^{B_i}))\ V_t^A \qquad (3.2)$$

where V_t^A is called here the *semi-membrane potential* due to neuron B and is defined as

$$V_t^A = g(\sum_{k=1}^{N_B(t)} h(t, T_k^B)). \qquad (3.3)$$

In order to account for the firings of neuron A that are due to V_t^A, we can think

of the factor $\frac{1}{\alpha}g(U_t + \sum_{i=2}^{n}\sum_{k=1}^{N_{B_i}(t)} h_i(t, T_k^{B_i}))$ as a continuous random variable Z, such that a random threshold $\theta(t)$ is formed, which is defined as

$$\theta(t) = Z\theta_o(t) \tag{3.4}$$

where $\theta_o(t)$ is a time function. Due to the refractory period r during which a neuron is unable to produce a successive spike, the time function can be taken as simple as

$$\theta_o(t) = \begin{cases} \infty, & T_k \leq t < T_k + r \\ \theta_o, & T_k + r \leq t < T_{k+1} \end{cases} \tag{3.5}$$

where $\theta_o > 0$ is a constant, and T_k and T_{k+1} are the times at which the k-th and $(k+1)$-st spike occur, respectively.

A spike occurs whenever the threshold is exceeded by the accumulated semi-membrane potential, i.e.,

$$\int_{t_0}^{t} V_\tau^A d\tau > \theta(t) \tag{3.6}$$

where t_0 is the instant of the preceding spike. Because the firing rate of a neuron is finite, the threshold is bounded below by the resting level $v_o > 0$. Denoting by $N_A(t)$ the number of spikes in train A during time interval $(0, t]$, a stochastic counting process $\{N_A(t) : t \geq 0\}$ is associated with spike train A with $N_A(0) = 0$. Let $\Delta N_A(t) = N_A(t + \Delta t) - N_A(t)$ be the number of spikes in an infinitesimal duration Δt. We say that a process is *orderly* if $P_r(\Delta N_A(T) > 1) = o(\Delta t)$.

Figure 3.2 shows a common input model structure where neurons A and B are simultaneously influenced by a source D which may represent a stimulus source or a neuron. The source generates stimuli or spikes with rate V_t^D.

Armed with these general neuron models, we are ready for the analysis of the interneuronal connectivities deduced from the stochastic firing of several neurons.

3.3 Analytical Results

In this section, we shall derive and elaborate on four basic results. We shall first consider the simple case of two observable neurons, and show how the connectivity between them can be expressed analytically in terms of the neuron model outlined above. We then consider the sources of uncertainty in this estimate and how they can be reduced through added information from neighboring neurons. Thirdly, we shall comment on the critical normalization procedures used to remove the confounding effects of stimulus artifacts. Finally, we show that the correlation peak due to a common input spreads more than that due to a synaptic connection.

In the following discussion, we will make use of the PST histogram of a single cell spike train which measures the firing rate of a neuron with respect to the stimulus onset. Each bin of the PST histogram is an unbiased estimator for the probability density of the average neuron firing over a short period Δt at instant t corresponding to that bin. Let $\Delta N_A(t) = N_A(t + \Delta t) - N_A(t)$ be the number of spikes in train A in time interval Δt. Let us denote by $P_A(t)$ the conditional firing probability density of the *postsynaptic* neuron given the history of the intensity process of the presynaptic neuron, $\mathcal{H}_t^B = \sigma\{V_s^B : s \leq t\}$,

and the history $\mathcal{N}_t^A = \sigma\{N_A(s) : s \leq t\}$ of spike train A, that is,

$$P_A(t) = \lim_{\Delta t \to 0} \frac{P_r(\Delta N_A(t) = 1 | \mathcal{H}_t^B, \mathcal{N}_t^A)}{\Delta t}. \tag{3.7}$$

The firing probability density of neuron A is defined as

$$P_r(A_t) = E[P_A(t)] = \lim_{\Delta t \to 0} \frac{P_r(\Delta N_A(t) = 1)}{\Delta t} \tag{3.8}$$

where the second equality is obtained by interchanging the limitation and the expectation operations. It is interchangeable because the firing rate of a neuron is finite. This means that there exists a random variable Q with $E|Q| < \infty$ such that

$$\frac{P_r(\Delta N_A(t) = 1 | \mathcal{H}_t^B, \mathcal{N}_t^A)}{\Delta t} \leq Q$$

almost surely for any $\Delta t > 0$. Hence this interchangeability is guaranteed by the dominated convergence theorem. This argument applies to every similar situation throughout Chapter 3.

Likewise, denote by $P_B(s)$ is the conditional firing probability density of the *presynaptic* neuron given the history of the intensity process of the presynaptic neuron and the history of spike train B, that is,

$$P_B(s) = \lim_{\Delta s \to 0} \frac{P_r(\Delta N_B(s) = 1 | \mathcal{H}_s^B, \mathcal{N}_s^B)}{\Delta s}. \tag{3.9}$$

We have

$$P_r(B_s) = E[P_B(s)] = \lim_{\Delta t \to 0} \frac{P_r(\Delta N_B(s) = 1)}{\Delta s}. \tag{3.10}$$

Note that the individual PST histograms of neurons A and B estimate $E[P_A(t)]\Delta t$ and $E[P_B(s)]\Delta s$, respectively, and that $P_B(s)$ is *not* defined symmetrically to $P_A(t)$.

Moreover, the joint PST histogram of the two neurons estimates $E[P_{AB}(t,s)]\Delta t\Delta s$ where $P_{AB}(t,s)$ represents the conditional joint probability density of firing of neurons A and B,

$$P_{AB}(t,s) = \lim_{\Delta t,\Delta s\to 0} \frac{P_r(\Delta N_A(t)=1, \Delta N_B(s)=1|\mathcal{H}^B_{\max(t,s)}, \mathcal{N}^A_t, \mathcal{N}^B_s)}{\Delta t\ \Delta s} \quad (3.11)$$

and

$$P_r(A_t, B_s) = E[P_{AB}(t,s)] = \lim_{\Delta t,\Delta s\to 0} \frac{P_r(\Delta N_A(t)=1, \Delta N_B(s)=1)}{\Delta t\ \Delta s}. \quad (3.12)$$

Recall that $h(t,s)$ represents the synaptic connectivity between neurons A and B. The four basic results derived are as follows.

Result 1. The joint probability density of firing of a presynaptic and postsynaptic neuron pair can be expressed as the product of individual firing probability densities and the pairwise connectivity, and a corrupting (uncertainty) factor due to other unobservable influences on the firing of A:

(i)

$$P_r(A_t, B_s) = P_r(A_t)P_r(B_s)\gamma(t,s)e^{h(t,s)} \quad (3.13)$$

where $\gamma(t,s)$ is the corrupting factor ($\gamma \geq 0$) given by

(ii)

$$\gamma(t,s) = \frac{E[f_A(t,\theta^A_t)P_B(s)]}{E[f_A(t,\theta^A_t)]\ E[P_B(s)]} \quad (3.14)$$

where

(iii)

$$f_A(t,\theta^A_t) = V^A_t\ \frac{f_{\theta^A_t}(a_t)}{1-F_{\theta^A_t}(a_t)} \quad (3.15)$$

where $a_t = \int_{t_0}^{t} V_\tau^A d\tau$, and $f_{\theta_t^A}(\cdot)$, $F_{\theta_t^A}(\cdot)$ are the density and the distribution functions of the threshold of neuron A, respectively.

Result 2. The uncertainty can be reduced (i.e., the corrupting factor can be made closer to 1) if more interacting neurons $C_1, C_2, \cdots, C_m$ are observed simultaneously. If $P_r(A_t, \mathbf{C}_t) \neq 0$ and $P_r(B_s, \mathbf{C}_t) \neq 0$, then the pairwise connectivity becomes

$$h(t,s) = \log \frac{P_r(A_t, B_s, \mathbf{C}_t) P_r(\mathbf{C}_t)}{P_r(A_t, \mathbf{C}_t) P_r(B_s, \mathbf{C}_t)} - \log \gamma^* \tag{3.16}$$

with $\mathbf{C}_t = \{\cap_{i=1}^{m} C_i \text{ fires in } [t, t+\Delta t)\}$, where

$$\gamma^*(t,s) = \frac{E[f_A(t, \theta_t^A) P_B(s) | \mathbf{C}_t]}{E[f_A(t, \theta_t^A) | \mathbf{C}_t] \; E[P_B(s) | \mathbf{C}_t]} \tag{3.17}$$

is a quantity satisfying $|\gamma^* - 1| \leq |\gamma - 1|$. If γ^* is very close to 1, then $\log \gamma^*$ can be negligited.

Result 3. In order to minimize the effects of the stimulus on the estimators of the connectivity, the normalized joint probability of firing given by :

$$N_p(t,s) = P_r(A_t, B_s) / P_r(A_t) P_r(B_s) \tag{3.18}$$

leads to estimators superior to those produced by the often employed *shuffle* method (normalization by difference):

$$N_d(t,s) = P_r(A_t, B_s) - P_r(A_t) P_r(B_s), \tag{3.19}$$

which is the quantity that the cross-covariance histogram estimates.

Result 4. This result is for the common input model illustrated in Fig. 3.2. We assume that the thresholds in neurons A and B are statistically independent.

Denote by $\mathcal{H}_t^D = \sigma\{V_s^D \ : \ s \leq t\}$ the history of the intensity process of the common input source D. Define

$$P_r(A_t|\mathcal{H}^D_{\max(t,s)}) \ = \ \lim_{\Delta t \to 0} \frac{P_r(\Delta N_A(t) = 1|\mathcal{H}^D_{\max(t,s)})}{\Delta t} \tag{3.20}$$

and

$$P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)}) \ = \ \lim_{\Delta t, \Delta s \to 0} \frac{P_r(\Delta N_A(t) = 1, \Delta N_B(s) = 1|\mathcal{H}^D_{\max(t,s)})}{\Delta t \ \Delta s}. \tag{3.21}$$

Assume further that $P_r(A_t|\mathcal{H}^D_{\max(t,s)}) \neq 0$ and $P_r(B_s|\mathcal{H}^D_{\max(t,s)}) \neq 0$, we have the following

(a) If D represents a deterministic stimulus source, then

$$\frac{P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)})}{P_r(A_t|\mathcal{H}_t^D) \ P_r(B_s|\mathcal{H}_s^D)} \ = \ 1. \tag{3.22}$$

(b) If D represents a neuron with firing rate V_t^D, then

$$\frac{P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)})}{P_r(A_t|\mathcal{H}_t^D) \ P_r(B_s|\mathcal{H}_s^D)} = \exp\{\int_0^{\min(t,s)} (e^{h_A(t,\tau)} - 1)(e^{h_B(s,\tau)} - 1) V_\tau^D d\tau\}. \tag{3.23}$$

3.3.1 Further Relationships

In order to discuss the derivation of the above stated results, we will need to utilize a few more relationships. Given a pair of interacting neurons (A and B), the following lemmas will play an important role in the analysis below. Let us first define an auxiliary function

$$f_B(t, \theta_t^B) = \lim_{\Delta t \to 0} \frac{P_r(\Delta N_B(t) = 1|\theta_t^B; \mathcal{H}_t^B, \mathcal{N}_t^B)}{\Delta t}. \tag{3.24}$$

Lemma 3.1. $P_B(t)$ can be expressed as a map from the semi-membrane potential space of neuron B onto $[0, \infty)$,

$$P_B(t) \;=\; V_t^B \; E_{b_t}[f_B(t, \theta_t^B)] \tag{3.25}$$

with

$$f_B(t, \theta_t^B) = V_t^B \; \frac{f_{\theta_t^B}(b_t)}{1 - F_{\theta_t^B}(b_t)} \tag{3.26}$$

where $b_t = \int_{t_0}^t V_\tau^B d\tau$, and $f_{\theta_t^B}(\cdot)$, $F_{\theta_t^B}(\cdot)$ are the density and the distribution functions of the threshold of neuron B, respectively. The expectation $E_{b_t}[\cdot]$ is taken with respect to θ_t^B. The function $P_B(\cdot)$ can have a very simple form. For example, if the threshold is an exponentially distributed independent random variable with mean λ, then $P_B(t) = \lambda V_t^B$. And in this case, $\{N_B(t) : t \geq 0\}$ is a doubly stochastic Poisson process.

Lemma 3.2. The conditional expectation of the product of the semi-membrane potential of neuron A and the firing rate of neuron B can be expressed as

$$E[V_t^A \frac{dN_B(s)}{ds}] \;=\; e^{h(t,s)} E[V_t^A P_B(s)] \tag{3.27}$$

The proofs of the lemmas are given in the Appendix. Lemma 3.1 gives the expression of the conditional firing probability density of the presynaptic neuron given the history of the intensity process of that neuron. Lemma 3.2 relates the connectivity, the membrane potential of the postsynaptic neuron, and the firing rate of the presynaptic neuron.

3.3.2 Discussion of Result 1

We will first need to derive an expression for the firing rate of the post-synaptic neuron (A). In general, the threshold θ_t^A of this neuron is not an independent variable, since it depends on all other unobservable inputs to the neuron. Given an arbitrary value for θ_t^A, we can write

$$P_r(\Delta N_A(t) = 1 | \theta_t^A = c) =$$

$$P_r(\int_t^{t+\Delta t} V_\tau^A d\tau \geq \theta_t^A | \theta_t^A = c) \simeq P_r(V_t^A \geq \frac{\theta_t^A}{\Delta t} | \theta_t^A = c). \tag{3.28}$$

Since θ_t^A is a positive threshold, by the Markov inequality we have

$$P_r(\Delta N_A(t) = 1 | \theta_t^A = c) \leq \frac{\Delta t}{c} E[V_t^A | \theta_t^A = c]. \tag{3.29}$$

Averaging for all possible θ_t^A and taking limit as Δt goes to zero result

$$P_r(A_t) = \frac{P_r(dN_A(t) = 1)}{dt} \leq E[\frac{1}{\theta_t^A} V_t^A] \tag{3.30}$$

where the expectation exists because θ_t^A is bounded below by the resting level $v_o > 0$. In fact, we have a precise expression as follows. By a symmetry to $f_B(t, \theta_t^B)$ defined in (3.24), we have

$$f_A(t, \theta_t^A) = \lim_{\Delta t \to 0} \frac{P_r(\Delta N_A(t) = 1 | \theta_t^A; \mathcal{H}_t^A, \mathcal{N}_t^A)}{\Delta t}. \tag{3.31}$$

Note that whereas $f_A(t, \theta_t^A)$ is *not* $P_A(t)$ defined in (3.7), by Lemma 3.1 we have

$$P_r(A_t) = E[f_A(t, \theta_t^A)] = E[V_t^A \frac{f_{\theta_t^A}(a_t)}{1 - F_{\theta_t^A}(a_t)}] \tag{3.32}$$

where $a_t = \int_{t_0}^{t} V_\tau^A d\tau$, and $f_{\theta_t^A}(\cdot)$, $F_{\theta_t^A}(\cdot)$ are the density and the distribution functions of the threshold of neuron A, respectively.

Similarly, the conditional joint probability of firing can be expressed as

$$P_r(\Delta N_A(t) = 1, \Delta N_B(s) = 1 | \theta_t^A = c) =$$

$$P_r(\int_t^{t+\Delta t} V_\tau^A d\tau \geq \theta_t^A, \Delta N_B(s) = 1 | \theta_t^A = c) \leq \frac{\Delta t}{c} E[V_t^A \Delta N_B(s) | \theta_t^A = c]. \tag{3.33}$$

By Lemma 3.2, we therefore have

$$P_r(A_t, B_s) = \frac{P_r(dN_A(t) = 1, dN_B(s) = 1)}{dt\, ds} \leq e^{h(t,s)} E[\frac{1}{\theta_t^A} V_t^A P_B(s)]. \tag{3.34}$$

As in Eq. (3.32) above, a more precise expression can be written as

$$P_r(A_t, B_s) = e^{h(t,s)} E[f_A(t, \theta_t^A) P_B(s)]. \tag{3.35}$$

Since the firing probability density of the presynaptic neuron is

$$P_r(B_s) = E[P_B(s)], \tag{3.36}$$

then combining equations (3.35), (3.32) and (3.36) gives Result 1 with

$$\gamma(t,s) = \frac{E[f_A(t, \theta_t^A) P_B(s)]}{E[f_A(t, \theta_t^A)]\; E[P_B(s)]}. \tag{3.37}$$

3.3.3 Discussion of Result 2

The factor $\gamma(t,s)$ reflects our ignorance of the input to neuron B, or that of the knowledge of the threshold θ_t^A. For a completely known input $\{V_s^B\}$ (hence

$P_B(s)$ is determined), $\gamma(t,s) = 1$; for a completely known threshold, $\gamma(t,s)$ is a constant. When the activity of more neurons are known, the uncertainty in the input and/or threshold decreases, and $\gamma(t,s)$ approaches 1. For instance, if the activities of more interacting neurons $(C_1, C_2, \cdots, C_m)$ are available, we can use a multiunit PST histogram in addition to the conventional joint and individual histograms to estimate

$$\frac{P_r(A_t, B_s, \mathbf{C}_t)P_r(\mathbf{C}_t)}{P_r(A_t, \mathbf{C}_t)P_r(B_s, \mathbf{C}_t)} = \frac{P_r(A_t, B_s|\mathbf{C}_t)}{P_r(A_t|\mathbf{C}_t)P_r(B_s|\mathbf{C}_t)} = \gamma^*(t,s)e^{h(t,s)} \tag{3.38}$$

where $\gamma^*(t,s)$ is defined in Eq. (3.17).

Because neurons $(C_1, C_2, \cdots, C_m)$ may contain information about $P_B(s)$ and/or θ_t^A (for instance, if these neurons influence the activity of either or both neurons A and B), observing more interacting neurons makes $f_A(t, \theta_t^A)$ less correlated with $P_B(s)$. Consequently, observing more neurons makes γ^* closer to one than γ is in Eq. (3.37), and hence the estimator for $h(t,s)$ is more reliable.

3.3.4 Discussion of Result 3

An important factor in correctly interpreting the correlations among the activities of different cells concerns the effects of the stimulus. Specifically, this refers to the fact that unconnected cells may exhibit strong correlations in their firings purely due to the fact that they are driven by the same stimulus. In order to eliminate these effects, some form of normalization is necessary. In Result 3 we show how the stimulus shuffle alone fails to accomplish this task. In order to illustrate this with explicit analytic expressions, three simplifying assumptions will be adopted concerning the properties of the postsynaptic

neuron threshold θ_t^A (used in Theorem 3.1 below) and the distribution of the presynaptic potential (used in Theorem 3.3). We start by stating two of these assumptions and the theorems associated with them, and then proceed to relate the correlation functions explicitly to the inter-neuronal connectivity ($h(t,s)$) in a pair of neurons (A and B).

Assumption 3.1. The random variable Z of the threshold in Eq. (3.4) is independent of V_t^A, and has an exponential pdf:

$$p_Z(z) = \begin{cases} \theta_o e^{-(\theta_o z - v_o)}, & v_o/\theta_o \leq z < \infty \\ 0, & z < v_o/\theta_o \end{cases} \tag{3.39}$$

where $\theta_o > 0$, and v_o is a resting level of the membrane potential.

This assumption is typically valid in cases where neuron A is only related to neuron B, i.e., it is weakly related to any other neuron. It can be verified that under Assumption 3.1, the output spike train of the neuron A is a doubly stochastic *Poisson* process $\{N_A(t) : t \geq 0\}$ with the intensity process $\{\Lambda_t^A : t \geq 0\}$, where

$$\Lambda_t^A = \begin{cases} 0, & T_k \leq t < T_k + r \\ V_t^A, & T_k + r \leq t < T_{k+1} \end{cases} \tag{3.40}$$

where r is the refractory period.

For a doubly stochastic Poisson process $\{N_A(t)\}$, we have

$$P_r(\Delta N_A(t) = 0 | \mathcal{H}_t^A, \mathcal{N}_t^A) = 1 - \Lambda_t^A \Delta t + o(\Delta t),) \tag{3.41}$$

$$P_r(\Delta N_A(t) = 1 | \mathcal{H}_t^A, \mathcal{N}_t^A) = \Lambda_t^A \Delta t + o(\Delta t), \tag{3.42}$$

$$P_r(\Delta N_A(t) > 1 | \mathcal{H}_t^A, \mathcal{N}_t^A) = o(\Delta t)) \tag{3.43}$$

Note that Λ_t^A depends on $\mathcal{N}_t^A$, and hence $\{N_A(t) : t \geq 0\}$ is a self-exciting process with the intensity function $E[\Lambda_t^A|\mathcal{N}_t^A]$, [35].

Assumption 3.2. The refractory period is much smaller than any interspike interval and hence is negligible.

Under this assumption Λ_t^A does not depend on $\mathcal{N}_t^A$, and hence the intensity process becomes the membrane potential process. And in this case, the conditional probabilities in (3.41), (3.42), and (3.43) do not depend on $\mathcal{N}_t^A$.

Theorem 3.1. Under Assumptions 3.1 and 3.2, $\{N_A(t) : t \geq 0\}$ is a doubly stochastic Poisson process with the intensity process $\{V_t^A : t \geq 0\}$. Furthermore, the conditional joint probability density of firing of neurons A and B can be expressed as

$$P_{AB}(t,s) \;=\; P_A(t)P_B(s)e^{h(t,s)} \tag{3.44}$$

for all t and all s, where $h(t,s)$ is the inter-neuronal connectivity with non-zero transmission delay.

Theorem 3.1 states that the joint probability density of firing can be expressed as the product of the individual firing densities and the connectivity. Thus the inter-neuronal connectivity $h(t,s)$ can be directly identified by

$$h(t,s) = \log P_{AB}(t,s) - \log P_A(t) - \log P_B(s). \tag{3.45}$$

This is an ideal case.

Experimentally, if only neurons A and B are recorded, the semi-membrane potential of the presynaptic neuron B is generally unknown (hence P_{AB}, P_A, and P_B are unknown), because the membrane potentials are unobservable in

extracellular recordings. Therefore, one must instead evaluate the normalized *unconditional* joint probability density

$$N_p(t,s) = \frac{E[P_{AB}(t,s)]}{E[P_A(t)]E[P_B(s)]}. \tag{3.46}$$

Let us recall that the PST histogram of neuron A estimates $E[P_A(t)]\Delta t$ and that of neuron B estimates $E[P_B(t)]\Delta t$, and the joint PST histogram estimates $E[P_{AB}(t,s)]\Delta t\Delta s$. Therefore, $N_p(t,s)$ can be estimated by these three histograms.

Lemma 3.3. Under Assumptions 3.1 and 3.2, the conditional firing probability density of the postsynaptic neuron can be expressed as

$$P_A(t) = E[V_t^A|\mathcal{H}_t^B] \;=\; \alpha \exp\{\int_0^t (e^{h(t,\tau)} - 1)V_\tau^B d\tau\}. \tag{3.47}$$

Theorem 3.2. If Assumptions 3.1 and 3.2 hold, the uncertainty γ in Result 1 can be expressed as

$$\gamma(t,s) = \frac{E[V_s^B \exp\{\int_0^t (e^{h(t,\tau)} - 1)V_\tau^B d\tau\}]}{E[V_s^B]\; E[\exp\{\int_0^t (e^{h(t,\tau)} - 1)V_\tau^B d\tau\}]}. \tag{3.48}$$

Theorem 3.3. If the semi-membrane potential process of the presynaptic neuron can be decomposed in the form of $V_t^B = Xf(t)$ where X is a positive random variable, and $f(t)$ is a deterministic time function, then

$$\gamma(t,s) = \frac{M'(\eta_t)}{E[X]\; M(\eta_t)} \tag{3.49}$$

where $M(\cdot)$ is denoted as the moment generating function of X, and $M'(\eta_t)$ is the first derivative of $M(\cdot)$ with respect to η_t which is expressed by

$$\eta_t = \int_0^t (e^{h(t,\tau)} - 1) f(\tau) d\tau. \tag{3.50}$$

Furthermore, $\gamma(t,s) \to 1$ when $Var(X) \to 0$.

Assumption 3.3. The presynaptic membrane potential X is Gamma distributed with parameters (λ, ν).

One consequence of Theorem 3.3 is that if Assumption 3.3 holds — a relatively common occurrence [6], [27], and [8] — the normalized unconditional joint probability density $N_p(t,s)$ can be explicitly evaluated in terms of these parameters as

$$N_p(t,s) = \frac{\lambda}{\lambda - \eta_t} e^{h(t,s)}, \quad \eta_t < \lambda. \tag{3.51}$$

Comparing this expression with that of Theorem 3.2 suggests that $\gamma(t,s) = \lambda/(\lambda - \eta_t)$. Therefore, for a given Gamma distribution (of degree ν), as the variance of X $(= \nu/\lambda^2)$ becomes smaller, λ increases, and $\gamma(t,s) \to 1$. In other words, the more is known about V_s^B (e.g., from recordings of additional neurons), the more accurate is the estimate of the connectivity between neurons A and B. We will illustrate these result through simulations later in section 3.5 (see Fig. 3.5).

If the membrane potential does not vary much for different stimulus presentation (small variance of X), then $\lambda \gg \eta_t$. Consequently, we have

$$N_p(t,s) \simeq e^{h(t,s)}. \tag{3.52}$$

This confirms the conclusions established in Result 2 earlier.

In contrast to the normalization used in Eq. (3.18), the conventional cross-covariance histogram (which is the modified joint PST diagram using the shuffling method) uses a *difference* normalization which estimates [11], [17]

$$N_d(t,s) = E[P_{AB}(t,s)] - E[P_A(t)]\, E[P_B(s)]. \tag{3.53}$$

In general, this expression is very complicated. However, if we make use of the assumptions in Theorem 3.3 for the intensity process of the presynaptic neuron (i.e., a Gamma distribution), it reduces to

$$N_d(t,s) = \alpha\nu f(s) M(\eta_t)\left(\frac{e^{h(t,s)}}{\lambda - \eta_t} - \frac{1}{\lambda}\right). \tag{3.54}$$

If the membrane potential is not varying too much for different stimulus presentation ($\lambda \gg \eta_t$), then $N_d(t,s)$ can be approximately written as

$$N_d(t,s) \simeq \frac{\alpha\nu f(s)}{\lambda}\left(e^{h(t,s)} - 1\right). \tag{3.55}$$

This expression suggests that identifying the connectivity here is considerably more difficult than that of the normalization $N_p(t,s)$ used earlier, since quantities α, ν, λ and function $f(s)$ are generally unknown. Nevertheless, Eq. (3.55) suggests that the shuffling method remains effective in indicating *the absence of a direct connection* (i.e., when $h(t,s)$ is very small), since in that case $N_d(t,s)$ is approximately zero regardless of the confounding terms (α, ν, λ and function $f(s)$). We will illustrate this in simulations in section 3.5 (see Fig. 3.6).

3.3.5 Discussion of Result 4

In the following discussion we consider the case where the observed neurons have no direct connections, but a common input source as depicted in Fig. 3.2. Hence neurons A and B are parallel with respect to the common source.

Gerstein pointed out in [11] that the *narrow peaks* in histograms would favor direct connection between two observed neurons over common input to the two neurons. This observation is confirmed quantitatively as the result indicates.

Now we prove part (a). If the common input source represents a deterministic stimulus, and if thresholds θ_t^A and θ_s^B are statistically independent, then neurons A and B fire independently for the given stimulus. In this case,

$$P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)}) = P_r(A_t|\mathcal{H}^D_{\max(t,s)}) \cdot P_r(B_s|\mathcal{H}^D_{\max(t,s)})$$

$$= P_r(A_t|\mathcal{H}^D_t) \cdot P_r(B_s|\mathcal{H}^D_s) \tag{3.56}$$

where the second equality is due to the fact that A_t (or B_s) does not depend on future information of the intensity process V_t^D (or V_s^D). This characteristic describes a physically realistic system. Therefore, part (a) of Result 4 is true.

Next, suppose that the common input represents another neuron with firing rate V_t^D, and spike train D is represented by $\{T_k^D : k = 1, 2, \cdots\}$ and $\{N_D(t) : t \geq 0\}$. For simplicity, we make use of Assumptions 3.1 and 3.2. Then by Theorem 3.1, spike trains A and B are represented by doubly stochastic Poisson processes. Therefore,

$$P_r(A_t|\mathcal{H}^D_t) = E[V_t^A|\mathcal{H}^D_t] \tag{3.57}$$

and

$$P_r(B_s|\mathcal{H}_s^D) = E[V_s^B|\mathcal{H}_s^D]. \tag{3.58}$$

Without loss of generality, we assume $t \geq s$. Thus

$$P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)}) = E[V_t^A\ V_s^B|\mathcal{H}_t^D]$$

$$= \alpha^2 E[\exp\{\sum_{k=1}^{N_D(s)}[h_A(t,T_k^D)+h_B(s,T_k^D)]\}\cdot\exp\{\sum_{k=N_D(s)+1}^{N_D(t)}h_A(t,T_k^D)\}]. \tag{3.59}$$

Define an event $D_{mn} = \{N_D(t) = n+m, N_D(s) = m\}$ which has the probability

$$P_r(D_{mn}) = \frac{[\int_0^s V_\tau^D d\tau]^m}{m!}\frac{[\int_s^t V_\tau^D d\tau]^n}{n!}e^{-\int_0^t V_\tau^D d\tau}. \tag{3.60}$$

Hence, we have

$$P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)}) = \alpha^2\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}E[\exp\{\sum_{k=1}^{N_D(s)}[h_A(t,T_k^D)+h_B(s,T_k^D)]\}|D_{mn}]$$

$$\times E[\exp\{\sum_{k=N_D(s)+1}^{N_D(t)}h_A(t,T_k^D)\}|D_{mn}]P_r(D_{mn})$$

$$= \alpha^2\exp\{\int_0^s e^{[h_A(t,\tau)+h_B(s,\tau)]}V_\tau^D d\tau + \int_s^t e^{h_A(t,\tau)}V_\tau^D d\tau - \int_0^t V_\tau^D d\tau\}$$

$$= \alpha^2\exp\{\int_0^s (e^{[h_A(t,\tau)+h_B(s,\tau)]}-1)V_\tau^D d\tau + \int_s^t (e^{h_A(t,\tau)}-1)V_\tau^D d\tau\}. \tag{3.61}$$

And using Lemma 3.3 proves part (b) of the result.

A simple glance at the above equation confirms that the joint PST histograms of two neurons are more spread with common input than with direct connection. By result 4 the area far from the principal diagonal in

the normalized joint PST diagram should be flat because $|t-s| \gg 1$ implies that one of the factor in the integral of (3.23) is close to zero.

Let us investigate further for the time invariant connectivities. Suppose that $h_A(t) = w_A e^{-\sigma t}$ and $h_B(t) = w_B e^{-\sigma t}$, and $V_t^D = 1$, $\forall\, t$. It turns out that Eqs. (3.23) and (3.61) can be expressed as

$$\frac{P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)})}{P_r(A_t|\mathcal{H}^D_t)\ P_r(B_s|\mathcal{H}^D_s)} = \exp\{\frac{1}{\sigma}\sum_{n=1}^{\infty}(a_n - 1)(1 - e^{-n\sigma(t\wedge s)})\frac{w_A^n}{n\cdot n!}\} \qquad (3.62)$$

and

$$P_r(A_t, B_s|\mathcal{H}^D_{\max(t,s)}) = \alpha^2 \exp\{\frac{1}{\sigma}\sum_{n=1}^{\infty}(a_n + 1)(1 - e^{-n\sigma(t\wedge s)})\frac{w_A^n}{n\cdot n!}\}, \qquad (3.63)$$

respectively, where

$$a_n = (a+1)^n - a^n\ ;\ \ a = \frac{w_B}{w_A}e^{-\sigma|t-s|}. \qquad (3.64)$$

Figure 3.3 compares three curves: one for the direct synaptic excitatory connectivity ($e^{h(t,s)}$), one for the common input connection without the normalization procedure (Eq. (3.63)), and the other for the common input connection with the normalization procedure (Eq. (3.62)). All the connectivities in Fig. 3.3 have the form

$$h_A(t) = h_B(t) = h(t) = 2.4e^{-100t},$$

and the time bin width is $\Delta t = 0.002$. It is shown that the common input correlation has a wider peak than the synaptic correlation does, and that the normalization procedure makes the common input correlation peak smaller.

3.4 Experimental Considerations

In the analysis of multineuronal connectivities, spike trains from several neurons are recorded in response to the repeated presentation (e.g., R times) of a stimulus. Spikes are usually sampled and parsed into (i.e., labeled by) small time bins, using the onset of the stimulus as the initial bin. The bin width Δt is always chosen to be so small that at most one spike may occur in each bin (which corresponds to the orderliness of the point process). Thus each spike train is converted into a discrete 0-1 process, and is further segmented into R segments, each for one stimulus presentation.

Let A_{rn} be the time bin corresponding to the n-th bin associated with the r-th stimulus presentation. A spike train can then be represented by a $R \times N$ random matrix $\mathbf{A}$ with elements $(A_{rn},\ r = 1, 2, \cdots, R;\ n = 1, 2, \cdots, N)$ — called here a *spike matrix*. Let us assume that the firing activity during each stimulus presentation is statistically independent. Therefore, each element is a random variable taking values $\{0, 1\}$, and the elements in the same column are independent and identically distributed.

The PST histogram $(H_n,\ n = 1, 2, \cdots, N)$ reflects the stimulus-locked firing rate of each single neuron, and it is formed by taking average over every column of the spike matrix,

$$H_n^A = \frac{1}{R}\sum_{r=1}^{R} A_{rn},\ \ n = 1, 2, \cdots, N. \tag{3.65}$$

The value of H_n^A counts the average spikes over R stimulus presentations in the n-th bin in a spike train A.

The joint PST scatter diagram of two neurons A and B (H_{mn}^{AB}, $m = 1,2\cdots,N$; $n = 1,2,\cdots,N$) measures the coincidence spikes in train A and in train B relative to stimulus onset. It is a two-dimensional histogram with one axis (m) for train A and the other axis (n) for train B, and hence it is an N square matrix $\mathbf{H}$. Element H_{mn}^{AB} represents the average count for coincidence of a spike in the m-th bin of train A and a spike in the n-th bin of train B over R stimulus presentations, that is,

$$H_{mn}^{AB} = \frac{1}{R}\sum_{r=1}^{R} A_{rm}B_{rn},\ m = 1,2\cdots,N;\ n = 1,2,\cdots,N \tag{3.66}$$

where A_{rm} and B_{rn} are the elements of spike matrices for trains A and B, respectively. Therefore, the matrix presentation of the joint PST scatter diagram is

$$\mathbf{H} = \frac{1}{R}\mathbf{A}^T\mathbf{B} \tag{3.67}$$

where T denotes transposition. The expanded joint PST histogram for multi-unit recordings (of M neurons) is then

$$H_{n_1 n_2\cdots n_M}^{C_1 C_2\cdots C_M} = \frac{1}{R}\sum_{r=1}^{R} C_{rn_1}^{1} C_{rn_2}^{2}\cdots C_{rn_M}^{M} \tag{3.68}$$

where $C_{rn_i}^{i}$, $(i = 1,2,\cdots,M)$, is the element of the spike matrix for the i-th neuron.

3.4.1 Using the Scatter Plot to Determine Neuronal Connectivities

The correlations between a pair of recorded neurons (A and B) can be computed from the experimental estimate of the expression of Result 1, i.e.,

$$h(t,s) = \log(\frac{N_p(t,s)}{\gamma(t,s)}) = \log(\frac{E[P_{AB}(t,s)]}{E[P_A(t)]\ E[P_B(s)]}) - \log(\gamma(t,s))$$

where $E[P_A(t)]$ and $E[P_B(s)]$ represent the PST histograms of firings of the neuron pair, $E[P_{AB}(t,s)]$ is their scatter plot, and γ (≥ 0) is the corrupting factor representing the uncertainty in the estimate due to the influences of other unobserved neurons and biophysical factors. Thus in terms of bin numbers m and n, the above equation can be written as

$$\log(\frac{H^{AB}_{mn}}{H^A_m\ H^B_n}) = h(m\Delta t, n\Delta t) + \log(\gamma(m\Delta t, n\Delta t)). \tag{3.69}$$

In the case of time invariant connectivities, $h(t,s)$ becomes $h(t-s)$, and the correlation peak becomes a band that runs parallel to the principal diagonal ($t-s=0$).[1]

In the practical application of Eq. (3.69), the confounding $\gamma(t,s)$ contributions are not known. However, the analysis shows that additional simultaneous recordings can be used to reduce these uncertainties. Therefore, by using the additional data, the improved estimator for $h(t,s)$ becomes

$$\log(\frac{H^{ABC_3\cdots C_M}_{mnm\cdots m}\ H^{C_3\cdots C_M}_{m\cdots m}}{H^{AC_3\cdots C_M}_{mm\cdots m}\ H^{BC_3\cdots C_M}_{nm\cdots m}}) = h(m\Delta t, n\Delta t) + \log(\gamma^*(m\Delta t, n\Delta t)) \tag{3.70}$$

[1]Note that one can detect further correlations in the *unnormalized* scatter plot, such as the more diffuse bands of time-invariant common inputs. Of course, these features are intentionally removed by the normalization since they do not reflect direct connectivities between the neuron pair (see Result 4).

where $H_{n_1 n_2 \cdots n_M}^{C_1 C_2 \cdots C_M}$ are simply the joint multidimensional scatter plots defined in Eq. (3.68), and the uncertainty factor γ^* (better than γ) is defined in Eq. (3.17). The estimates of Eqs. (3.69) and (3.70) are illustrated in network simulations in section 3.5.

3.4.2 Establishing Confidence Measures on the Estimates

The histograms are random variables subject to fluctuations. Hence, it is important to determine upper and lower bounds such that we assume a connection between neurons A and B whenever these bounds are surpassed. By the law of large numbers, H_{mn}^{AB} converges to $E[P_{AB}(t,s)]$, so does H_m^A to $E[P_A(t)]$ and H_n^B to $E[P_B(s)]$ almost surely as $R \to \infty$. Therefore, if neurons A and B are independent, by theorems on limiting distributions,

$$\frac{H_{mn}^{AB}}{H_m^A H_n^B} \to 1 \ \ as \ \ R \to \infty \tag{3.71}$$

almost surely.

The hypothesis $\mathcal{H}_0$ is that the two neurons are statistically independent, which is supported by

$$E[P_{AB}(t,s)] = E[P_A(t)]\ E[P_B(s)]. \tag{3.72}$$

And the alternative hypothesis $\mathcal{H}_1$ is that the two neurons depend, which is described by

$$E[P_{AB}(t,s)] \neq E[P_A(t)]\ E[P_B(s)]. \tag{3.73}$$

One expects $H_{mn}^{AB}/(H_m^A H_n^B)$ to be close to 1 if hypothesis $\mathcal{H}_0$ is true. Conversely, if the amount it deviates from 1 exceeds a bound b, one accepts hypothesis $\mathcal{H}_1$.

Now for a given significance level α, we need to find the bound b satisfying

$$P_r(|\frac{H_{mn}^{AB}}{H_m^A H_n^B} - 1| > b \mid H_m^A, H_n^B; \mathcal{H}_0) = \alpha. \tag{3.74}$$

The hypothesis testing is stated as the following theorem.

Theorem 3.4. Let b be a bound which divides a critical region for the hypothesis testing. One announces that there is a dependence between the two observed neurons if

$$|\frac{H_{mn}^{AB}}{H_m^A H_n^B} - 1| > b. \tag{3.75}$$

For the given significance level α of false announcement of dependence, the bound can be approximately calculated by

$$b \simeq \varepsilon_b \sqrt{\frac{1 - H_m^A H_n^B}{R H_m^A H_n^B}} \tag{3.76}$$

where the value of ε_b is determined from

$$\Phi(\varepsilon_b) = 1 - \frac{\alpha}{2} \tag{3.77}$$

and $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$.

The function $\Phi(x)$ is usually available as the standard normal distribution table. For example, $\alpha = 0.05$ gives $\varepsilon_b = 1.96$.

The above theorem implies that element $H_{mn}^{AB}/(H_m^A H_n^B)$ of the normalized joint PST diagram has a conditional expectation value 1 and an approximate conditional variance

$$\sigma_{mn}^2 \simeq \frac{1 - H_m^A H_n^B}{R H_m^A H_n^B} \tag{3.78}$$

given the values of H_m^A and H_n^B under hypothesis $\mathcal{H}_0$. Since H_m^A and H_n^B are usually very small and R is fairly large, this approximation is close to a recent result by Palm et al. [30] where their conditional variance is

$$\sigma_{mn}^2 = \frac{(1 - H_m^A)(1 - H_n^B)}{(R-1)H_m^A H_n^B} \tag{3.79}$$

under hypothesis $\mathcal{H}_0$.

The bound dividing the hypothesis regions can be made more useful in neural networks with time invariant connectivities. Let w_m reflect the fluctuation in the normalized joint PST diagrams such that

$$\frac{H_{mn}^{AB}}{H_m^A H_n^B} = \gamma(m\Delta t, n\Delta t)e^{h((m-n)\Delta t)} + w_m \ , \tag{3.80}$$

and the mean of w_n is zero. Let $k = m - n$. A collapsed version can be generated by averaging over diagonals of the normalized joint PST diagram. This collapsed version is a 1-dimensional histogram G_k expressed by

$$G_k = \frac{1}{N - |k|} \sum_{n=\max(1,1-k)}^{\min(N,N-k)} \frac{H_{n+k,n}^{AB}}{H_{n+k}^A H_n^B}$$

$$k = -N+1, \cdots, -1, 0, 1, \cdots, N-1 \tag{3.81}$$

where $k = 0$ is the collapsed point of the principal diagonal.

Since averaging reduces the fluctuations (the average of w_m has a smaller variance), G_k is a better estimator for the time invariant connectivity $h(t,s) = h(t-s)$. This enables us to establish a bound such that

$$P_r(|G_k - 1| > b_k \mid \mathcal{H}_0) = \alpha. \tag{3.82}$$

Theorem 3.5. Given a significance level α, let b_k be a bound of critical region satisfying the above equation, then b_k may be approximately written as

$$b_k \simeq \frac{\sqrt{\sum_{n=\max(1,1-k)}^{\min(N,N-k)} \sigma_{n+k,n}^2}}{N-|k|}\,\varepsilon_b \tag{3.83}$$

where ε_b is the the same as in Theorem 3.4, and σ_{mn}^2 is given by (3.78). Furthermore, b_k will reduce to

$$b_k \simeq \frac{b}{\sqrt{N-|k|}} \tag{3.84}$$

when σ_{mn}^2's are taken as constants.

This theorem indicates that the critical region is enlarged (the bound value decreases) when the collapsed version of the normalized joint PST histogram is used.

3.5 Simulations and Discussion

In order to illustrate the nature of the estimates, uncertainties, and bounds derived earlier, we show the results from simulations of networks of excitatory and inhibitory neurons. The neuron model used for the simulations is depicted in Fig. 3.1(c) where the nonlinearity $g(x) = \alpha e^x$ and the random threshold has an exponential distribution with mean 1.

In the first case (Fig. 3.4), pairwise excitatory and inhibitory, time invariant connections are estimated using the normalized scatter plots; the uncertainty

factor (γ) is equal to 1. The upper plots show the two-dimensional normalized scatter plots. The correlations appear as bands along the principal diagonal because $h(t,s)$ is time-invariant. Hence, the scatter plot can be collapsed along this axis to produce the lower histograms. Note that time-variations in $h(t,s)$ (e.g., due to poststimulus adaptation) do not allow this reduction. Consequently, it should only be performed on the portions of the neural record that display obvious stationary behavior. In both simulations of Fig. 3.4, the predicted analytical estimates are also plotted for comparison, together with the bound lines for the confidence measures (determined by Theorem 3.5).

In order to illustrate the effects of the uncertainty factor γ, we examine in Fig. 3.5 the interactions among three neurons with time invariant connectivities. Here, neuron A is inhibited by neuron B and excited by neuron C, and neuron C is in turn excited by neuron B. Because of the interactions between B and C, the threshold in neuron A is no longer independent of the firings of B. Thus, if we attempt to identify the connectivity between neurons A and B from pairwise recordings, the estimates will be contaminated by the γ uncertainty factor. The top curve in Fig. 3.5 first shows the "target" theoretical connectivity obtained from the multi-recording estimate given by formula (3.38) with $\gamma^*(t,s) = 1$ (i.e., $e^{h_{AB}(t-s)}$). If neuron C is ignored, the pairwise estimate of $e^{h_{AB}(t-s)}$ is shown as the middle curve in Fig. 3.5 (corresponding to formula (3.69)). The correlation is so distorted that actual inhibition becomes false excitation because of the strong excitatory activity from neuron C. To correct the erroneous correlation, we have to use the information from the third neuron. The tripartite correlation according to formula (3.70) is displayed at bottom of Fig. 3.5, which is much closer to the analytical estimate.

Fig. 3.6 compares the preferred normalization with the difference normalization (shuffle method) under two situations. In the absence of a direct connection, the shuffle method provides accurate indication of the lack of synaptic inputs between the two neurons. However, in the presence of a direct connection, the shuffle method fails to remove completely the stimulus correlations as indicated by the deviation from the analytical results. Instead, the normalization suggested in this paper performs well in both cases.

In conclusion, the above simulations confirm the proposed theory. The neuron model adopted is quite general because (1) the synaptic connectivity $h(t,s)$ represents a time-varying system; (2) the processes representing spike trains are not necessarily Poisson processes, and (3) the nonlinear function $g(x) = \alpha e^x$ is an approximation of $\alpha e^x/1 + \alpha e^x$ when $\alpha e^x \ll 1$, meaning that the neuron is operating at low firing rates. Moreover, our analytical Results 1 and 2 do not dependent on any further assumptions. Although the three simplifying assumptions were made in order to see Result 3 more clearly, we did not use these assumptions in the simulations of Fig. 3.6.

The analysis presented in this paper also points to the following sobering conclusion: For multiunit correlation analysis to play a useful role in establishing the basic circuitry of the nervous system, new technologies have to be developed for *stable, multiunit* recordings. These requirements stem from the need for extended simultaneous recordings from many cells in order to construct adequate scatter histograms and to minimize inherent uncertainty due to unobserved but related activities. Unfortunately, neither of these requirements are easily met at present, although extensive efforts towards this goal are underway through the use of silicon-based microelectrode arrays [22].

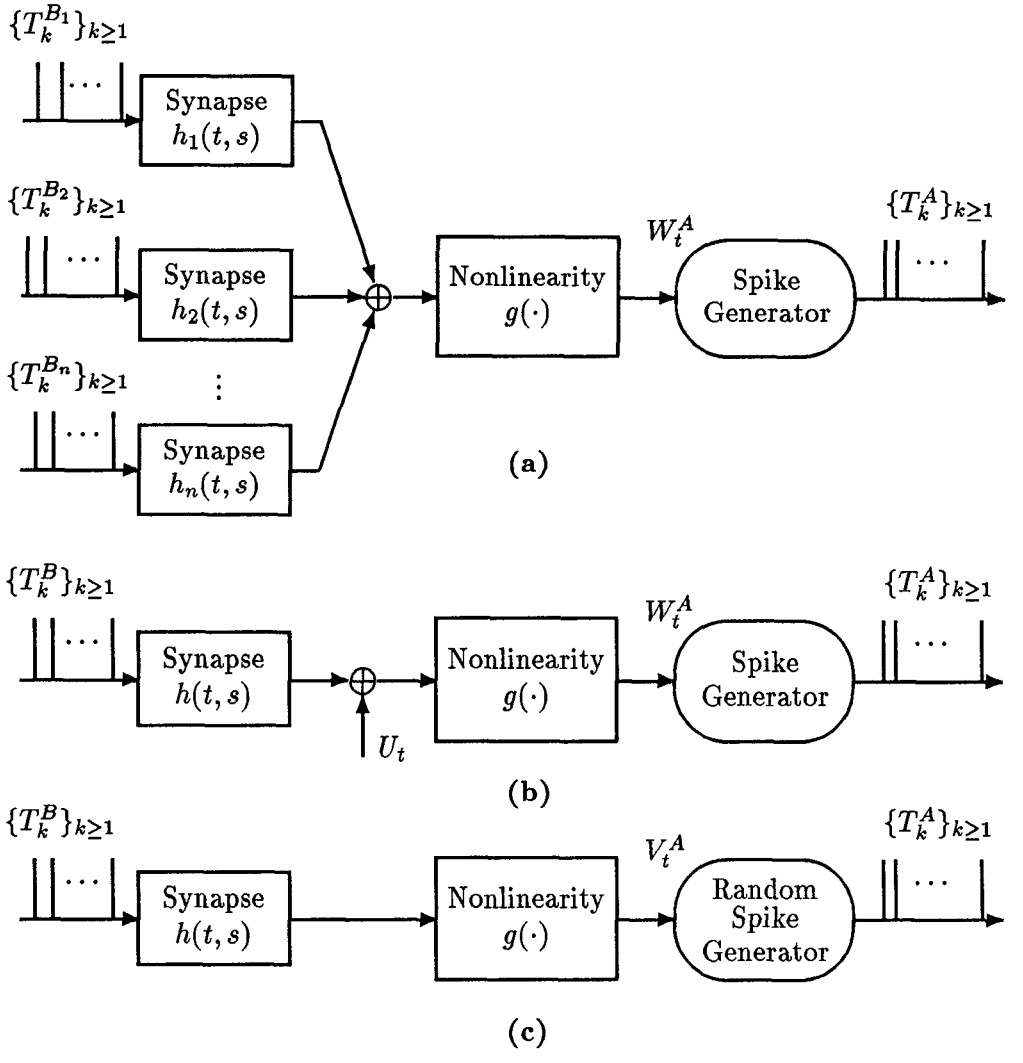


Figure 3.3: A dynamical nonlinear neuron model, where neuron A is considered as the postsynaptic neuron.

(a) Neuron A is influenced by presynaptic neurons $B_1, B_2, \cdots, B_n$.

(b) A synaptic connection between neurons A and B; the influences of other neurons on neuron A are summarized by U_t.

(c) An equivalent probabilistic version of the neuron model. The impact of the random input U_t is now moved to the spike generator where the threshold becomes random.

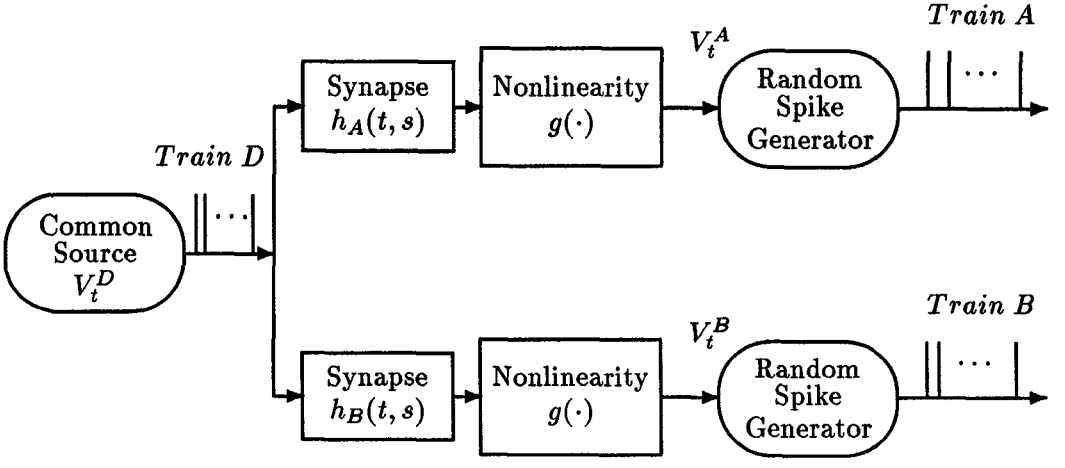


Figure 3.2: A pair of Neurons (A and B) are stimulated by a common input source, train D. The common input may represent a stimulus source or a neuron. The average rate of train D is V_t^D.

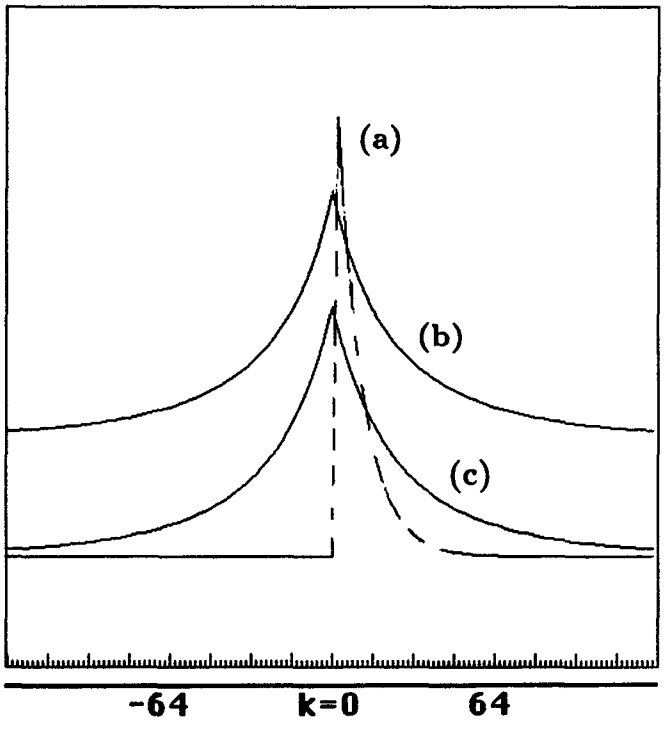


Figure 3.3: The common input correlation vs. the snaptic correlation. The time axis represents $k = t - s$. All the connectivities have the form of $h_A(t) = h_B(t) = h(t) = 2.4e^{-100t}$, and the time bin width $\Delta t = 0.002$.

(a) Narrow correlation peak due to the synaptic connectivity.

(b) Wide correlation peak due to the common input connection.

(c) The normalization procedure makes the wide peak smaller.

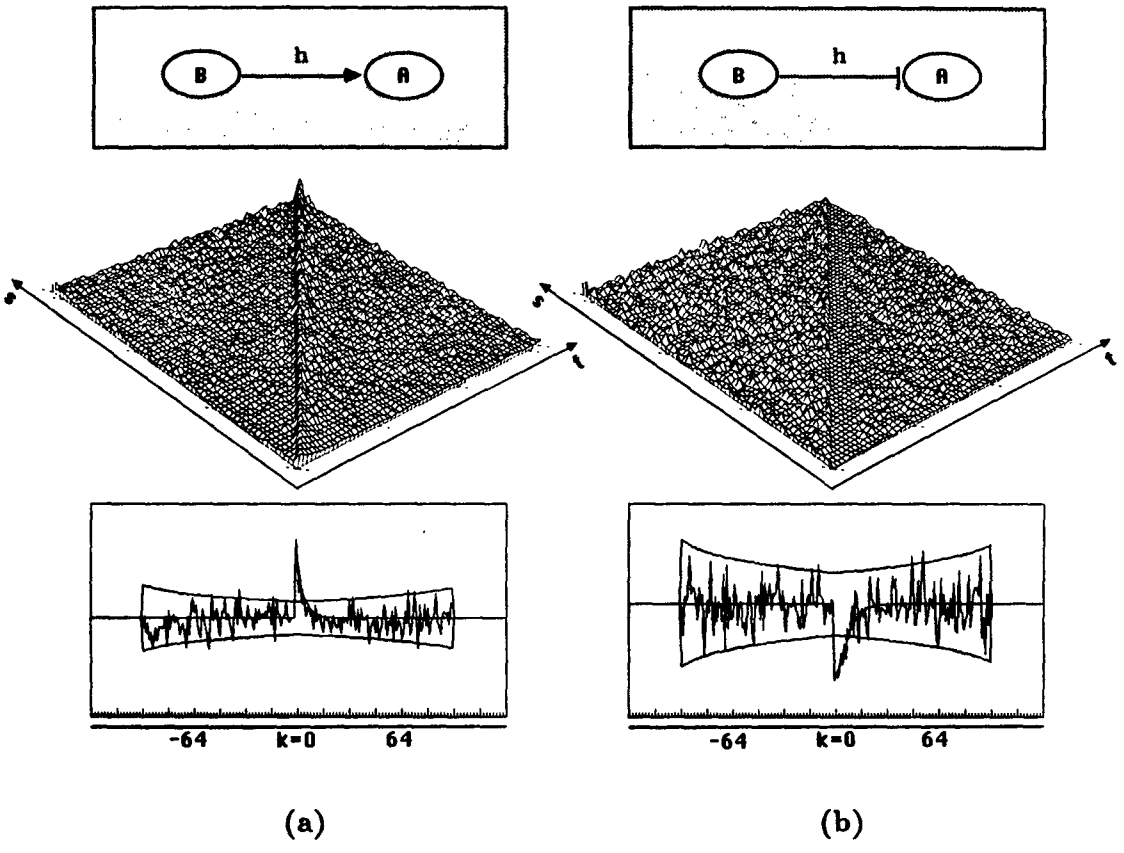


Figure 3.4: Simulations for pairwise excitatory and inhibitory correlations.
(a) Excitatory coupling $h(t,s) = 0.8e^{-20(t-s)}$, $t > s$. Shown is the two-dimensional normalized scatter plot generated by the spike trains of the two neurons; below it is the histogram G_k that results from collapsing the scatter plot along the principal diagonal. It corresponds to the function $N_p(k) = e^{h(k)}$. The upper and lower bound lines represent the 95 % confidence measure.
(b) Inhibitory coupling, similar to (a) for $h(t,s) = -3.0e^{-20(t-s)}$, $t > s$.

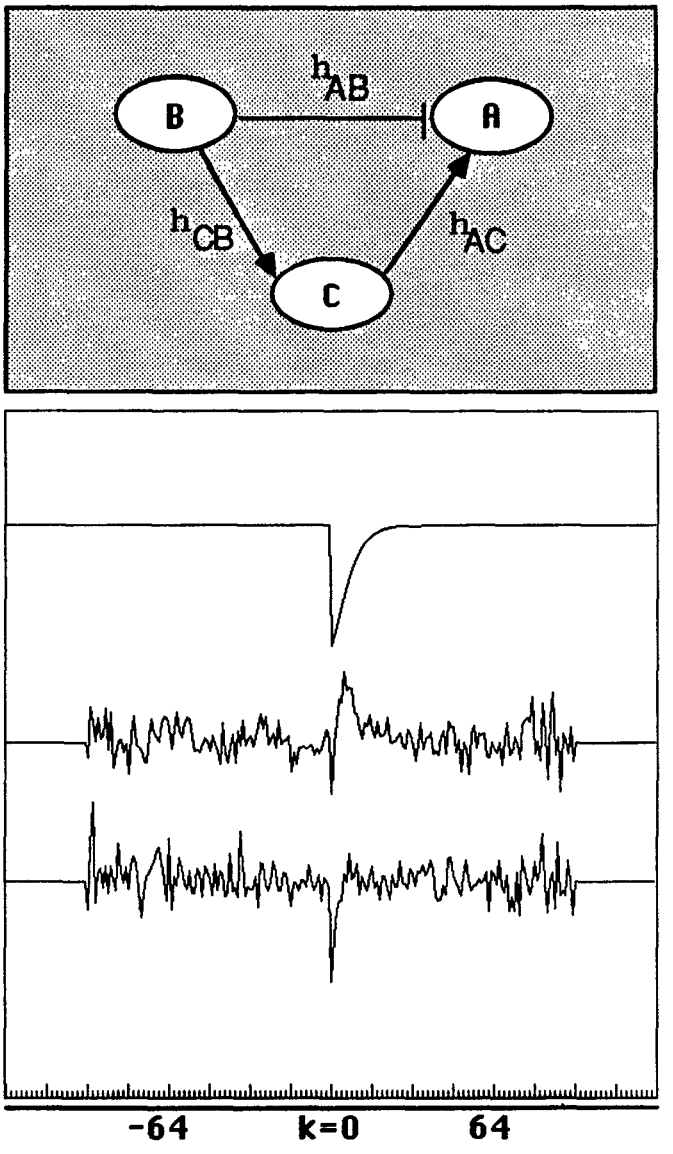


Figure 3.5: Interaction among three neurons. The network structure is displayed on the top graph: neuron B inhibits neuron A and excites neuron C, and neuron C excites neuron A. $h_{AB}(t) = -1.8e^{-20t}$, $h_{AC}(t) = 3.6e^{-20t}$, and $h_{CB}(t) = 2.0e^{-20t}$. The top curve gives the theoretical connectivity from formula (3.38) with $\gamma^*(t,s)$. The middle one is the correlation curve corresponding to formula (3.69) generated from spike trains A and B only. The correlation is so distorted that actual inhibition becomes a false excitation (which is actually due to a strong excitatory input from neuron C). The bottom curve shows the tripartite correlation according to formula (3.70), which displays the correct inhibitory sign for the connectivity.

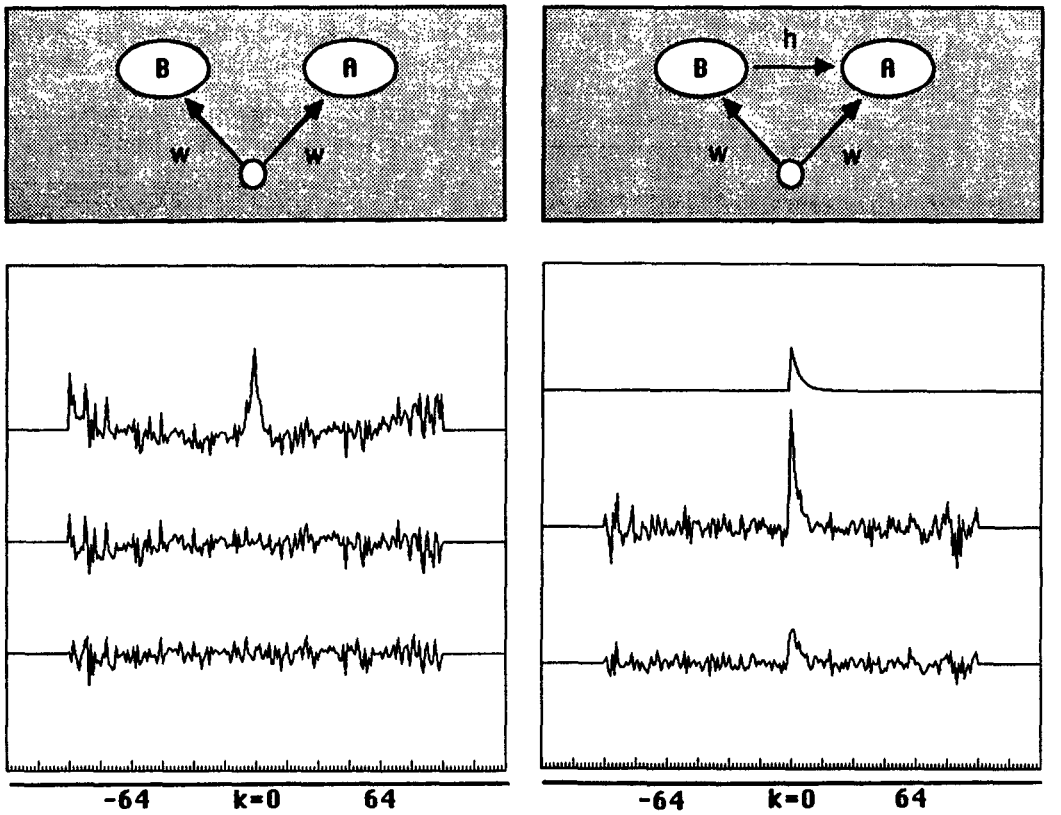


Figure 3.6: Comparison of the preferred with the difference normalizations.
(a) The absence of a direct connection case ($h = 0$): neurons A and B have a common input source — a neuron driven by a stimulus. The connection strength from the common input is $w = 1$. The top curve gives the collapsed version of the joint PST histogram without any normalization. The correlation peak is purely due to stimulus effects. The middle curve represents the difference normalized correlation. The bottom curve shows the preferred normalized correlation curve. Both methods perform well in indicating the absence of connection between A and B.
(b) The presence of a direct connection case ($h \neq 0$): neurons A and B have a common input source as in (a), and in addition, a direct synaptic connectivity from B to A, $h_{AB}(t) = 0.4e^{-20t}$. The top curve gives the theoretical correlation predicted from $N_p(k) = e^{h(k)}$. The middle curve shows the difference normalized correlation. Although the connectivity is weak (only 0.4), the large sharp peak in the correlation leads to a false impression of high excitatory connectivity, which is in fact due to stimulus effects. The bottom curve shows the preferred normalized correlation, which is very close to the theoretical function $0.4e^{-20t}$.

CHAPTER
FOUR
CONCLUDING REMARKS

This dissertation aimed to develop the theoretical and experimental means to study the nature of the neural networks of the nervous system. It covered two areas: the detection and classification of neural signals, and the synaptic connectivity identification of neural networks.

To study the functional connectivity, reliable recording and identification of the simultaneous activities of a group of neurons is essential. In Chapter 2, a totally automated system for neural spike detection and classification was presented which did not require *a priori* assumptions about spike shape or timing. Such a system is desired to overcome some of limitations presented by newly developed extracellular microelectrode arrays where the automation and the signal processing speed are essential.

The system is divided into two parts: a learning subsystem and a real-time detection and classification subsystem. Because its bases have shapes similar to neural spikes, the Haar transform can be used to detect spikes without the knowledge of spike templates. The learning subsystem, comprising a Haar

transform detection scheme, a feature learning phase, and a template learning phase, extracts templates for each separable spike class. The real-time detection and classification subsystem identifies spikes in the noisy neural trace and sorts them into classes, according to the templates and the statistics of the background noise. Three fast algorithms are proposed for the real-time sorting subsystem, and comparisons are made among different schemes. Performance of the system is illustrated by using it to classify spikes in segments of neural activity recorded from monkey motor cortex and from guinea pig and ferret auditory cortices. The system is implemented without human supervision and therefore is suitable for real-time multichannel recording.

In Chapter 3, analytical and experimental methods have been provided for estimating inter-neuronal connectivities from simultaneous recordings of multiple neurons (after separation).

The results are based on detailed, yet flexible nonlinear neuron models in which spike trains are modeled as general doubly stochastic point processes. The expressions derived can be used with nonstationary or stationary records, and can be readily extended from pairwise to multineuron estimates. Furthermore, we have shown analytically how the estimates are improved as more neurons are sampled, and derived the appropriate normalizations to eliminate stimulus-related correlations. In short, the following results have been obtained:

(1) The connectivity can be approximately estimated by the pairwise joint PST histogram and individual histograms.

(2) The connectivity identification can be improved if a third, a forth or more interacting neurons are examed.

(3) It is very important to choose an appropriate normalization method used in the correlation histograms to remove efficiently the stimulus effects.
(4) The correlation peak spreads more for common input connections than for synaptic connections.

Finally, we illustrate how to use analytical expressions for interpreting experiment data, and give explicit confidence measures for detecting inter-neuronal connectivity.

This dissertation suggests that several areas need further investigation. In the spike classification part, it is not infrequent that two or more spikes from different neurons occur simultaneously, causing a superposition of these spike in the recording trace, and hence reducing the reliability of spike separation. Therefore, developing a real-time algorithm to deal with the superposition and integrating it into the automated spike sorting system are important objectives.

In the connectivity identification aspect, the identification accuracy depends on the amount of data collected. This results in two problems: One is the need for huge amount of data; the other is the computational burden. We wish to investigate a new approach for connectivity identification, which is universal for stationary or nonstationary neural firings, and is in addition extremely efficient. This new approach will utilize a hidden Markov model to estimate the intensity process (membrane potentials) of a doubly stochastic process (spike train) so that the synaptic connection between a pair of neurons is readily revealed. At first, we could use the observed spike train of the post-synaptic neuron to estimate the semi-membrane potential process V_t (the states of the neuron at each instant). Then the estimate will be compared with the actual (unknown)

states determined by the observed spike train $\{T_k\}_{k\geq 1}$ of the presynaptic neuron expressed as

$$V_t = g(\sum_{k=1}^{N(t)} h(t, T_k))$$

so as to find the form of the synaptic connectivity $h(\cdot, \cdot)$.

The future objectives of research will focus on investigating the organization of the nervous system. And based on further neurobiological evidence, we hope to develop models with more biological fidelity that can perform various cognitive tasks.

APPENDIX

In this appendix, we prove the theorems and the lemmas in the text.

Proof of Theorem 2.1:

We need to prove that there exists a unique minimum of the objective function $J = \theta P_F + P_M$, if not all $s(k)$ are zero and if $\theta > 1$.

Let us set

$$\frac{\partial J}{\partial x_n} = 0, \; n = 1, 2, \cdots, M. \tag{0.1}$$

Since $P_F = \prod_{m=1}^{M} \Phi(-x_m)$, and $P_M = 1 - \prod_{m=1}^{M} \Phi(\alpha_m - x_m)$, where $x_m = \eta_m/\sigma$, $\alpha_m = s(m)/\sigma$, and $\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, we have

$$\theta\phi(-x_n) \prod_{m \neq n}^{M} \Phi(-x_m) \; = \; \phi(\alpha_n - x_n) \prod_{m \neq n}^{M} \Phi(\alpha_m - x_m), \; n = 1, 2, \cdots, M. \tag{0.2}$$

Using the facts that $\theta > 0$, $0 < \Phi(x) < 1$, and $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, therefore, the M simultaneous equations are equivalent to

$$\nu_n e^{-x_n^2/2} \; = \; e^{-(x_n - \alpha_n)^2/2}, \; \; n = 1, 2, \cdots, M. \tag{0.3}$$

where ν_n's are positive numbers. Suppose that $s(n) \neq 0$, hence $\alpha_n \neq 0$, then

Eq. (0.2) has one and only one solution, which implies that there is only one extremum for J.

To see the extremum is the minimum, notice that

$$\lim_{x_m \to \infty} J = 1, \ \forall m \tag{0.4}$$

and

$$\lim_{x_1 \to -\infty} \lim_{x_2 \to -\infty} \cdots \lim_{x_n \to \infty} J = \theta. \tag{0.5}$$

It suffices to show that there exists an $\underline{x}^*$, such that $J(\underline{x}^*) < 1$. It is easy to see that for fixed $x_2, x_3, \cdots, x_M$, J can be expressed as

$$J = \theta\gamma \int_{-\infty}^{-x_1} \phi(x)dx + 1 - \beta \int_{-\infty}^{\alpha_1 - x_1} \phi(x)dx \tag{0.6}$$

$$= (\theta\gamma - \beta) \int_{-\infty}^{-x_1} \phi(x)dx - \beta \int_{-x_1}^{\alpha_1 - x_1} \phi(x)dx + 1 \tag{0.7}$$

where $0 < \gamma < \beta < 1$, and $\alpha_1 > 0$. Because $\Phi(-x) < e^{-2x^2}$, $for\ x > 0$, we have

$$J < (\theta\gamma - \beta)e^{-2x_1^2} - \frac{\beta\alpha_1}{\sqrt{2\pi}} e^{-x_1^2/2} + 1 < 1 \tag{0.8}$$

for x_1 sufficiently large. This completes the proof.

Proof of Lemma 3.1:

The threshold of neuron B, which is a continuous random variable, has the probability density function and the distribution function denoted $f_{\theta_t^B}(x)$ and $F_{\theta_t^B}(x)$, respectively. Let $b_t = \int_{t_0}^{t} V_\tau^B d\tau$, where t_0 is the occurrence instant of the previous spike. From definition of $P_B(t)$ we have

$$P_B(t) = E_{b_t}[f_B(t, \theta_t^B)] \tag{0.9}$$

where the expectation $E_{b_t}[\cdot]$ is taken with respect to θ_t^B. And

$$f_B(t,\theta_t^B) = \lim_{\Delta t \to 0} \frac{P_r(b_{t+\Delta t} \geq \theta_t^B | b_t < \theta_t^B; \mathcal{H}_t^B, \mathcal{N}_t^B)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P_r(b_t < \theta_t^B \leq b_{t+\Delta t} | \mathcal{H}_t^B, \mathcal{N}_t^B)}{\Delta t \; P_r(b_t < \theta_t^B | \mathcal{H}_t^B, \mathcal{N}_t^B)} = V_t^B \frac{f_{\theta_t^B}(b_t)}{1 - F_{\theta_t^B}(b_t)} \tag{0.10}$$

Furthermore, if the threshold is exponentially distributed with mean λ, then $f_{\theta_t^B}(b_t)/1 - F_{\theta_t^B}(b_t) = \lambda$, and hence $P_B(t) = \lambda V_t^B$. In this case, $P_B(t)$ does not depend on $\{\mathcal{N}_B(t)\}$, and $\{N_B(t)\}_{t \geq 0}$ evolves without aftereffects.

Proof of Lemma 3.2:

Because $\Delta N_B(t)$ can take values 0 and 1 only, by Eq. (3.3) we have

$$E[V_t^A \frac{\Delta N_B(s)}{\Delta s} | \mathcal{H}_{t \vee s}^B, \mathcal{N}_s^B] =$$

$$E[\alpha \exp\{\sum_{k=1}^{N_B(t)} h(t, T_k^B)\} | \Delta N_B(s) = 1; \mathcal{H}_{t \vee s}^B, \mathcal{N}_s^B] \frac{P_r(\Delta N_B(s) = 1 | \mathcal{H}_s^B, \mathcal{N}_s^B)}{\Delta s}. \tag{0.11}$$

For $t > s$, the conditional expectation in the above equation can be written as

$$E[\alpha \exp\{\sum_{k=1}^{N_B(t)} h(t, T_k^B)\} | \Delta N_B(s) = 1; \mathcal{H}_t^B, \mathcal{H}_s^B] =$$

$$E[\alpha \exp\{\sum_{k=1}^{N_B(s)} h(t, T_k^B)\} \exp\{h(t, s + \Delta s)\} \exp\{\sum_{k=N_B(s+\Delta s)+1}^{N_B(t)} h(t, T_k^B)\} | \mathcal{H}_t^B, \mathcal{N}_s^B] \tag{0.12}$$

which becomes

$$e^{h(t,s)} E[V_t^A | \mathcal{H}_t^B, \mathcal{N}_s^B] \tag{0.13}$$

as Δs goes to 0. Since $P_B(s)$ is a measurable function with respect to $\sigma(\mathcal{H}_s^B \times \mathcal{N}_s^B)$, we obtain

$$E[V_t^A \frac{dN_B(s)}{ds}] = e^{h(t,s)} E[E[V_t^A | \mathcal{H}_t^B, \mathcal{N}_s^B] P_B(s)] = e^{h(t,s)} E[V_t^A P_B(s)]. \quad (0.14)$$

For $t \leq s$, we have

$$E[V_t^A \frac{\Delta N_B(s)}{\Delta s} | \mathcal{H}_s^B, \mathcal{N}_s^B] = E[V_t^A | \mathcal{H}_s^B, \mathcal{N}_s^B] \; E[\frac{\Delta N_B(s)}{\Delta s} | \mathcal{H}_s^B, \mathcal{N}_s^B]$$

$$= E[V_t^A | \mathcal{H}_s^B, \mathcal{N}_s^B] \; \frac{P_r(\Delta N_B(s) = 1 | \mathcal{H}_s^B, \mathcal{N}_s^B)}{\Delta s}, \quad (0.15)$$

hence

$$E[V_t^A \frac{dN_B(s)}{ds}] = E[E[V_t^A | \mathcal{H}_s^B, \mathcal{N}_s^B] P_B(s)] = E[V_t^A P_B(s)]. \quad (0.16)$$

Since $h(t,s)$ represents a synaptic connectivity, which is a causal system with non-zero transmission delay, $h(t,s) = 0$ for $t \leq s$. Thus Lemma 3.2 holds for all t and s.

Proof of Theorem 3.1:

Suppose that Assumptions 3.1 and 3.2 hold, and that the threshold θ_t has an exponential distribution with mean 1. By (3.42),

$$\lim_{\Delta t \to 0} \frac{P_r(\Delta N_A(t) = 1 | \mathcal{H}_t^A, \mathcal{N}_t^A)}{\Delta t} = V_t^A, \quad (0.17)$$

hence spike train $N_A(t)$ represents a doubly stochastic Poisson process with the intensity process $\{V_t^A : t \geq 0\}$. Therefore, by Eqs. (3.40) and (3.42) we have

$$P_A(t) = E[V_t^A | \mathcal{H}_t^B]. \quad (0.18)$$

We choose Δt and Δs such that $s < s + \Delta s < t < t + \Delta t$, or $t < t + \Delta t < s < s + \Delta s$. Because Poisson process is an independent increments process, the conditional probability given the firing histories of neurons A and B can be split into

$$P_r(\Delta N_A(t) = 1, \Delta N_B(s) = 1 | \mathcal{H}_t^A, \mathcal{H}_{t \vee s}^B)$$

$$= P_r(\Delta N_A(t) = 1 | \mathcal{H}_t^A)\ P_r(\Delta N_B(s) = 1 | \mathcal{H}_t^A, \mathcal{H}_{t \vee s}^B). \tag{0.19}$$

By Eqs. (3.40) and (3.42), the first factor is

$$P_r(\Delta N_A(t) = 1 | \mathcal{H}_t^A) = V_t^A \Delta t + o(\Delta t). \tag{0.20}$$

We write the second factor as

$$P_r(\Delta N_B(s) = 1 | \mathcal{H}_t^A, \mathcal{H}_{t \vee s}^B) = E[\Delta N_B(s) | \mathcal{H}_t^A, \mathcal{H}_{t \vee s}^B], \tag{0.21}$$

and we have

$$P_r(\Delta N_A(t) = 1, \Delta N_B(s) = 1 | \mathcal{H}_t^A, \mathcal{H}_{t \vee s}^B) = E[V_t^A \Delta N_B(s) | \mathcal{H}_t^A, \mathcal{H}_{t \vee s}^B] + o(\Delta t). \tag{0.22}$$

By taking average over the σ-field $\mathcal{H}_t^A$, we obtain

$$P_{AB}(t, s) = E[V_t^A \frac{dN_B(s)}{ds} | \mathcal{H}_{t \vee s}^B], \tag{0.23}$$

which is, by the proof of Lemma 3.2,

$$P_{AB}(t, s) = e^{h(t,s)} E[V_t^A | \mathcal{H}_t^B] P_B(s) = e^{h(t,s)} P_A(t) P_B(s). \tag{0.24}$$

Proof of Lemma 3.3:

Under Assumptions 3.1 and 3.2, $P_A(t)$ becomes the conditional expectation of the postsynaptic membrane potential, which can be expressed as

$$E[V_t^A|\mathcal{H}_t^B] = E[\alpha \exp\{\sum_{k=1}^{N_B(t)} h(t, T_k^B)\}|\mathcal{H}_t^B] \tag{0.25}$$

Define an event $D_n = \{N_B(t) = n\}$, which has a conditional Poisson distribution

$$P_r(D_n|\mathcal{H}_t^B) = \frac{[\int_0^t V_\tau^B d\tau]^n}{n!} e^{-\int_0^t V_\tau^B d\tau}. \tag{0.26}$$

Because $\{N_B(t) : t \geq 0\}$ is an inhomogeneous Poisson process with the associated point process $\{T_k : k = 1, 2, \cdots\}$ for the given realization of the intensity process $\{V_t^B : t \geq 0\}$, it can be shown that [24]

$$E[\exp\{\sum_{k=1}^{N_B(t)} h(t, T_k^B)\}|N_B(t) = n, \mathcal{H}_t^B] = [\frac{\int_0^t e^{h(t,\tau)} V_\tau^B d\tau}{\int_0^t V_\tau^B d\tau}]^n. \tag{0.27}$$

Therefore, we have

$$E[E[\exp\{\sum_{k=1}^{N_B(t)} h(t, T_k^B)\}|N_B(t) = n, \mathcal{H}_t^B]]$$

$$= \sum_{n=0}^{\infty} E[\exp\{\sum_{k=1}^{N_B(t)} h(t, T_k^B)\}|N_B(t) = n, \mathcal{H}_t^B] P_r(D_n|\mathcal{H}_t^B)$$

$$= \sum_{n=0}^{\infty} \frac{[\int_0^t V_\tau^B d\tau]^n}{n!} e^{-\int_0^t V_\tau^B d\tau}, \tag{0.28}$$

and the summation in the above equation is a series expression for an exponential function. Thus rewriting it proves the lemma.

Proof of Theorem 3.2:

By Lemma 3.1, $P_B(s) = V_s^B$; by Theorem 3.1, $f_A(t, \theta_t^A) = V_t^A$. Since V_t^A and V_s^B are always positive, we write

$$\gamma(t,s) = \frac{E[V_t^A V_s^B]}{E[V_t^A]E[V_s^B]}. \tag{0.29}$$

Note that $E[V_t^A] = E[P_A(t)]$ and $E[V_t^A V_s^B] = E[E[V_t^A|\mathcal{H}_{t\vee s}^B]V_s^B] = E[P_A(t)V_s^B]$. This is because $E[V_t^A|\mathcal{H}_{t\vee s}^B] = P_A(t)$. Hence, applying Lemma 3.3 completes the proof.

Proof of Theorem 3.3:

In the expression (3.48), the numerator of $\gamma(t,s)$ can be written as $E[Xf(s)e^{X\eta_t}]$; the denominator can be written as $E[Xf(s)]E[e^{X\eta_t}]$. Eq. (3.49) is true by canceling $f(s)$. Furthermore, we want to show that as the variance of X decays to zero, $M'(\eta_t) \to \mu e^{\mu\eta_t}$ and $M(\eta_t) \to e^{\mu\eta_t}$, and consequently $\gamma(t,s) \to 1$.

Let $f_\mu(x, \sigma^2)$ be the pdf of X with mean μ and variance σ^2. Define

$$\phi_{[1/\sigma^2]}(x-\mu) = f_\mu(x, \sigma^2) \tag{0.30}$$

where $[x]$ = the largest integer no bigger than x, hence $\{\phi_k(x-\mu) : k = 0, 1, \cdots\}$ is a Dirac sequence with shrinking support [23]. Assume that the moment generating function of X is defined on A as

$$M(t) = \int e^{xt} f_\mu(x, \sigma^2)dx. \tag{0.31}$$

First, we prove that $M(t) \to e^{\mu t}$. Define

$$M_k(t) = \int e^{xt}\phi_k(x-\mu)dx. \tag{0.32}$$

It suffices to show that $\lim_{k\to\infty} M_k(t) = e^{\mu t}$. Since

$$e^{\mu t} = e^{\mu t}\int \phi_k(x-\mu)dx = \int e^{\mu t}\phi_k(x-\mu)dx, \tag{0.33}$$

we have

$$M_k(t) - e^{\mu t} = \int [e^{xt} - e^{\mu t}]\phi_k(x-\mu)dx$$

$$= \int_{|x-\mu|<\delta} + \int_{|x-\mu|\geq\delta} [e^{xt} - e^{\mu t}]\phi_k(x-\mu)dx. \tag{0.34}$$

For any $\epsilon > 0$, there exists δ such that if $|x-\mu| < \epsilon$ then for any $t \in A$ we have $|e^{xt} - e^{\mu t}| < \epsilon$. For k large, the support of ϕ_k is contained in the interval of radius δ centered at μ, whence the integral expressing $M_k(t) - e^{\mu t}$ is concentrated on that interval, and is obviously bounded up by ϵ. This proves that $\lim_{k\to\infty} M_k(t) = e^{\mu t}$.

By the same token, we prove that

$$\lim_{k\to\infty} \int xe^{xt}\phi_k(x-\mu)dx = \mu e^{\mu t} \tag{0.35}$$

which completes the proof.

Proof of Theorem 3.4:

For a given significance level α, we need to find a bound b satisfying

$$P_r(|\frac{H^{AB}_{mn}}{H^A_m H^B_n} - 1| > b \mid H^A_m, H^B_n; \mathcal{H}_0) = \alpha. \tag{0.36}$$

Let us remember that RH_{mn}^{AB} is binomially distributed with parameters $(R, E[P_{AB}(t,s)])$ and that $H_{mn}^{AB} \rightarrow H_m^A H_n^B$ almost surely under $\mathcal{H}_0$. By the central limiting theorem and theorems on limiting distributions,

$$\frac{RH_{mn}^{AB} - RE[H_{mn}^{AB}]}{\sqrt{RH_{mn}^{AB}(1 - H_{mn}^{AB})}} \rightarrow N(0,1) \ \ as \ \ R \rightarrow \infty \tag{0.37}$$

where $N(0,1)$ is denoted as a standard Gaussian random variable. Then if spike trains A and B are uncorrelated, we approximately write

$$\frac{RH_{mn}^{AB} - RE[H_{mn}^{AB}]}{\sqrt{RH_{mn}^{AB}(1 - H_{mn}^{AB})}} \simeq \frac{RH_{mn}^{AB} - RH_m^A H_n^B}{\sqrt{RH_m^A H_n^B(1 - H_m^A H_n^B)}} \tag{0.38}$$

This means that Eq. (0.36) can be approximately written as

$$P_r(|N(0,1)| > \varepsilon_b \mid \mathcal{H}_0) = \alpha \tag{0.39}$$

where

$$\varepsilon_b \simeq \frac{bRH_m^A H_n^B}{\sqrt{RH_m^A H_n^B(1 - H_m^A H_n^B)}} \tag{0.40}$$

which results in an expression of the bound as

$$b \simeq \varepsilon_b \sqrt{\frac{1 - H_m^A H_n^B}{RH_m^A H_n^B}}. \tag{0.41}$$

The value of ε_b is determined by

$$\Phi(\varepsilon_b) = 1 - \frac{\alpha}{2} \tag{0.42}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$.

The above arguments imply that element $H^{AB}_{mn}/(H^A_m H^B_n)$ of the normalized joint PST diagram has a conditional expectation value 1 and an approximate conditional variance

$$\sigma^2_{mn} \simeq \frac{1 - H^A_m H^B_n}{R H^A_m H^B_n} \tag{0.43}$$

under hypothesis $\mathcal{H}_0$.

Proof of Theorem 3.5:

Let us note that under hypothesis $\mathcal{H}_0$, $H^{AB}_{mn}/(H^A_m H^B_n)$ is approximately Gaussian distributed with mean 1 and variance σ^2_{mn}. Hence

$$|G_k - 1| \simeq |\frac{1}{N - |k|} \sum_{n=\max(1,1-k)}^{\min(N,N-k)} \sigma_{n+k,n}\ N_n(0,1)| \tag{0.44}$$

where each $N_n(0,1)$ approximately has a standard Gaussian distribution expressed by

$$N_n(0,1) = \frac{R H^{AB}_{n+k,n} - R H^A_{n+k} H^B_n}{\sqrt{R H^A_{n+k} H^B_n (1 - H^A_{n+k} H^B_n)}}, \tag{0.45}$$

and σ^2_{mn} is given in Eq. (3.78). Therefore, $G_k - 1$ is approximately Gaussian distributed with zero-mean and variance

$$Var(G_k - 1) = \frac{1}{(N - |k|)^2} \sum_{n=\max(1,1-k)}^{\min(N,N-k)} \sigma^2_{n+k,n} \tag{0.46}$$

where mutual independence of $N_n(0,1)$ is assumed. Let

$$\varepsilon_b = \frac{b_k}{\sqrt{Var(G_k - 1)}}, \tag{0.47}$$

we obtain

$$P_r(|G_k - 1| > b_k \mid H_m^A, H_n^B; \mathcal{H}_0) = 2(1 - \Phi(\varepsilon_b)). \tag{0.48}$$

If all σ_{mn}^2's are the same, observing the bound b in Theorem 3.4 completes the proof.

BIBLIOGRAPHY

[1] M. Abeles and M. H. Goldstein, "Multispike train analysis," *Proc. IEEE*, vol. 65, pp. 762–772, 1977.

[2] N. Ahmed, T. Natarajan, and K. R. Rao , "Some considerations of the modified Walsh-Hadamard and Haar transforms," *Proc. 1973 Symp. Applications of Walsh Functions*, pp. 91–95.

[3] H. C. Andrews and K. L. Caspari, "A generalized technique for spectral analysis," *IEEE Trans. Computers*, vol. C-19, pp. 16–25, 1970.

[4] J. Bak and E. M. Schmidt, "An analog delay circuit for on-line visual confirmation of discriminated neuroelectric signals," *IEEE Trans. Biomed. Eng.*, vol. BME-24, pp. 69–71, 1977.

[5] J. Bak and E. M. Schmidt, "An improved time-amplitude window discriminator," *IEEE Trans. Biomed. Eng.*, vol. BME-24, pp. 486–489, 1977.

[6] P. B. Bishop, W. R. Levick and W. O. Williams, "Statistical analysis of the dark discharge of lateral geniculate neurones," *J. Physiol.*, vol. 170, pp. 598–612, 1964.

[7] H. Van den Boogaard, G. Hesselmans and P. Johannesma, "System identification based on point processes and correlation densities. I. The nonrefractory neuron model," *Math. Biosci.* vol. 80, pp. 143–171, 1986.

[8] M. J. Correia and J. P. Landolt, "A point process analysis of the spontaneous activity of anterior semicircular canal units in the anesthetized pigeon," *Biol. Cybern.*, vol. 27, pp. 199–213, 1977.

[9] J. C. Dill, P. C. Lockeman and K. Naka, "An attempt to analyze multiunit recordings," *Electroenceph. Clin. Neurophysiol.*, vol. 28, pp. 79–82, 1970.

[10] G. J. Dinning and A. C. Sanderson, "Real-time classification of multiunit neural signals using reduced feature sets," *IEEE Trans. Biomed. Eng.*, vol. BME-28, pp. 804–812, 1981.

[11] G. L. Gerstein, "Functional association of neurons: Detection and interpretation," pp. 648–661 in *The Neurosciences. Second Study Program*, F. O. Schmitt(ed). Rockefeller Univ. Press, New York, 1970.

[12] G. L. Gerstein and W. A. Clark, "Simultaneous studies of firing patterns in several neurons," *Science*, vol. 143, pp. 1325–1327, 1964.

[13] G. L. Gerstein and D. H. Perkel, "Simultanously recorded trains of action potentials: Analysis and functional interpretation," *Science (Washington)*, vol. 148, pp. 828–830, 1969.

[14] S. Grossberg, "Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround neural networks," *Biol. Cybern.*, vol. 20, pp. 69–98, 1975.

[15] S. Grossberg and E. Mingolla, "Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading," *Psych. Rev.*, vol. 92, pp. 173–211, 1985.

[16] A. Haar, "Zur theorie der orthogonalen funktionensysteme," *Math. Ann.*, vol. 69(1910), pp. 334–371, 1910; vol. 71(1912), pp. 38–53, 1912.

[17] M. K. Habib, P. K. Sen, "Non-stationary stochastic point-process models in neurophysiology with applications to learning," *Biostatistics: statistics in biomedical, public health and environmental sciences*, P. K. Sen (ed), Elsevier/North-Holland, Amsterdam, pp. 481–509, 1985.

[18] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, vol. 79, pp. 2554–2558, 1982.

[19] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci. USA*, vol. 81, pp. 3088–3092, 1984.

[20] C. K. Knox, "Cross-correlation functions for a neuronal model," *Biophys. J.*, vol. 14, pp. 567–582, 1974.

[21] T. Kohonen, *Self-Organization and Associative Memory*, Berlin: Springer-Verlag, 1984.

[22] J. Krüger, "Simultaneous individual recordings from many cerebral neurons: techniques and results," *Rev. Physiol. Biochem. Pharmacol.*, vol. 98, pp. 177–233, 1983.

[23] S. Lang, *Real Analysis*, 2nd ed., Addison-Wesley, London, 1983, p. 358.

[24] H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences (Vol. II) Random Noise, Signals, and Dynamic Systems*, John Wiley, 1979, p. 591.

[25] G. D. McCann, "Interactive computer strategies for living nervous system research," *IEEE Trans. Biomed. Eng.*, vol. BME-20, pp. 1–11, 1973.

[26] D. J. Mishelevich, "On-line real-time digital computer separation of extracellular neuroelectric signals," *IEEE Trans. Biomed. Eng.*, vol. BME-17, pp. 147–150, 1970.

[27] H. Nakahama, H. Suzuki, M. Yamamoto, S. Aikawa and S. Nishioka, "A statistical analysis of spontaneous activity of central single neurons," *Physiol. Behav.*, vol. 3, pp. 745–752, 1968.

[28] R. O'Connell, W. A. Kocsis, and R. L. Schoenfeld, "Minicomputer identification and timing of nerve impulses mixed in a single recording channel," *Proc. IEEE*, vol. 61, pp. 1615–1621, 1973.

[29] M. N. Oguztoreli and R. B. Stein, "Optimal filtering of nerve signals," *Biol. Cybern.*, vol. 27, pp. 41–48, 1977.

[30] G. Palm, A. M. H. J. Aertsen and G. L. Gerstein, "On the significance of correlations among neuronal spike trains," Biol. Cybern., vol. 59, pp. 1–11, 1988.

[31] D. H. Perkel, G. L. Gerstein and G. P. Moore, "Neuronal spike trains and stochastic point processes, I. The single spike train," *Biophys. J.*, vol. 7, pp. 391–418, 1967.

[32] D. H. Perkel, G. L. Gerstein and G. P. Moore, "Neuronal spike trains and stochastic point processes, II. Simultaneous spike train," *Biophys. J.*, vol. 7, pp. 419–440, 1967.

[33] V. J. Prochazka and H. H. Kornhuber, "On-line multi-unit sorting with resolution of superposition potentials," *Electroenceph. Clin. Neurophysiol.*, vol. 34, pp. 91–93, 1973.

[34] W. M. Roberts, "Optimal recognition of neuronal waveforms," *Biol. Cybern.*, vol. 35, pp. 73–80, 1979.

[35] D. L. Snyder, *Random Point Processes*, John Wiley, New York, 1975.

[36] M. D. Srinath and P. K. Rajasekaran, *An Introduction to Statistical Signal Processing with Applications*, John Wiley & Sons, New York, 1979, p. 79–83.

[37] D. Stagg, "Computer acquisition of multiunit nerve-spike signals," *Med. Biol. Eng.*, vol. 11, pp. 340–347, 1973.

[38] R. B. Stein, S. Andreassen and M. N. Oguztoreli, "Mathematical analysis of optimal multichannel filtering for nerve signals," *Biol. Cybern.*, vol. 32, pp. 19–24, 1979.

[39] R. B. Stein, S. Andreassen and M. N. Oguztorelia, "Application of optimal multichannel filtering to simulated nerve signals," *Biol. Cybern.*, vol. 32, pp. 25–33, 1979.

[40] I. H. M. van Stokkum, P. I. M. Johannesma and J. J. Eggermont, "Representation of time-dependent correlation and recurrence time functions," *Biol. Cybern.*, vol. 55, pp. 17–24, 1986.

[41] D. W. Tank and J. J. Hopfield, "Collective computation in neuronlike circuits," *Scientific American*, pp. 104–114, Dec. 1987.

[42] X. Yang, S. A. Shamma, "A totally automated system for the detection and classification of neural spikes," *IEEE Trans. Biomed. Eng.*, vol. BME-35, pp. 806–816, 1988.