Explorations in Neurally Inspired Computing

by

Anthony Montemayor

A Thesis submitted to the Faculty

in partial fulfillment of the requirements for the

BACHELOR OF ARTS

Accepted

Paul Shields, Thesis Advisor

William Dunbar, Second Reader

Mary B. Marcy, Provost and Vice President

Simon's Rock College of Bard
Great Barrington, Massachusetts

2007-2008

Acknowledgments

This thesis is dedicated to:

My thesis advisor, Paul Shields -- your words of sympathy and encouragement meant more to me than you know.

My faculty advisor, Bill Dunbar -- I know I didn't make it easy, but your patience with my procrastination over the years leaves me with an extreme debt of gratitude that I probably can't ever really repay.

David LaBerge, my source of inspiration -- your classes and conversations set me on the path of neuroscience. Your enthusiasm and humor were infective, and made every encounter a joy.

My friends and peers, Ian Peters, Kyle Chao, Mindy Isser, Monty Hill, Sam Wilensky, Stephanie Barefield, Umar Alizai, Vincent Rozalski, and many others -- your support and gentle prodding kept me going when things got hard. You were what made the stress bearable and the time fly.

My family, who had the patience and trust to allow me to come to this hectic place so far from home, and whose love offered me the stability to pursue my ambitions.

Table of Contents

Introduction Page 1

Chapter 1: A First Approach

The Biology Page 3

The Computing Page 5

Chapter 2: A Second Approach

The Biology Page 15

The Computing Page 20

Conclusions Page 28

Bibliography Page 32

Abstract

The field of artificial intelligence strives to automate problem solving usually requiring human input. For decades many researchers have tried to do so by looking at models of neural activity. I review some major efforts and results in the modeling of neural processes; from simple single and multi-layer feed-forward networks to biologically realistic conductance based models. Working from basic biological ideas I systematically motivate each model from the simple to the complex. I present a geometrical interpretation of feed-forward network activity, and show that any functional mapping can be recreated. I then review methods to interpret qualitative behavior in the more complex conductance based models by use of dynamical systems theory. I offer results illustrating the computational power and ability to faithfully recreate neural outputs, and discuss some ramifications and applications to experimental modeling and problem solving.

Introduction

Humankind is blessed with a powerful trait that allows them to interact with the world above and beyond most other species of this earth. Our intelligence is a tool that has allowed us to achieve dominion over the globe. This trait pushes us to always look for better and easier solutions to the problems facing us. With the dawn of the age of the computer it is natural to look for ways to automate the problem-solving usually requiring our input, but how limited is the potential for computer automation? Can computers obtain a level of intelligence equal to or greater than ours? The field of computer science called artificial intelligence attempts to answer these questions.

Artificial intelligence has several varied and vague definitions. (see Russell and Norvig for a summary) What these definitions hold in common is the description of machine intelligence as capable of what we deem intelligent or rational from the standard of human behavior. One natural approach to such a feat is to try to create models of the way our minds work, and attempt to abstract key features into machine architecture. But without the tools to completely map or even measure every biological variable at work in our brains and bodies, we must be content with simplifications instead of full-fledged emulations. As with any model we seek to capture what we think is important for intelligence and throw out the rest. Whichever model we choose must then be tested. What are its capabilities and limitations? Does it capture, successfully, the aspects of

intelligence we have sought to recreate? Is the behavior of the model plausible within the framework of our notion of intelligent/rational behavior?

Beyond specific biological or time constraints even the simplest models created following this approach of abstracted brain modeling can be shown to be incredibly powerful. It is this power that has time and again brought tremendous interest to the field of artificial neural networks (i.e., computer models of neural activity in the brain). If one can tame the power of the dynamic information handling at use in the brain, many problems that were thought beyond automation are made quite tractable. When a machine can learn and apply any informational relation in data sets, one needs to ask, "Isn't this intelligent behavior?"

I will not answer this question for the reader, but will merely introduce some of the results that clearly illustrate how even the simplest models of neural activity implemented in machines allows the automation of a robust and complex set of abilities. With this toolset established, I will then show the reader that more specific models can capture the versatility of real neurons. If we can create tractable and faithful models of the information processing at work in our minds, is it really so far-fetched that machines can recreate the intelligent behavior that humans and other species exhibit?

Chapter One

A First Approach

The Biology

To model human intelligence one must first develop an understanding of its origins. Disregarding dualistic views, the seat of human intelligence is the brain. Then, taking a reductionist approach, the behavior of the brain should be able to be explained by the behavior of the sum of its parts. What are the parts of the brain, and how do they process information?

The nervous system is made up of two broad classes of cells: neurons and glia. Glia are classically treated as support cells. They are responsible for the blood brain barrier, a safety measure for protecting the brain from chemicals in the blood stream, for controlling blood supply and oxygenation of the brain's cells, for repairing cellular damage, and electrical insulation of a neuron's axon. (Volterra and Meldolesi; Fields) More recently they have been seen to actually affect neural activity itself via chemical instead of electrical activity. (Kozlov et al.) Because of this difference in signal type, their contribution to neural processing wasn't able to be resolved until recent advances in imaging techniques. (see Koizumi et al.; Poznanski and Riera; Simard et al.; Volterra and Meldolesi)

Neurons, the other type of cell, are the ones at which most modeling attempts are aimed. A typical neuron is made up of several dendrites, a cell body (soma), and an axon (see Figure 1). Neurons are connected to one another usually by chemical synapses and rarely by direct contact of their membranes. Electrical activity in a presynaptic neuron makes its synapse release chemicals that affect the postsynaptic neuron's electrical behavior.

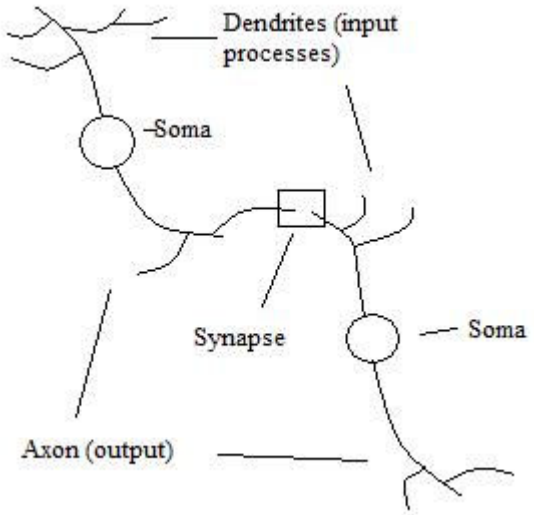


Figure 1

Electrical activity from the spike of a presynaptic neuron transfers via the synapse usually to the dendrites of the postsynaptic neuron. These neurons are only diagrammatic. Morphologically speaking, such neurons are not the majority in the brain's cortex.

A neuron's activity is its electrical state. If a neuron receives enough stimuli from other neurons via its dendrites it will reach its threshold. This threshold is the electrical potential necessary to trigger openings along its membrane to activate and allow ion flow across the membrane. This flow creates a fast change of the cellular potential called a spike. This spike propagates down the axon and can trigger the synapses to release their chemical transmitters to affect other neurons. The nature of the response of the synapses and the postsynaptic neuron can be altered by cellular mechanisms responding to specific

chemical and electrical context. This alteration, in essence, weights the signal from the presynaptic neuron to have a stronger or weaker effect on the postsynaptic neuron.

This process of neurotransmission is typically believed to be the primary mode of information processing in the brain. A neuron receives information from other neurons and processes those signals via the nature of its electrochemical response and threshold requirements. Then it propagates the processed information to other neurons through its axon by spiking activity.

The Computing

This typical activity of neurons leads quite naturally to a computational/empirical model. A "neuron" will take a weighted sum of its inputs and compare it to a threshold. If met, the neuron will send its signal to other neurons.

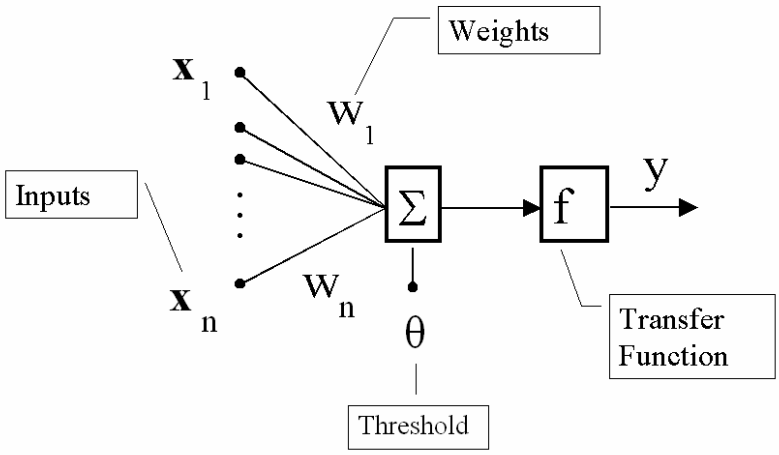


Figure 2
Diagram of an artificial neuron.

Inputs from other "neurons" or from external sources are weighted and summed and compared to the threshold, theta, which can be viewed as a constant input $w_0 * x_0$ where the x_0 line is always on. Then, because the soma/axon can initiate a spike based on the internal neurobiology, we will allow the artificial neuron the ability to transform the weighted sum. If one views a spike as a binary event then one example of such a neuron would be the following.

$$f(w,x)=\mathrm{sgn}(w_0+w_1x_1+\ldots+w_nx_n)$$

$$\mathrm{sgn}(z)=\{1 \text{ if } z \geq 0 \text{ , } 0 \text{ if } z < 0\}$$

$$w_0+w_1x_1+\ldots+w_nx_n \geq 0$$

$$w_1x_1+\ldots+w_nx_n \geq \theta$$

To simplify notation I will use $\mathbf{w} = \langle w_0, w_1, \ldots, w_n \rangle'$ and $\mathbf{x} = \langle x_0, x_1, \ldots, x_n \rangle'$ where the primes denote transposition.

This model can be interpreted geometrically. The weight vectors define a hyperplane, and if the weighted sum lies above or on this plane the neuron spikes (outputs 1). If below the plane, the neuron outputs 0. An example for two-dimensional input is shown in figure 3. The $w_1, w_2, \ldots, w_n$ orient the plane, and the w_0 determines the offset from the origin.

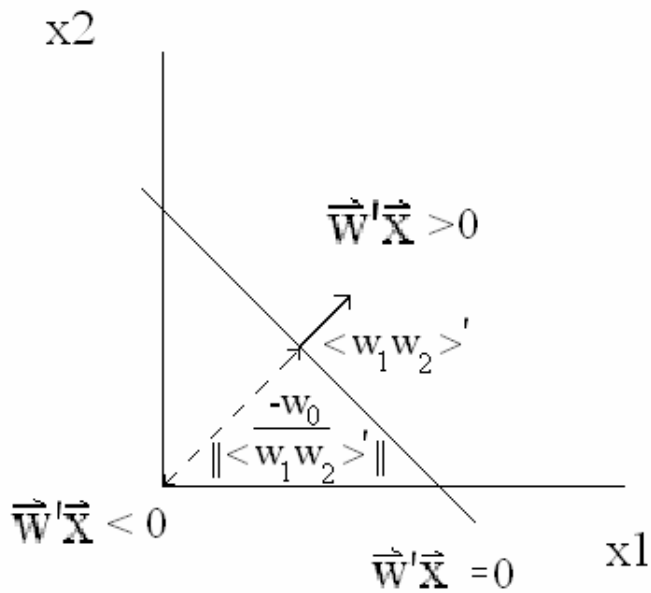


Figure 3
Decision boundary for two-dimensional input. The prime represents transposition of the vector. (adapted from Bishop)

One can extend this idea into a single-layer network. Multiple neurons will receive the same inputs, and the output will be based on the neuron that most surpasses its threshold. This leads to a simply connected and convex decision region. If two input vectors, $\mathbf{x}_a$ and $\mathbf{x}_b$, both lie in the region that most activates the kth neuron then we know that $\mathbf{w}_k\mathbf{'x} > \mathbf{w}_j\mathbf{'x}$ for any other neuron for those vectors. Parameterizing the line segment between those vectors we have

$t*\mathbf{x}_a + (1-t)*\mathbf{x}_b$, $0 \leq t \leq 1$

$t*\mathbf{w}_k\mathbf{'x}_a + (1-t)*\mathbf{w}_k\mathbf{'x}_b$ is then the activation for the kth neuron.

$\mathbf{w}_k\mathbf{'x}_a > \mathbf{w}_j\mathbf{'x}_a$ and $\mathbf{w}_k\mathbf{'x}_b > \mathbf{w}_j\mathbf{'x}_b$ for all $j \neq k$, by assumption.

And so $t*\mathbf{w}_k\mathbf{'x}_a + (1-t)*\mathbf{w}_k\mathbf{'x}_b > t*\mathbf{w}_j\mathbf{'x}_a + (1-t)*\mathbf{w}_j\mathbf{'x}_b$ for all $j \neq k$

So any point on the connecting line also lies in the region that most activates the kth neuron. This shows simple-connectedness and convexity. (Bishop)

So, the network can be interpreted geometrically as cutting up the input space into classes, and depending on the form of the transfer function it can be fairly limited in the problems it can correctly classify. From the argumentation above these classes are separated by linear boundaries in two dimensions, or hyperplanar boundaries in higher dimensions. These classes must also be simply-connected and convex. Such a limitation prevents a single layer network from solving the classic exclusive-or, XOR, problem. This simple Boolean function isn't linearly separable, and so cannot be correctly classified by our single-layer network.

x	y	xor
1	1	0
1	0	1
0	1	1
0	0	0

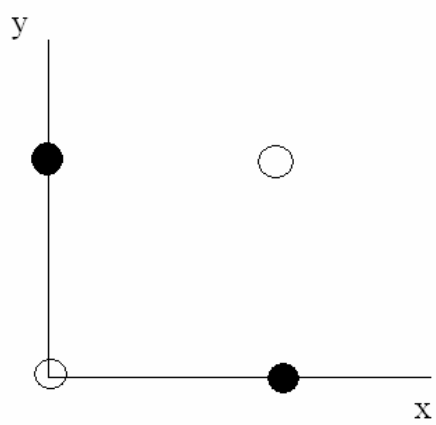


Figure 4
The Boolean function XOR. No linear decision boundary can successfully separate the two classes of points.

To get around this limitation, we need to adjust our model. Something that we overlooked when we were abstracting from biology of real neurons is that neurons might operate on inputs even before the sum over inputs is taken. When the input signal from

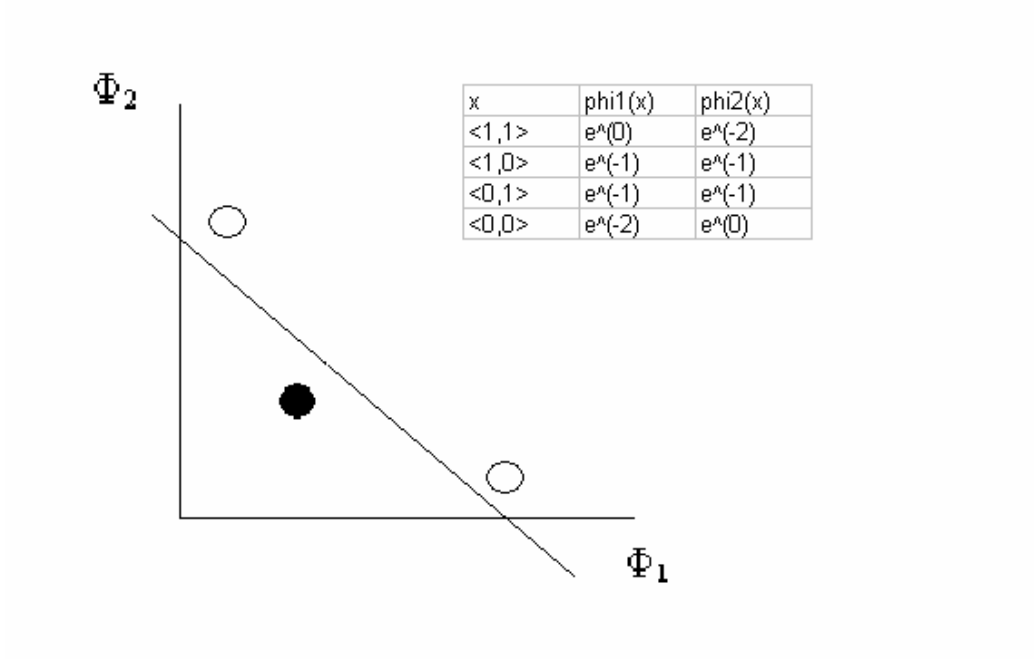
other neurons first interfaces via the synapse, it can be transformed by some function. So instead of needing linear separability in the original input space, we only need linear separability in the transformed space.

Consider the functions:

$$\Phi_1(\mathbf{x}) = \exp\{ - \|\mathbf{x} - \mathbf{t}_1\|^2\},\ \mathbf{t}_1 = <1,1>'$$

$$\Phi_2(\mathbf{x}) = \exp\{ - \|\mathbf{x} - \mathbf{t}_2\|^2\},\ \mathbf{t}_2 = <0,0>'$$

Now if we take the input vectors from the XOR problem, and transform them with Φ_1 and Φ_2, they are linearly separable (see Figure 5).



x	phi1(x)	phi2(x)
<1,1>	e^(0)	e^(-2)
<1,0>	e^(-1)	e^(-1)
<0,1>	e^(-1)	e^(-1)
<0,0>	e^(-2)	e^(0)

Figure 5
XOR is linearly separable in a transformed input space.

This idea changes our model. The output from the kth neuron becomes $f(\mathbf{w}_k' \ \mathbf{\Phi}(\mathbf{x}))$ instead of $f(\mathbf{w}_k'\mathbf{x})$. The problem now becomes how we choose these transformation functions. No matter how we choose, if they are a fixed set, there will be

problems that can't be solved. This limitation was illuminated by Minsky and Papert in 1969 with their book *Perceptrons*. Because of their argumentation and failed claims of network abilities in attempted application much of neural network research halted for years. (Olazaran) If the **Φ** functions are fixed, then the number of them to successfully complete a problem increases, sometimes exponentially, with the dimension of the problem. (Bishop) This isn't computationally acceptable, and so we must allow the **Φ** to be adaptive. Another way to achieve this is to allow multiple layers of adaptive weights.

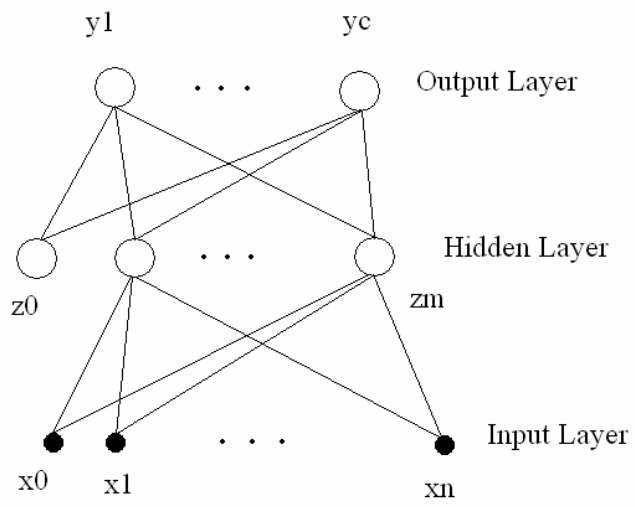


Figure 6
Multiple layer network (adapted from Bishop)

With multiple layers we add tremendous functionality. With binary inputs and outputs one can show that *any* Boolean function, including XOR can be modeled. Let's show XOR using a 2-layer network.

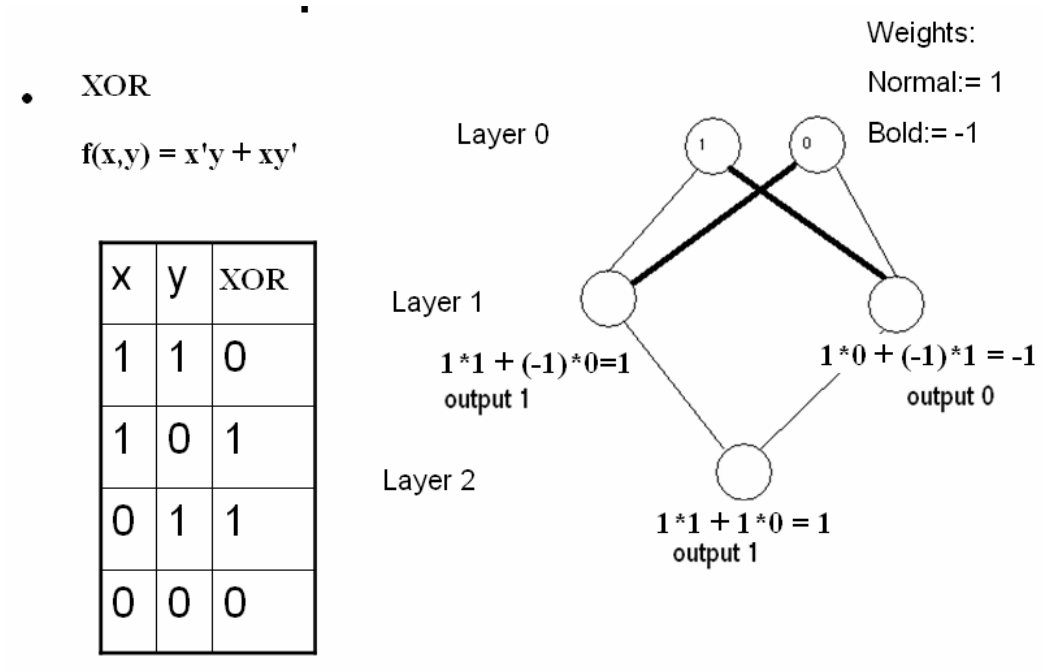


Figure 7
XOR in disjunctive normal form with table (Left), the corresponding network (Right)

Given that any Boolean function can be expressed in what is called disjunctive normal form (an "or" of "ands"), constructing a corresponding network is relatively easy. As shown in Figure 7, for instance, XOR is the same as the disjunctive normal form xy′ + x′y. You construct one node for each "and"-group in the hidden layer with the weights determined by whether each term in the product is negated or not. If negated you have a weight of negative one, and if in the positive you have a weight of a positive one. With this scheme you check for any possible satisfaction of the Boolean function. If any one of the hidden units says that its term checks then the final output layer is passed a one that puts it over its threshold of ½.

This establishes that a 2-layer neural network can perform any computation that a binary computer can perform. Although a network can be shown to exist for any Boolean function, it is computationally impractical to have a neural network set up for any

conceivable Boolean function. The benefit of neural networks is in their adaptive weights. One should be able to train a network from a limited learning set into an optimal setup to compute these functions. As they stand in our current model this condition isn't met. (see Bishop for learning algorithms and further discussion) Despite this lack in efficiency, that these compact networks can theoretically simulate anything a computer can do is cause for closer study at the least.

What if our network wasn't limited to discrete or binary input and output? What are the capabilities of a 2-layer network? As I have mentioned, a single-layer network can be interpreted as splitting up the input space by a hyperplane. If one combines the ideas from this cutting with hyperplanes and the Boolean function computation, one can easily envision the types of construction you can make with a continuous state 2-layer neural network. At the least, one can cut off any convex region of input space with boundaries formed by hyperplanes. You essentially use the hidden layer to form all the hyperplane decision boundaries and then use the next adaptive layer to form logical "ands" or "ors". If you use an "and" construction you isolate where all the hyperplanes intersect to yield a convex region.

One can create more exotic decision regions, but you still don't have full generality of modeling functional mappings. To do this you need yet another adaptive layer of weights. The first two layers construct hyperplane divisions in the form of small cubes. Then the last layer can do a logical "or" of these cubes, and yield a classification

decision. If the cubes are small enough any decision boundary can be formed. (See Bishop for a more precise constructive argument)

Again as in the Boolean function case this construction is far from optimal. In order for the cubes to be small enough to be of much worth the networks would have to be enormous. And besides its size, the boundary setup would have to be fixed in advance. (Bishop) In order to get around this problem one has to study learning algorithms that go into changing the weights/transformation functions of these networks. This problem is beyond the scope of this paper.

Artificial neural networks have the potential to solve a great many problems in artificial intelligence research. With the ability to model any mapping, they can automate any complex process of information processing. The only problem is how to optimize the setup of the network to handle the specific needs of the problem in an acceptable amount of time.

Simple feed-forward networks (with information flow proceeding only down towards the next layer) have wide-spread application. They can perform facial recognition, model experimental data in color and taste perception (Churchland), model psychological conditioning and memory (Gluck and Myers), and are widespread in data/statistical analysis in general (Bishop). Much has been done in the effort to make these networks more versatile. One can allow recursion (back-propagation of outputs),

complicated learning algorithms based on error minimization, apply statistical concepts such as Bayesian theory in guiding classification training, and more (see Arbib).

Needless to say, neural network research is a rich field. One can study many results that develop from modifications of these simple network constructions. The original hope behind the construction of these systems was to capture some of the computational power that nature has implemented in neurons and the brain. While it is worthwhile to work bottom-up in the sense of adding variability to our abstract neural system (via considerations mentioned above), the work I find most exciting is the efforts in working top-down by constructing more accurate models of real neural activity. If one can understand the mechanisms at work in the solutions developed by nature and evolution, then perhaps we will better understand the types of abstracted neural systems we need to use for specific problem sets. In the vein of this approach, it again seems best to move away from our abstracted model to look more closely at how real neurons act on their inputs and outputs.

Chapter Two

A Second Approach

The Biology

One typically illustrates a neuron as I did in Figure 1, but even this isn't universal as there are many types of neurons. Neurons are made up of dendrites, a soma, and an axon. A neuron fires by use of concentrations of ions within and without. In its resting state a certain amount of potassium ions are within the cell, and outside an amount of sodium ions (both types of ions are positively charged). As the amount of collective charge is uneven the outside is positive relative to the cell. So one says the resting potential is negative for the cell.

If there are enough excitatory signals received in the right way, so that they create the necessary voltage to reverse the desired states of the ions, then gates along the cell membrane will open allowing sodium to flow in, and potassium to flow out at a slower rate. At this stage the neuron is said to be depolarized, and the inside is then positive relative to the outside. Then the sodium gates close, but the potassium gates remain open. This allows the potassium to exit the cell, and is said to repolarize the cell. Even once the resting potential is reached the potential drops a little more as the potassium gates don't close fast enough. Then the cell works to restore resting potential. (Wolfe)

Immediately after firing there is a period called the refractory period where the neuron cannot fire again. This ion gating and refractory period behavior propagates down the axon until it reaches the synapse where it may trigger vesicles (sacs) to release their contents of neurotransmitters to other neurons. If the conditions are right the receiving neuron might fire. Although this description says a little more about the mechanism of firing and signal transfer than previously, it essentially is the same on or off, firing or not, behavior that is perfectly reflected in our simple neural network model. However, the interactions between neurons get a lot more interesting.

Each neuron has about 10^4 synapses. That is a vast number of connections. And within all of those the neurons react in different ways to different situations. Neurons, in order to propagate firing, must receive a certain amount of excitation from the surrounding environment. This necessary excitation energy is called the threshold. Usually one neuron's spike won't be enough to propagate to other neurons. The neurotransmitters sent by the presynaptic neuron (the one that originally fired) on average yield about 1mv to the potential of the postsynaptic neuron (the one receiving the signal). It takes about 20-30mv to spike the postsynaptic neuron. (Gerstner and Kistler) So within a short time several inputs have to come in to initiate another firing.

It still seems the neurons act as these little summing processors. But, in fact, the summing is nowhere near as simple and clean as checking for a prerequisite number of excitatory signal connections and subtracting the inhibitory. The addition tends to be nonlinear. Receiving one type of signal alters how future signals will be received.

Furthermore *where* the incoming signal touches the cell—the ends of the dendrites, at the dendritic spines, or the soma itself—alters how it sums with previous inputs. As well as the physical location, the configuration of input signals is important to neural response. Even the temporal configuration of inputs alter their effect on the neuron. This milieu of activity even brings into question the idea of a threshold. The nonlinear responses to inputs make a solid voltage threshold to initiate spiking only a simplification. (Izhikevich)

The nonlinearity involved in signal processing adds several layers of variability to what is usually conceived as a binary state unit. This variability moves us further and further away from easily modeling the behavior of such a system. Even beyond considerations of inhibitive and excitatory controls with regard to synapse/dendrite geometry (see Segev and London for an overview and further reading) our simplistic model fails in several other regards. For one, signals do not always move in just one direction. Sometimes a neuron can send signals backwards (soma to dendrite), creating a feedback mechanism that could help with neuronal learning. (Koch et al.) Yet another of the several failings has to do with the ion gates involved. Although the voltage-triggered sodium and potassium gates are at the forefront of most simplistic models of single neurons, there are a dozen or more other types of ion gates in the neurons of complex organisms. (Koch et al.) One example is the calcium-ion gates. Calcium gates behave similarly as the potassium and sodium gates, but have interesting properties of their own.

Besides acting on the electric properties of the neuron, calcium plays a very large role in affecting other cellular mechanisms. Calcium levels outside the cell can affect other ion gates (like the potassium gates previously mentioned), hence changing a neuron's response to stimulus and altering signal propagation on the individual neuron level. (Gerstner and Kistler) They can act as signaling chemicals, and their concentrations may be involved in modulation of synaptic strength. (Koch et al.)

On top of voltage-triggered and calcium-activated channels, one needs to take into consideration the role of neurotransmitters and other chemicals on the electrical state of neurons. Neurotransmitters can open or close other gateways in the membrane of the neuron to change its receptivity to signals. Chemicals such as NMDA (N-methyl-D-aspartic acid) and AMPA (alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) among others can affect neuron signal receptivity over long periods of time (long term potentiation or depression of neurons). (Segev and London)

Not only do neurons have far more variables than simply firing or not firing, we aren't even completely sure what type of activity represents information. Although early experiments with stimulus intensity to the peripheral nervous system had strong correspondence to increases in firing rate, we can't be sure that this type of behavior carries over to the entire nervous system. (Fotheringhame and Young) As I briefly mentioned above, there are several variables in neuronal activity. Any or all of these might somehow be information bearers. Besides rate coding, which can be problematic considering the speed of response times, there is some evidence that pulse coding - seeing

timing of individual spikes as vitally important to information carrying - has use in the brain. (Fotheringhame and Young) Although in some cases rate and pulse coding can be seen as equivalent, since the spiking rate is built upon the timing of individual spikes, there still are important differences in computational power. (See Gerstner and Kistler for details)

Yet another cause for modeling concern is that neurons can exhibit other activities beyond individual spike variability. Based on the time that the ion gates respond to electrochemical context, neurons can output periodic spikes. That is, neurons can output multiple spikes instead of singular spikes even without further input from other neurons. This behavior isn't allowed for in our simplistic artificial neural network from chapter one. This multiple spiking, or bursting, behavior is quite prominent in the brain, and has been proposed to explain a variety of effects on informational processing and on consciousness and attention. (Fotheringhame and Young; Gerstner and Kistler)

Let us investigate some of the models that have been made in the attempt to model real neurons more accurately, and which of these models can account for the behaviors not explained in the computational model from chapter one. Which variables allow for the robustness and strength we associate with real neural systems? Then what does this mean for instilling computer systems with "intelligent" behavior, and understanding the power of the brain?

The Computing

One could combine the approach from chapter one with the additional biological variability we wish to instill into the model. The weighted sums over the inputs will become synaptic weights on incoming current from other nodes. Then the activation of each node will be determined in continuous time by the electrical potential of the neuron/node. Using basic circuit analogues we can obtain expressions for how the electrical state will change over time in response to injected currents. This type of modification of our original network concept is used quite often. One example, which I will define exactly as Dezhe Z. Jin does, is given below.

$$\tau \frac{dV_j}{dt} = L_j + I_j - V_j - \sum_{i=1}^{N}\sum_{k=1}^{\infty} [G_{j,i}^{E} V_j + G_{j,i}^{I}(V_j - E_I)]\tau\delta(t - t_i^{(k)}) \qquad \text{(Jin)}$$

N – the number of neurons
$V_j(t)$ – membrane potential
L_j – resting membrane potential
I_j – external input to the jth neuron
$G_{j,i}^{E}$ - conductance of excitatory synapse from neuron i to j
$G_{j,i}^{I}$ - conductance of inhibitory synapse from neuron i to j
E_I – reversal potential of the inhibitory synapse (E_E is set to 0)
$t_i^{(k)}$ - time of the kth spike of neuron i
τ – membrane time constant
Θ_j – threshold of neuron j
R_j – reset potential (when $V_j > \Theta_j$ the potential is reset to R_j)
δ(x) – delta Dirac function (infinity at x = 0, 0 x $\neq$ 0) and $\int_{-\infty}^{\infty} \delta(x)dx = 1$

This equation defines how the potential changes over time. Input from other neurons is weighted via the membrane's permissibility to current – its conductance, G.

Then based on the accumulation of these infinitesimally brief spikes and external input the neuron can either spike or continue accumulating electrical potential; negative or positive based on the overall inhibitory or excitatory input. Notice that this model still uses the simplification of a threshold parameter, and avoids all considerations of realistic potential evolution during a spike event by use of the delta Dirac function and reset potential. This approximation is fairly common and an intuitive abstraction because the time scales involved in spike events compared to a neuron's sub-threshold potential evolution are extremely small.

Jin shows that a network of these neurons, with some additional assumptions, converges on a periodic spiking scheme based on the initial state of the neurons. These assumptions are that the network is fully connected, globally inhibitive, and that spiking events are stable and non-simultaneous. That is, a series of spiking events, if perturbed, don't push the timing of the next spike outside a certain range derived by Jin. Under these assumptions Jin showed that a spike series will repeat in perpetuity. (Jin) This can have great application in modeling problems that require the recreation of stable and patterned output. One example of such a problem is in the song generation of birds where the assumptions are plausible (see Hahnloser et al. for an investigation into the periodic spike schemes of song generation; Doya and Sejnowki for considerations in modeling and learning).

Despite the model's added variability as far as continuous time potential evolution, it still seems quite similar to our model of a neuron from chapter one. We will

see that without modeling the contributions of the various ion gates more accurately we lose a great deal in the versatility of spiking response evident in the brain. If the spiking response determines the information processing of a neuron, then we can't use such a simplified model to elucidate the capabilities of biological neural networks.

In order to improve the above model we need to incorporate the ability of ion channels to open or close in response to current. So, instead of inputs contributing current in the form $G(V\text{-}E)$, we need to include a variable keeping track of the ion channel's state. To achieve this we can average over the state of several ion gates to see the permissibility of the membrane to current. I will follow the definitions of Eugene M. Izhikevich.

$$I = \overline{g}p(V - E)$$

I is the current, $\overline{g}$ is the maximal population conductance, p is the average proportion of open ion channels, V is the electrical potential of the membrane, and E is the reversal potential, which determines the direction of current based on the concentrations of the various ion species.

$$p = m^a h^b$$

This equation is the average proportion of open ion channels across a membrane with a activation gates and b inactivation gates. m is the probability of an activation gate being open, and h is the probability of an inactivation gate being open. Every ion channel has an associated number of gates. For an ion channel to be open all of its gates must be open. Hence the probability of ion channel being open, and consequently the average proportion of open ion channels across a large population, will be as given. Each species

of voltage-gated ion channels can be described in this manner. (see Izhikevich for a summary of experimentally obtained values) With these models of ion channel behavior our model of an individual neuron's response to current is as follows.

$$C\frac{dV}{dt} = C\dot{V} = I - \sum_i \overline{g}_i m_i^{a_i} h_i^{b_i} (V - E_i)$$

$$\frac{dm_i}{dt} = \dot{m}_i = (m_{i,\infty}(V) - m_i) / \tau_{m,i}(V) \ , \ \ i = 1,\ldots,n$$

$$\frac{dh_i}{dt} = \dot{h}_i = (h_{i,\infty}(V) - h_i) / \tau_{h,i}(V), \ \ i = 1,\ldots,n$$

C is a constant called the capacitance. The functional forms of $m_{i,\infty}(V)$, $h_{i,\infty}(V)$, $\tau_{m,i}(V)$, and $\tau_{h,i}(V)$ are experimentally determined, where $m_{i,\infty}(V)$ and $h_{i,\infty}(V)$ are sigmoid, and $\tau_{m,i}(V)$ and $\tau_{h,i}(V)$ are unimodal. (Izhikevich)

Alan Lloyd Hodgkin and Andrew Huxley were some of the first to use this technique to describe spiking dynamics in neurons. They used this approach to model spiking behavior in the axon of a giant squid, and the predictions of the model were correct to within 10 % of observed values. (Churchland and Sejnowski) For this work they won a Nobel Prize in 1963.

This model, however, is not always amenable to finding explicit solutions. So without explicit solutions we must content ourselves with numerical approximations or

general behaviors. To find these general behaviors I will introduce some basic dynamic systems theory following closely the introduction by Izhikevich.

A one-dimensional dynamic system is in the form

$$\dot{V} = F(V), \qquad V(0) = V_0 \in R\,.$$

The progression of the variable V depends on the form of the function F. So in order to make educated conclusions about the behavior of V given some starting value V_0 we can look at the graph of F.

For example, consider the system

$$\dot{V} = \sin(V)$$

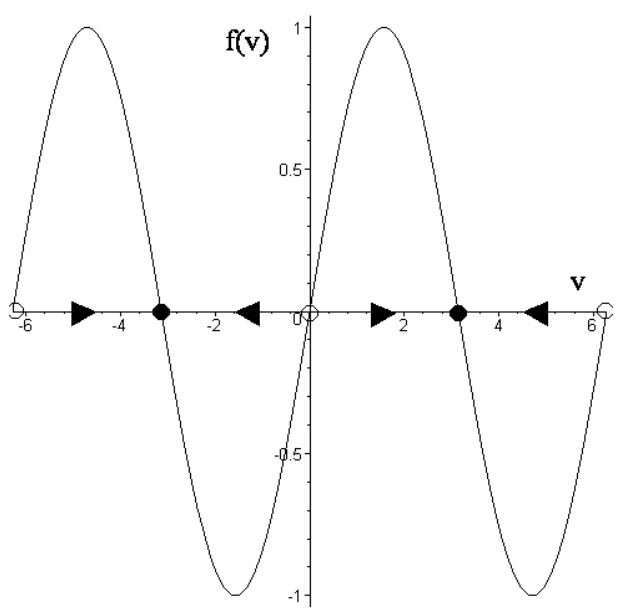


Figure 8
A graph of sin(v) showing the direction of flow and fixed points on the horizontal axis.

As this is a one-dimensional system we are only concerned with the values of V lying on the horizontal axis. The functional form of F, in this case the sine function, fully determines the qualitative behavior of V at any initial condition. For example, if V_0 is at $\pi/2$, then V will increase because sine is positive there. This increase will continue until it gets near the value of π. At π sine is zero, and so the change in V approaches zero. Similarly if V is $-\pi/4$ then V will decrease towards $-\pi$ because sine is negative in this region, and zero at $-\pi$. Now if it happened that V_0 is at the origin then we would expect no change in V, but if perturbed slightly then V_0 is drawn away to either side. If V_0 is at $\pm\pi$, and slightly perturbed, then it will be drawn back to the point.

This behavior defines the stability of the zeroes, or fixed points, of F. If when slightly perturbed, V is repelled, the fixed point is called unstable. If V is drawn back in, the fixed point is called stable. This only leaves one other possibility – a mix of the two. When this happens, the fixed point is called half-stable. An example of this third case is illustrated in Figure 9.

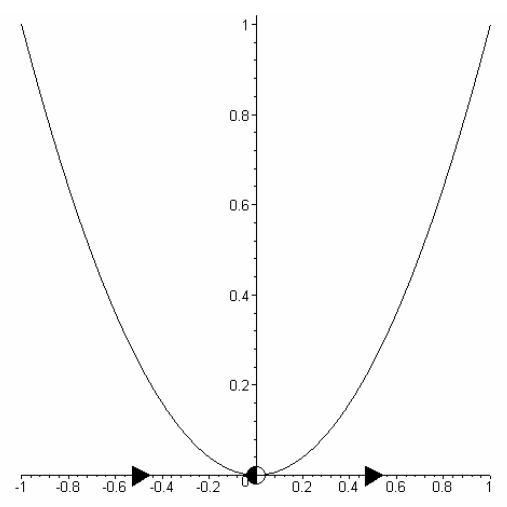


Figure 9
Illustration of half-stable fixed point ($f(v) = v$^2)

Now we are ready to look at what happens as we inject current into the neuron. The system becomes $\dot{V} = I + F(V)$. So now, depending on the sign of I, our F curve is shifted up or down. For $F(V)=V^2$, we have the following situations.

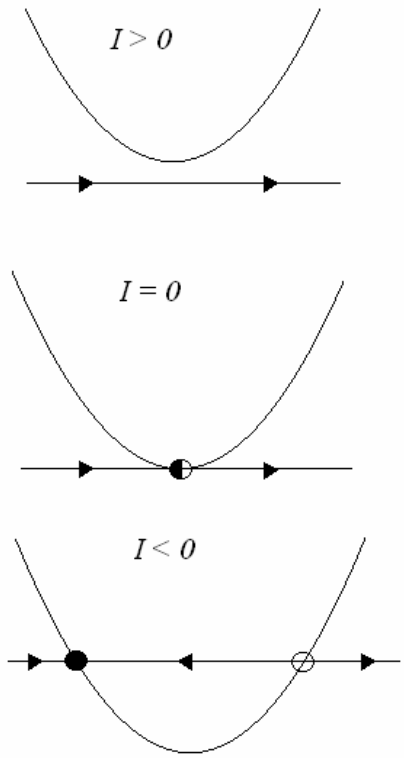


Figure 10
Phase portraits for $\dot{V} = I + F(V)$, as I changes

As can be clearly seen in Figure 10, changing the injected current from negative to positive causes the appearance of fixed points. When the phase portrait of a dynamic system changes in its overall qualitative behavior, it is said to have undergone a bifurcation. (Strogatz) In higher dimensions things obviously can get more complicated, but the ideas are the same as in one dimension. Depending on the forms of the functions governing the change in the variables, we have certain fixed points or fixed cycles that can be changed by altering some bifurcation parameter. This change in equilibria brings

about different qualitative behavior. And so it is exactly these bifurcations that define the type of computational properties a neuron can have.

If one can establish that a neuron's dynamics correspond to a model with similar assemblies of fixed points that can undergo the same bifurcations, then we have a valid model. As models are solely concerned with capturing key aspects of qualitative behavior, if we have a simple model that corresponds to the more complex, biologically accurate one then we can instead study the simpler one.

It is this idea of simplification via capturing the key qualitative behaviors that Izhikevich follows in his use of "canonical models". He classifies certain characteristics of potential response to different types of stimulation in terms of the bifurcation necessary to explain that behavior. Then based on which bifurcations are needed he uses the corresponding canonical model.

This is an exciting way to approach the problem of neural modeling, and can reliably reproduce much of the variety of spiking behaviors evident in the brain. Based on this theory models have been created of motor and sensory systems (Kopell and Ermentrout), associative memory (Singer), and neural synchrony. (Hoppensteadt and Izhikevich)

Conclusions

We have established both that computers can use artificial neural networks to model any functional mapping and, through the work of Izhikevich and others, that most activity in a single neuron can be modeled tractably. While consciousness has yet to be reliably identified in neural activity, if consciousness can be identified then it stands to reason that it can be modeled. Neurons in certain regimes can be simplified to specific archetypes, or canonical models. With the aid of these archetypes we can capture the functional mappings that don't fall out easily from the abstracted neural networks of chapter one. While dynamical system models of networked neurons are orders of magnitude harder to simulate (see Mirollo and Strogatz), applying dynamical system theory elucidates key features that can guide simplification of these structures. (Ermentrout)

In the brain much research has been done trying to illustrate the neural correlates of conscious behavior in terms of attentional modulation and pattern memory. Even the some of the simplest conductance based models can capture temporal patterning. This is illustrated by the model used by Dezhe Z. Jin. When one incorporates more plausible models of complete potential evolution in the neuron this behavior reappears. Neurons that can generate spike trains with arbitrary frequency can be simulated by dynamic system models of oscillators. Then when these oscillators are connected in networks their behavior is governed by their synchronicity.

The idea of neural synchrony is not alien to the field of neuroscience. Synchrony across neural systems has been proposed to be responsible for coincidence detection and feature binding (Singer; Crick and Koch; see Gray for a review), the formation and recollection of memory states (see Izhikevich; Érdi and Szalisznyó), and even perceptual organization (Finkel et al.). Different synchronous firing states could possibly correspond to a memory state. Then the configuration of these memory states in the activation space could represent more complex relations. (see Churchland; Phillips) This configuration could reflect the presence of multiple attractor states – such as multiple stable fixed points or stable cycles. (see Hertz for methods of computing with attractors) With the qualitative models of such dynamic behavior, even this can be modeled. (see Kuramoto for network behavior of coupled oscillators)

Even without the exact neural correlates of information handling much has been done showing some potential information processing capabilities of various electrochemical and morphological aspects of the neuron. Dendritic trees can be interpreted as performing some fundamental coincidence detection and logic operations. (Koch) Population codes and learning mechanisms in neurons can be shown to perform a variety of statistical operations and coordinate transforms crucial to muscle maps and visual scene parsing. (see Dayan and Abbott; Amirikian and Georgepoulos; Simoncelli and Olshausen) Even chaos in biological systems might have a higher functional role than noise in information handling. (Aihara; Glass)

We don't yet know the exact neural correlates of all the informational processing capabilities at work in our mind, and so we can't yet implement full emulations in machine architecture, but investigations such as the ones mentioned above and playing with the more powerful models from chapter two might allow us to make educated decisions about what experiments in neuroscience we should do next.

If we can determine that some sensory or attentional system needs a form of functional manipulation that requires the presence of an ion channel with a characteristic bifurcation type, then we can pursue pharmaceutical means to block that channel and measure the affect on that system.

Even these neural models might not yet be a complete description of information processing in the brain. As I mentioned previously, the glia also contribute to neural dynamics, but are not included in most models used today of biological information handling. Perhaps modeling efforts in the future will follow recent experimental evidence in their role in neuromodulation.

While there are certainly philosophical debates about the true understanding of a computer system that can mimic, or even fully emulate, the behavior of humans (see Clark and Eliasmith), the goal for most of artificial intelligence research is not about making conscious computers, but systems that can make sensible choices in dynamic situations. I feel that research into artificial neural networks is quite persuasive in establishing this possibility. I harbor no illusions that such implementation will be easy,

but it is possible. This possibility is what continues research in the field, and was my motivation for presenting these results.

Bibliography

Aihara, Kazuyuki. "Chaos in Neural Systems". Arbib 208-12.

Amirikian, Bagrat, and Georgopoulos, Apostolos P. "Motor Cortex: Coding and Decoding of Directional Operations". Arbib 690-96.

Arbib, Michael A., ed. The Handbook of Brain Theory and Neural Networks. 2nd ed. Cambridge: MIT Press, 2003.

Bishop, Christopher M. Neural Networks for Pattern Recognition. Oxford University Press, 1995.

Churchland, Patricia Smith. Brain-Wise Studies in Neurophilosophy. Cambridge: MIT Press, 2002.

Churchland, Patricia Smith, and Sejnowski, Terrence J. The Computational Brain. Cambridge: MIT Press, 1992.

Clark, Andy, and Chris Eliasmith. “Philosophical Issues in Brain Theory and Connectionism.” Arbib 886-88.

Crick, Francis, and Koch, Christof. "Towards a Neurobiological Theory of Consciousness". Seminars in Neuroscience 2: 263-275, 1990).

Dayan, Peter, and Abbott, LF. Theoretical Neuroscience Computational and Mathematical Modeling of Neural Systems. MIT Press, 2001.

Doya, Kenji, and Sejnowski, Terrence J. "A Computational Model of Avian Song Learning". Gazzaniga 469-482.

Érdi, Péter, and Szalisznyó, Krisztina. "Hippocampal Rhythm Generation". Arbib 533-39.

Ermentrout, Bard. "Phase-Plane Analysis of Neural Nets". Arbib 881-85.

Fields, R. Douglas. “The Other Half of the Brain.” Scientific American Apr. 2004: 55-61.

Finkel, Leif H., Yen, Shih-Cheng, and Menschik, Elliot. "Synchronization: The Computation Currency of Cognition". Penn State Publications, 1998.

Fotheringhame, David K., and Malcolm P. Young. “Neural Coding Schemes for Sensory Representation: Theoretical Proposals and Empirical Evidence.” Rugg 47-76.

Gazzaniga, Michael S., ed. The New Cognitive Neurosciences. 2nd ed. Cambridge: MIT Press, 1999.

Gerstner, Wulfram, and Werner M. Kistler. Spiking Neural Models Single Neurons, Populations, Plasticity. Cambridge University Press, 2002.

Glass, Leon. "Chaos in Biological Systems". Arbib 205-08.

Gluck, Mark A., and Myers, Catherine E. Gateway to Memory An Introduction to Neural Network Modeling of the Hippocampus and Learning. Cambridge: MIT Press, 2001.

Gray, Charles M. "The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well". Neuron 24: 31-47, 1999.

Hahnloser, Richard H.R., Kozhevnikov, Alexay A., and Fee, Michale S. "An Ultra-Sparse Code Underlies the Generation of Neural Sequences in a Songbird". Nature 419: 65-70, 2002.

Hertz, John. “Computing with Attractors.” Arbib 248-52.

Hoppensteadt, Frank C., and Izhikevich, Eugene M. Weakly Connected Neural Networks. New York: Springer-Verlag, 1997.

Izhikevich, Eugene M. Dynamical Systems in Neuroscience The Geometry of Excitability and Bursting. Cambridge: MIT Press, 2007.

Jin, Dezhe Z. "Fast Convergence of Spike Sequences to Periodic Patterns in Recurrent Networks". Physical Review Letters 89: 20, 2002.

Koch, Christof. Biophysics of Computation Information Processing in Single Neurons. Oxford University Press, 1999.

Koch, Christof, Chun-Hui Mo, and William Softky. “Single-Cell Models.” Arbib 1044-49.

Schuichi Koizumi, Kayoko Fujishita, Makoto Tsuda, Yukari Shigemoto-Mogami, Kazuhide Inoue, "Dynamic Inhibition of Excitatory Synaptic Transmission by Astrocyte-Derived ATP in Hippocampal Cultures". PNAS 100: 11023-11028, 2003.

Kopell, Nancy, and Ermentrout, G. Bard. "Chains of Oscillators in Motor and Sensory Systems". Arbib 201-05.

Kozlov AS, Angulo MC, Audinat E, Charpak S, "Target Cell-Specific Modulation of Neuronal Activity by Astrocytes". PNAS 103: 26 10058-10063, 2006.

Kuramoto, Yoshiki. "Collective Behavior of Coupled Phase Oscillators". Arbib 223-26.

Minsky, M. L., and Papert, S. A. Perceptrons.Expanded ed. Cambridge: MIT Press, 1990.

Mirollo, R.E., Strogatz, S.H. "Synchronization of Pulse-Coded Biological Oscillators. SIAM Journal of Applied Math 50: 1645-1662, 1990.

Olazaran, Mikel. "A Sociological Study of the Official History of the Perceptrons Controversy". Social Studies of Science 26: 3 611-59, 1996.

Phillips, William A. “Theories of Cortical Computation.” Rugg 11-46.

Poznanski Roman R, Riera Jorge J, "fMRI Models of Dendritic and Astrocytic Networks". Journal of Integrative Neuroscience 5: 2 273-326, 2006.

Rugg, Michael D., ed. Cognitive Neuroscience. Cambridge: MIT Press, 1996.

Russell, Stuart, and Norvig, Peter. Artificial Intelligence A Modern Approach. 2nd ed. Pearson Education, 2003.

Segev, Idan, and Michael London. “Dendritic Processing.” Arbib 324-32.

Simard Marie, Arcuino Gregory, Takano Takahiro, Liu Qing Song, Nedergaard Maiken, "Signaling at the gliovascular interface". The Journal of Neuroscience 23: 27 9254-9262, October 8, 2003.

Simoncelli, Eero P., and Olshausen, Bruno A. "Natural Image Statistics and Neural Representation". Annual Reviews in Neuroscience 24:1193-216, 2001.

Singer, Wolf. "Synchronization, Binding and Expectancy". Arbib 1136-43.

Strogatz, Steven H. Nonlinear Dynamics and Chaos. Westview Press, 1994.

Volterra Andrea, Meldolesi Jacopo, "Astrocytes, from Brain Glue to Communication Elements: The Revolution Continues". Nature Reviews Neuroscience 6; 626-640, August 2005.

Wolfe, Stephen L. Introduction to Cell and Molecular Biology. Belmont: Wadsworth, 1995.