Preface

Theoretical analysis and computational modeling are important tools for characterizing what nervous systems do and determining how and why they do it. Neuroscience encompasses approaches that range from molecular studies to human psychophysics. Theoretical neuroscience encourages cross-talk between the many layers of this broad field by constructing compact representations of what has been learned at various levels, building bridges between these different descriptions, and identifying potential unifying concepts and principles. In this book, we present the basic methods used for these purposes and discuss examples in which theoretical approaches have yielded insight into nervous system function.

The questions what, how, and why are addressed by the three types of models discussed in this book: descriptive, mechanistic, and interpretive. Descriptive models are used to summarize large amounts of experimental data compactly yet accurately. These models may be based loosely on biophysical, anatomical, and physiological results, but their primary purpose is to describe a phenomenon, not to account for it. Noteworthy examples of this type of modeling are found in chapter 2, where we discuss mathematical descriptions of how neurons in the retina, lateral geniculate nucleus of the thalamus, and primary visual cortex respond to the spatiotemporal patterns of visual images.

Mechanistic models account for nervous system activity on the basis of neuronal morphology, physiology, and circuitry. Such models often form a bridge between descriptive models acting at different levels. For example, in chapter 7, descriptive models of the neurons, synapses, and circuitry of primary visual cortex are used to build mechanistic models of the neural responses discussed descriptively in chapter 2. Mechanistic models are part of a general reductionist approach that uses knowledge of the underlying biophysical processes to explain neural responses.

Interpretive models explore the functional roles of neural circuits. For example, in chapter 4 we present interpretive models of the visual responses analyzed descriptively and mechanistically in chapters 2 and 7. These models account for the response properties of neurons early in the mammalian visual pathway by showing that they maximize the information conveyed about natural visual scenes subject to noise and other relevant constraints. In general, interpretive models relate neuronal response properties to basic computational principles.

Theoretical neuroscience attempts to deal with issues ranging from the dynamics of single ion channels to cognition. A great variety of different modeling approaches are needed to deal with such a wide range of phenomena. People contemplating neural modeling often have difficulty identifying the appropriate level of model to apply to their particular problem. A frequent mistake is to assume that a more detailed model is a superior one. Because models act as bridges between levels of understanding, they must be detailed enough to make contact with the lower level yet simple enough to yield clear results at the higher level. Like a bridge, a model only needs to reach between the levels being spanned. A far-ranging understanding of neural systems is more likely to be achieved by constructing many short bridges than a single long one.

Theoretical neuroscience is obviously based on the belief that mathematics can elucidate nervous system function. Unfortunately, mathematics can sometimes seem more an obstacle than an aid to understanding. We have not hesitated to employ the level of mathematics needed to be precise and rigorous. At times, this may stretch the patience of some of our readers. We encourage such readers to consult the mathematical appendix, but also to persevere and attempt to understand the implications and consequences of a difficult derivation, even if its steps are unclear. In other words, even if the mathematics at some places is too obscure, the associated sections should not be abandoned. A list of mathematical texts containing relevant material appears at the end of this preface.

While some introductory material is included in this book, it is by no means a neuroscience textbook. Readers needing additional background material should consult standard neuroscience texts such as those listed at the end of the preface.

Each chapter contains two reference lists. The first offers suggestions for further reading. The second provides information about works cited within the chapter and related studies to which reference is not directly made. In order to maintain the flow of the text, we have kept citations within the chapters to a bare minimum. In fact, we have undoubtedly dipped below the level needed to keep our friends and colleagues happy. We apologize for all the excellent work that we have failed to cite. A list of general texts on computational and theoretical neuroscience is given at the end of the preface.

This book is organized around general principles, not around specific systems such as the hippocampus or cerebellum, which have been the targets of substantial modeling. Deciding what to leave out has been a very difficult aspect of writing this book, and we again apologize for our omissions.

Theoretical neuroscience, like any skill, can only be mastered with practice. We have provided exercises for this purpose and urge the reader to do them. In addition, it will be highly instructive for the reader to construct the models discussed in the text and explore their properties beyond what we have been able to do in the available space.

This book has three parts. Part I covers how information about stimuli and responses is encoded in, and can be decoded from, the temporal patterns of action potentials produced by one or more neurons. Part II presents biophysical models of neurons, synapses, and feedforward and recurrently coupled networks. Part III discusses the roles of adaptation and synaptic plasticity in forming and modifying neuronal responses.

Chapter 1 analyzes how neurons encode information by firing sequences of action potentials. It covers methods for interpreting spike trains and understanding their relationship to the stimuli or responses they encode, statistical models of spiking, and issues such as the role of correlations in neural coding. A range of mathematical techniques, including Fourier transforms and Poisson processes, is introduced and employed.

Many of the methods discussed in chapter 1 are applied to early stages of the visual system in **Chapter 2**. The descriptive models presented summarize the properties of cells in the retina, thalamus, and primary visual cortex.

Chapter 3 complements the encoding discussion of chapter 1 by considering ways that information can be decoded from sequences of action potentials and from the activities of populations of neurons. Statistical ideas such as receiver-operator characteristic curves and the Fisher information are introduced.

Chapter 4 treats the application of Shannon information theory to neural encoding. Optimization of the entropy and mutual information are applied to the encoding of natural stimuli by single neurons and neural populations, and the trade-off between signal and noise is discussed.

Chapters 5 and **6** present biophysical models of neurons and synapses. These include the Hodgkin-Huxley model for the generation and propagation of action potentials, descriptions of neurotransmitter- and voltage-gated channels and conductances, and analytical and numerical computational methods for calculating the membrane potentials of neurons with complex morphologies. Reduced descriptions are developed, including integrate-and-fire and firing-rate models. The chapter introduces and uses phase-plane methods and stability analysis applied to nonlinear differential equations.

Chapter 7 deals with models of networks of neurons. Starting from a general description of coupled excitatory and inhibitory neurons, it applies feedforward models to the computation of coordinate transformations in parietal cortex. Recurrent network models are then developed, and a variety of computations that they can carried out, including short- and long-term memory and statistical inference, are discussed. The chapter also considers oscillatory networks, using a model of the olfactory bulb as an example.

Chapter 8 discusses models of synaptic plasticity and their relationship to statistical methods such as principal component analysis. Hebbian models of activity-dependent development are introduced. The chapter also treats the delta rule, a model for plasticity that is frequently used for supervised learning. The consequences of Hebbian learning are analyzed by computing eigenvectors and eigenvalues of covariance matrices calculated from input statistics.

Chapter 9 presents an analysis of selectivity maps, in particular, the regular ways that selectivities to orientation, location, and ocularity of visual inputs are mapped across primary visual cortex. Optical imaging experiments, together with manipulations such as monocular deprivation, provide a wealth of data about the nature and plasticity of these maps, which have been a fruitful target for modeling.

Chapter 10 covers reinforcement learning, the ways that animals learn to predict events of importance in their environment, such as the delivery of reward and punishment, and use those predictions to choose appropriate actions. Neurons that release the neurotransmitter dopamine have been implicated in reward learning, and their behavior is modeled. The chapter also develops the computational context for reinforcement learning, which is an engineering technique for optimal control called dynamic programming.

Chapter 11 is concerned with representational learning, the way that input activities come to be recoded as a result of experience. The main idea is to represent inputs according to their possible causes in the world. The methods discussed in the chapter determine causes by performing versions of the statistical technique of density estimation.

The **Appendix** provides reviews of some of the mathematical methods used in the book.

We are very grateful to the large numbers of colleagues and students at Brandeis, GCNU and MIT who have painstakingly read, commented on, and criticised, a host of versions of all the chapters. We particularly thank Zhaoping Li, Mark Kvale, and Mark Goldman for being our most consistent critics and Karen Abbott for help with figures and references. In addition, we are grateful to Matt Beal, Bill Bialek, Emery Brown, Frances Chance, Pat Churchland, Yang Dan, Bard Ermentrout, Ed Erwin, David Foster, Marcus Frean, Ralph Freeman, Fabrizio Gabbiani, Zoubin Ghahramani, Federico Girosi, Geoff Goodhill, David Heeger, Geoff Hinton, Mike Jordan, Sham Kakade, Szabolcs Káli, Christof Koch, Eve Marder, Guy Mayraz, Ken Miller, Read Montague, Quaid Morris, Sacha Nelson, Klaus Obermayer, Bruno Olshausen, Randy O'Reilly, Mark Plumbley, Tomaso Poggio, Alex Pouget, Max Riesenhuber, Emilio Salinas, Terry Sejnowski, Sebastian Seung, Mike Shadlen, Fiona Stevens, Nick Swindale, Gina Turrigiano, Carl van Vreeswijk, Martin Wainwright, Xiao-Jing Wang, Chris Williams, David Willshaw, Charlie Wilson, Danny Young and Rich Zemel. We also thank Michael Rutter at MIT Press for his consistent commitment and studied patience.

Suggested Background Reading

Computational Neuroscience

Churchland, PS & Sejnowski, TJ (1992) *The Computational Brain*. Cambridge, MA:MIT Press.

Koch, C (1998) *Biophysics of Computation: Information Processing in Single Neurons*. New York:Oxford University Press.

Koch, C & Segev, I, editors (1998) *Methods in Neuronal Modeling: From Synapses to Networks*. Cambridge, MA:MIT Press.

Rieke, FM, Warland, D, de Ruyter van Steveninck, R & Bialek, W (1997) *Spikes: Exploring the Neural Code*. Cambridge, MA:MIT Press.

Tuckwell, HC (1988) *Introduction to Theoretical Neurobiology*. Cambridge, UK:Cambridge University Press.

Neuroscience Texts

Bear, MF, Connors, BW & Paradiso, MA (1996) *Neuroscience: Exploring the Brain*. Baltimore:Williams and Wilkins.

Hille, B (1992) *Ionic Channels of Excitable Membranes*. Sunderland, MA:Sinauer.

Johnston, D & Wu, SM (1995) *Foundations of Cellular Neurophysiology*. Cambridge, MA:MIT Press.

Kandel, ER, Schwartz, JH & Jessel, TM, editors (1991) *Principles of Neural Science*. New York:Elsevier.

Levitan, IB & Kaczmarek, LK (1997) *The Neuron: Cell and Molecular Biology*. New York:Oxford University Press.

Nicholiis, JG, Martin, AR & Wallace, BG (1992) *From Neuron to Brain: A Cellular and Molecular Approach to the Function of the Nervous System*. Sunderland MA:Sinauer.

Shepherd, GM, editor (1998) *The Synaptic Organization of the Brain*. New York:Oxford University Press.

Zigmond, MJ, Bloom, FE, Landis, SC, Roberts JL & Squire LR, editors (1999) *Fundamental Neuroscience*. San Diego:Academic Press.

Mathematics Texts

Boas, ML (1966) *Mathematical Methods in the Physical Sciences.* New York, NY:Wiley.

Cover, TM & Thomas, JA (1991) *Elements of Information Theory.* New York, NY:Wiley.

Feller, W (1968) *An Introduction to Probability Theory and its Application, 3rd edition.* New York, NT:Wiley.

Gershenfeld, N (1999) *The Nature of Mathematical Modeling.* Cambridge, UK:CUP.

Jordan, DW & Smith, P (1977) *Nonlinear Ordinary Differential Equations.* Oxford, England:Clarendon Press.

Lindgren, BW (1993) *Statistical Theory, 4th edition.* New York, NY:Chapman & Hall.

Oppenheim, AV & Willsky, AS with Nawab, H (1997) *Signals and Systems, 2nd edition.* Upper Saddle River, NJ:Prentice Hall.

Press, WH, Teukolsky, SA, Vetterling, WT & Flannery, BP (1992) *Numerical Recipes in C*. Cambridge, UK:Cambridge University Press.

Strang, G (1976) *Linear Algebra and Its Applications.* New York, NY:Academic Press.