

# The Neuronal Communication System

Jeffrey Mehlman

Electrical Engineering Department, Stanford University  
Stanford, California 94305  
jmehlman at stanford dot edu

**Index Terms**—Perceptron, Neuron, Dendritic Integration, Neural pattern recognition, Synaptic Correlation, Synaptic Weight Distribution, Spike-timing Dependent Plasticity, Quadratic Optimization

**Abstract**—The brain communicates via electrical signaling. The molecular mechanisms which drive this communication are quite complex. In this paper, we compare intra-neuron signaling properties with communications networks, and find that there are many similarities. We review some existing neuroscientific findings which exemplify this neuron-communication system link. Additionally, we review existing mathematical models and theories for neural functionality and communication. In particular, we look at long-term synaptic plasticity mechanisms which amplify important synaptic inputs and attenuate noisy ones over time. We prepare a basic model based upon communication theoretic principles, and derive a mathematical approximation to it. We find that our model is largely based on correlation of input patterns, which matches quite well with some existing neuroscientific theories as well as neural network principles. With simulation, we find that the model predicts some of the basic properties effects of synaptic plasticity.

## I. INTRODUCTION

Beginning with Hodgkin & Huxley and their peers in the 1930s [1], it was theorized and then shown that the basis for neuronal communication was electrical in nature. Since then, electrical engineers have been both contributors to and benefactors of neuroscientific discoveries. An entire field of electrical engineering, called "neural networks," has risen out of this interdisciplinary field, in which we mimic the ability of the brain to learn through simplistic models of neurons and networks of them. Much as the smallest element of the integrated circuit is a transistor, the neuron seems to be the smallest computational element in the brain. While we can characterize the functionality of a transistor quite well, we are still learning about many of the neurons internal behaviors.

From the perspective of communications theory, both intra- and inter-neuron communication is an interesting topic. Here, we focus on intra-neuron communication. The rest of the paper is organized as follows: In section II we cover some basic biophysical characteristics of the neuron. In section III, we will cover some mathematical basics, related work, and specific biophysical characteristics of the neuron which are interesting from a communications perspective. In section IV, we will describe our basic system model in detail, and review some of it's links with existing theories and systems. In section V we will review some of our results with this model. We summarize our findings as follows: against our metric of signal-to-interference (SIR) ratio, we find that the optimal neural characteristics

can match quite well with existing theories and experimental measurements.

## II. NEURAL BASICS

The human brain is composed primary of two types of cells: neurons and glia. As shown in Figure 1, a single neuron is comprised of three main sections - a dendritic tree or arbor, a soma (cell body), and an axon [2]. The axon generally functions as the "output" and the dendritic tree takes "inputs" from other neuron's axonal outputs. The interface between an axon terminal and a dendritic tree is commonly called a *synapse*. The space between the axon terminal and the dendrite is called the *synaptic cleft*.

Glia are generally believed to be support cells for neuronal function, and are extremely interesting from an evolutionary perspective. The axonal membrane is generally a very leaky one, and as an unprotected axon shrinks in diameter the propagation speed of APs decreases. Two evolutionary paths have been developed to overcome this issue. In some animals, important axons are enlarged in order to support speedy (but leaky) propagation of APs. Squid axons are an example of this, and were the basis of Hodgkin and Huxley's experiments because of their large size. For a small number of neurons, this approach works, but it is limited by the fact that the neuron's energy consumption and volume must be increased to provide AP speed.

Some glia (Schwann cells) have evolved to be a second approach to reduce leakage and increase propagation speed of APs. These cells grow myelin sheaths around the axons of most neurons in the brain. The sheathing serves to reduce energy loss in the signal thereby increasing propagation speed. The axon plus myelin diameter required to support the same propagation speed as an un-sheated axon is orders of magnitude smaller, while additionally requiring less energy. This is one of the many examples of ways that the human brain has evolved to support incredible computational abilities as very low energy consumption.

Nearly all of the inter-neuron signaling in the human brain takes places via electrochemical transfer. When specific criterion are met, a given neuron will initiate an *Action Potential* (AP) at the axonal hillock (in the soma). An AP is an electrical pulse which is carried by ionic currents (mostly Sodium, Calcium, and Potassium ions). The pulse propagates through the entire cell body, and is maintained down the length of the axon by periodically spaced amplification sites where myelin sheathing is not present, called *Nodes of Ranvier*. We can think

of these sites as selection amplify-and-forward nodes. So long as the signal exceeds some voltage threshold, enough ionic channels will open such that a runaway effect occurs and the pulse is effectively amplified. It seems that these node spacings are optimized to be as far apart as possible, while maintaining a high probability of re-amplification [2].

When the AP reaches the axon terminal, it releases chemicals called *neurotransmitter* which migrate extracellularly across the synaptic cleft and have either an excitatory or inhibitory effect on the dendrite. In the excitatory case, the dendrite synapse generates an electrical current by temporarily altering properties of the cell membrane to allow ionic currents to flow into the cell. In the inhibitory case, the dendrite can both alter the properties of the cell membrane to have a shunting effect and suppress any currents generated nearby as well as generate a negative (outgoing) current. These pulses are termed Post Synaptic Potentials (PSPs). One neuron can have tens of thousands of synaptic inputs, with most inputs being excitatory.

While the equations that govern the Hodgkin-Huxley model are complex, they result in a required (approximate) "threshold voltage" at the soma / axon hillock in order to generate an AP. The voltage in the soma at any given point is largely a function of the dendritic inputs from other neurons. A simplistic view is that the excitatory PSPs (EPSPs) and inhibitory PSPs (IPSPs) at a given point in time summate in the soma, and the threshold voltage criterion is applied to determine whether or not an AP should be generated.

While a single action potential does not consume much energy, one must consider that the human brain contains on the order of 100 billion neurons. In fact, the average human brain consumes approximately 20% of the total energy consumed by the entire body, while accounting for only 2% of it's total mass. Nearly half of the energy consumed in the brain is due to action potential generation and propagation [3]. While these statistics vary across different animals, one fact is clear: the brain consumes a lot of energy with respect to it's mass and size. None the less, it is an incredibly efficient computational system, which outperforms state-of-the-art, energy-hungry computers at many tasks. Researchers here at Stanford seek to harness this efficient computing with silicon-based neural circuits.

We can trace the energy efficacy of the brain down to it's smallest constituents: neurons. Clearly, controlling the average AP generation rate is an important factor in overall energy conservation. There is a considerable body of existing research which considers the tradeoff between volumetric/energy conservation and information storage. Here, we consider each neuron as a small computational device with many inputs and one output. The functional purpose is to compress information from the thousands of inputs down to a single decision at the axon.

Even this idea is a partial simplification. Neurons can encode information about the state of it's synaptic inputs across multiple pulses with dynamic rate-modulation strategies. A number of information theorists and theoretical neuroscientists have also attempted to apply Shannon's theorems to these *neural codes*. The transfer function and associated information

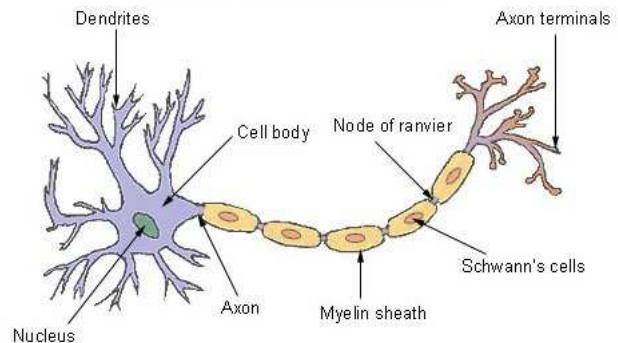


Fig. 1. The Neuron

compression algorithm for a neuron is very complex and to this day is not well understood.

### III. BACKGROUND & RELATED WORK

One example of neural energy and volume optimization is demonstrated in [5]. The authors determine through simulation that there should exist a minimum limit on diameter, and thus volume, of the axon. At the molecular level, electrochemical signals are generated and propagated by the stochastic conditions of ion channels. The state of channel activation is a function of membrane voltage and subject to thermodynamics. Thus, some number of channels are always "active" and allowing current to flow. In normally functioning axons, these stochastic conditions contribute to a noise floor.

As the diameter of the axon decreases, the effect of a single "open" channel increases and the noise floor increases. At some minimal diameter, generation of "spontaneous APs" (in other words, bit errors due to significant channel noise) occur due to the influence of stochastic channel dynamics. This is illustrated in figure 1 (A,C) from this paper. The authors found that the SAP rate was largely insensitive to biophysical parameters other than the axonal diameter. Data mined from dissection of actual brain tissue supports the simulation-based  $0.1\mu m$  limit. As the volume decreases, energy cost to propagate APs also decreases. Thus, operating near the noise-limit provides savings in both volume and energy consumption.

We can think of this as some hard signal-to-noise limit at which the output (the axon) begins to make many wrong decisions. There are a plethora of neuronal characteristics which we can compare to communications theory and systems. For many years now, information theory has been utilized to attempt to determine the information capacity of a neuron with questionable results.

Originally, it was believed that the dendritic tree functioned as a set of passive cables, whereby each input was independently propagated through the tree to the soma and summed linearly there. This model is called the "point" neuron or peceptron, and is described mathematically as

$$r = \sum_{i=1}^n w_i x_i \quad (1)$$

where  $w_i$  and  $x_i$  are the synaptic weights and synaptic inputs respectively, and  $r$  is the signal seen at the soma [6].

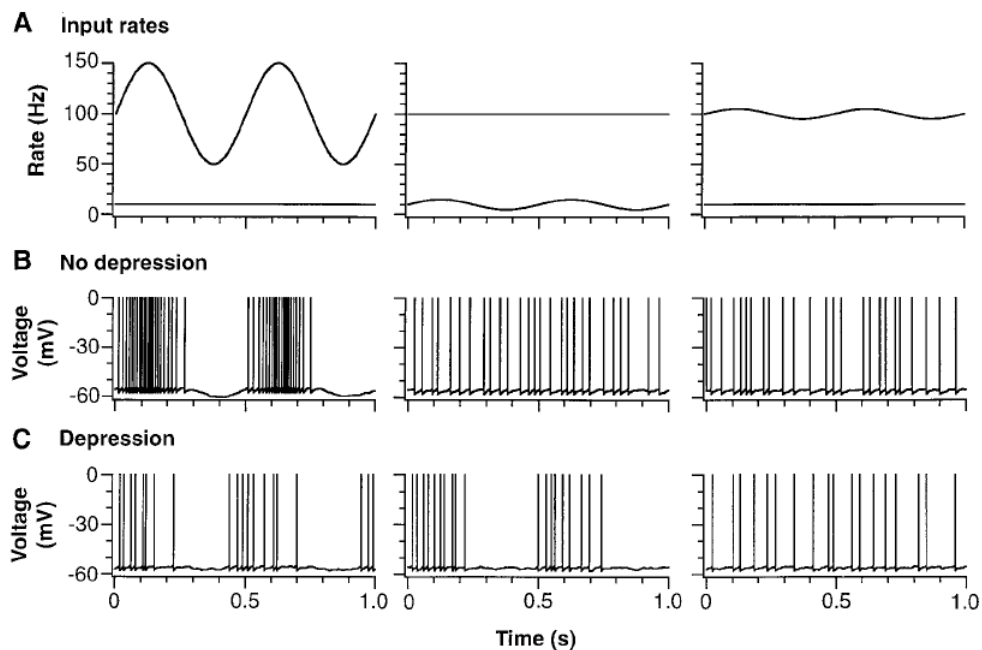


Fig. 2. Simulation: Synaptic Depression leads to better input differentiation [4]

If  $r$  exceeds some threshold, the axon hillock generates an Action Potential (AP), which propagates down the axon as well as back up into the dendritic tree. This back-propagating action potential (BPAP) is of great interest and importance to synaptic weight modification (plasticity) and other feedback-control mechanisms.

In recent years, researchers have also discovered that function of the dendritic tree itself is much richer than originally believed. In fact, it has been shown in simulation and through experiment that each dendritic branch functions as a computational subunit which can perform supra- or sub-linear summations of dendritic PSPs on that branch [6]–[11]. This functionality along with some other proposed abilities of the dendritic tree lead to a much richer neural model, with greatly enhanced computational capability. We discuss this topic in greater detail in section III-D.

#### A. Synaptic Plasticity

Morphological studies of Purkinje and cortical neurons in the human brains have illuminated their extensive dendritic arborizations. These cells can accept on the order of 100,000 synaptic inputs. While the number is much lower for other cell types, it seems unlikely that all of these inputs could contribute to a single output. Supporting this fact is research into response patterns for these neurons which finds that they respond only to specific subsets of inputs. Some studies also indicate that anywhere from 80-90% of the synaptic inputs to these cells are inactive and produce no noticeable voltage change at the cell soma. These ineffective inputs are called "silent synapses." This raises an important question: How do these cells determine which inputs are important? It seems very unlikely that random connections at the cellular level could govern very orderly higher brain functions.

Much more likely is that some form of selectivity is at work in determining the importance of each synaptic connection. It is believed that this occurs through synaptic weight modification via a wide variety of biological mechanisms. This feature is generally termed *Synaptic Plasticity*, and many neuroscientists believe that it is the molecular basis for memory and learning. In 1949, Dr. Donald Hebb was one of the first to propose such a mechanism [12], now appropriately called Hebbian learning. The basic theory is often succinctly stated as: "cells that fire together, wire together."

Research on this phenomenon *in vivo* or *in vitro* is complicated by the seemingly wide variety of participating biological mechanisms. There are believed to be both long-term (> minutes) and short-term (milliseconds to seconds) forms of synaptic plasticity which lead to gain or attenuation of a given synaptic input weight, and these mechanisms seem to be sometimes competing!

Additionally, plasticity occurs in to both feed-forward and feed-back directions [13]. Post-synaptic synapses can communicate with pre-synaptic axon terminals via retrograde chemical messengers that travel extracellularly back across the synaptic cleft.

Parametrically speaking,  $w_i$  in equation 1 is modified based upon it's interaction with the neuron. In biological research, weight gain is termed *potentiation* and attenuation is termed *depression*.

In [14], using in-vitro measurements, the authors first developed a quantitative model for short-term plasticity in rat cortical cells and then built a simulation model for it. Measurements of short-term potentiation and depression were fit to a product of exponential functions with different time constants. In a second, landmark paper, the same authors used that model to demonstrate the efficacy of short-term synaptic depression as a dynamic gain control mechanism for cortical

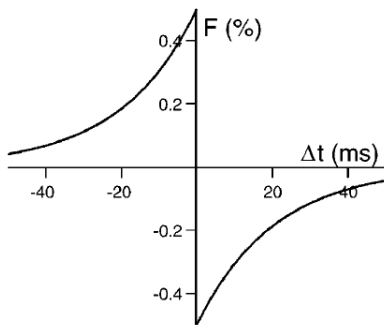


Fig. 3. Spike Timing Dependent Plasticity: Modification Function

neurons [4]. Cortical neurons, as aforementioned, have tens of thousands of inputs with firing rates ranging from 1 to 200 Hz. How these neurons differentiated between slowly firing inputs and quickly firing ones was not well understood prior to this discovery.

The effect of depression is shown in figure 2. The simulation used two sets of independent, Poisson process inputs with mean rates of 10 Hz and 100 Hz respectively. Simulation runs are divided by column. In descending order, the rows plot the mean firing rate for both input groups, the output spike train without depression, and the output spike train with depression.

In the first simulation run (column A), the mean firing rate of the 100 Hz group was sinusoidally modulated by 50% around the mean and the 10 Hz group was held constant. In the second (column B), the two input groups were swapped. In the third (column C), the 100Hz group was modulated by 5Hz (which was the same modification for the second simulation run). We see that the depression-enabled output trains for simulations one and two are nearly identical, and that depression has no effect on the output train for simulation three. The effect of depression on the output spike train is quite clear: it allows for the neuron to differentiate between relative changes in mean synaptic firing rates.

This phenomenon can be explained as follows: For each synaptic input, the balance between its mean firing rate and synaptic depression will lead to a steady-state synaptic weight which is inversely proportional to the mean firing rate. Through this balance, relative (%) changes of each input will be weighted equally at the cell soma. It is a form of power control, where higher average firing rates have lower peak powers, and therefore all firing rates have relatively constant average power.

Another of these observed and seemingly very important mechanisms is spike-timing dependent plasticity (STDP). Some neuroscientists believe that STDP is the cellular mechanism for learning and memory. Its functionality can be simply described as follows: If a synaptic EPSP at synapse  $x_i$  precedes a post-synaptic AP within some time constraint, the synaptic weight  $w_i$  is increased. If a synaptic EPSP follows a post-synaptic AP, its weight  $w_i$  is decreased. A proposed STDP function is shown in figure 3 [15]. In effect, STDP stochastically increases the importance of synaptic inputs which are correlated with output AP generation, and decreases

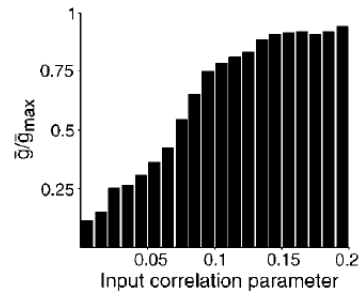


Fig. 4. Spike Timing Dependent Plasticity: Weight vs. Correlation

the importance of inputs which are not. It treats the latter inputs as noise and drives their importance down to some minimum weight value (possibly zero!). This is a form of *unsupervised* Hebbian learning.

In [15], the authors go on to characterize the effect of STDP on inputs with specified correlation. They show that as correlation increases, the inputs tend to have more significant weight. For every input, a 'correlation multiplier'  $c_i$  was specified. The correlation between any two inputs was then  $c_i c_j$ . In figure 4, they perform a simulation run with 200 inputs and STDP. We see the effect of the correlation on the synaptic weights - more correlated inputs have larger gain. Note that in these simulations, a maximum ( $g_{max}$ ) and minimum (0) gain for the synapses was specified.

### B. Synaptic Weight Distribution

It is clear that our brains have been driven down towards minimal energy consumption whilst maintaining their incredible computational power. Part of the story behind this functionality is now clear. Synaptic plasticity allows for neurons to separate important input patterns from unimportant ones, thus reducing the AP generation rate and associated energy costs. The next question is: towards what optimization metric is synaptic plasticity working? A number of recent research efforts have been directed at understanding the effect of synaptic plasticity, and its possible optimization goals.

One such approach is investigation of the synaptic weight distribution for a neuron. In [16], the authors chronicle a number of recent studies of synaptic weight distributions in various neuronal types, including one of their own. Most studies include only the effect of excitatory synapses and ignore inhibitory ones. In their study, theoretical approaches were used to derive an optimal weight distribution based upon information theoretic concepts originally posed by Gardner in the late 1980s.

The authors find that the critical capacity of a perceptron-model neuron was 0.83 bits per synapse [17], which is lower than the original 2 bits per synapse predicted by the original noiseless models. This later result was, however, supported by later findings. Critical capacity is defined as the maximum number of patterns stored divided by the number of inputs. The authors of [17] extended Gardner's model to include an error-regime (some patterns stored with non-zero error) in [18], and utilized these findings in deriving an estimate of capacity for a Purkinje cell.

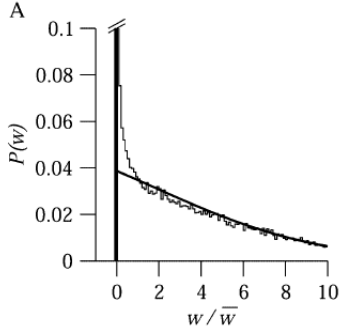


Fig. 5. Optimal Synaptic Weight Distribution

An example weight distribution is shown in figure 5 above. It is given by the equation [17]

$$P(W) = H(-B)\delta(W) + \frac{1}{\sqrt{2\pi}W_s} e^{-\frac{1}{2W_s^2}(W+BW_s)^2} \Theta(W) \quad (2)$$

It is a delta pulse at zero (these are called silent synapses), with value  $H(-B)$ . The right-hand term is a gaussian tail of non-negative synaptic weights. The general form of this distribution was found to fit fairly well with observed distributions. The entire synaptic weight distribution is a function of the noise SD and the desired BER for the system. By fitting these parameters to measured synaptic weights for a Purkinje cell, they estimated its noise, BER, and information capacity.

A related and equally interesting finding is that a significant proportion of synapses generate no noticeable membrane voltage change at the cell soma. As aforementioned, these have been termed "silent synapses," and are linked with the effects of synaptic plasticity. The proportion of silent synapses increases as the noise level is increased, which is also of interest. It is possible that some form of waterfilling across correlation is taking place, such that inputs with more significant correlation are allocated more power, and inputs below some correlation/noise threshold are driven down to zero. As shown in the correlation vs. synaptic weight plot, this idea is supported by simulation results. Unfortunately, we were unable to derive a model which clearly fit this conceptual idea. This modeling is left for future work.

### C. Information Theory and Neuronal Computation

There have been many attempts to characterize the performance of a neuron with respect to information theoretic concepts. Cover originally considered the problem of pattern recognition for a perceptron in the late 1950s [19], and found that for a noiseless perceptron the number of stored patterns should be  $P=2N$ , where  $N$  is the number of inputs. The problem was considered geometrically, where randomly drawn input sets were considered in a hypercube. So long as a hyperplane could be drawn to separate the inputs into two distinct classes, the pattern was considered stored. The same result was rederived in a famous paper by Gardner, utilizing a statistical mechanics approach [20].

More recent efforts have been geared at utilizing information theory to explain some of the natural biological optimizations which take place. Signal-to-Noise ratios and Shannon

capacity are often used to measure the quality of memory and information storage in neural networks and perceptrons. In [21], the authors consider the impact of discrete-valued synapses on the quality of information storage. Some neurological research indicates that synapses take values only in discrete steps. The authors analyze the relationship between information storage and the number of synapse states.

One very interesting and recent result deals with a small biological two-layer neural network which is modeled at a functional level [22]. It consists of a many granule cell input stage with a Purkinje cell decision stage. These networks are primarily involved in motor control. The granule cell inputs provide a number of primary signals, each with independent signal and noise levels,  $p_i = \alpha_i s + n_i$ .

The signal  $s_i$  and noise  $n_i$  and are mean 0 and variance 1 and  $\sigma_i$ , respectively. One granule cell input carries an error signal which provides the learning mechanism for the Purkinje cell. They theorize that a certain gain,  $\gamma$ , for every input of interest is desired at the Purkinje cell output, and develop a model which uses a "covariance rule" between the error and primary signals to modify the gain for each input in the Purkinje cell layer.

They find that the optimal weighting/gain vector components are given by

$$w_i^{opt} = \frac{\alpha_i/\sigma_i^2}{\gamma + \sum_{j \neq i} \alpha_j^2/\sigma_j^2} \quad (3)$$

which bears close resemblance to power control results in communication networks with interference. Using these rules they simulate the system and find that undesired noisy inputs are driven to zero over time. They argue that biologically, these noise signals force their associated synapses to become "silent" (a recurring idea!). In addition, the covariance update rule they utilize bears significant resemblance to the Foschini-Miljanic algorithm for power control in multi-user interfering networks.

Another interesting information-based model is proposed in [23]. They develop an optimization-based model which predicts the existence of synaptic plasticity. They propose that neurons attempt to meet an optimization goal

$$\mathcal{L} = I - \gamma D - \lambda \Psi \quad (4)$$

where  $\mathcal{L}$  is the optimization goal,  $I$  is the information stored by the neuron,  $D$  is the distance of the AP generation rate from some target rate, and  $\Psi$  is an energy cost term for maintenance of large synaptic inputs. They find that a gradient-ascent rule for approaching the optimization goal accurately predicts the existence of synaptic weight modification. Synaptic weight distributions are generated after STDP simulation which show input selectivity and match some of the other weight distributions predicted by different methods or measured experimentally.

In [24], the authors develop a Shannon-information model where the signal to noise ratio and volumetric ratio are exponentially proportional to each other. There is a body of neuroscience research which indicates that synaptic volume is both proportional to the size of gain  $w_i$  for that synapse

as well as the energy costs of that synapse. They treat the function of the neuron as a memory storage device, where communication occurs over time. Information is fed into the system at one point in time, and retrieved at a later point. They characterize the impact of noise associated with this process, accounting for both write- and read-time sources of noise. Their results indicate that large quantities of small, noisy synapses are preferred to singular, larger synapses. This is support by experimental findings which indicate some pre-synaptic neurons make numerous connections with post-synaptic targets at a variety of locations in the dendritic tree.

#### D. Dendritic Tree Morphology & Computation

Much of the aforementioned research treats the neuron as a perceptron with linear summation of dendritic inputs. It also largely ignores the effect and importance of IPSPs. More recent research has also shown that in fact the basic dendritic functionality is far richer than initially thought [11]. The dendritic tree is now believed to be a significant contributor to the computational abilities of the neuron [25], and is theorized to be responsible for a number of neural functionalities.

As depicted in figure 6, the authors propose that the neuronal system should be modeled as a two-layer neural integration and computational system. Each dendritic branch performs it's own computation, defined mathematically as

$$b_i = s(f(d_{i,1}, \dots, d_{i,n})) \quad (5)$$

where  $b_i$ ,  $s(-)$ ,  $f(-)$ , and  $d_{i,j}$  are the output of the  $i^{th}$  branch, sigmoidal amplification function, dendritic arithmetic function, and  $j^{th}$  dendritic input on the  $i^{th}$  branch, respectively. The function  $s$  is a sigmoidal function, which determines whether the amplifier for that branch will indeed coherently amplify-and-forward the signal. The function  $f$  determines how each set of branch inputs sums together, although inter-branch summation seems to be nearly linear.

Additionally, the dendritic tree has a number of filtering capabilities which can impact the computational ability of the neuron. A single sharp EPSP spike in the dendritic tree is subject to low-pass filtering effects through the tree's morphology, and it's energy is spread at the soma. EPSPs from different branches can last for differing lengths of time at differing peak amplitudes based upon their location in the tree. In [26], a recently developed transmission cable equation for the dendritic tree is laid out. It is the first to include some of the higher-frequency low pass characteristics of the neuron. In sub-threshold (i.e. linear) regions, this cable equation matches well with experimental measurements.

Some of the newer models are also attempting to account for the impact of IPSPs on the computational ability of the neuron, though these research efforts are in their infancy. In [6], [7], the authors develop arithmetic models for EPSP and EPSP/IPSP summation, respectively. A model which incorporates all of these results into a single model has not yet been developed.

#### IV. SYSTEM MODEL

The main goal of the system modeling in this paper is two-fold:

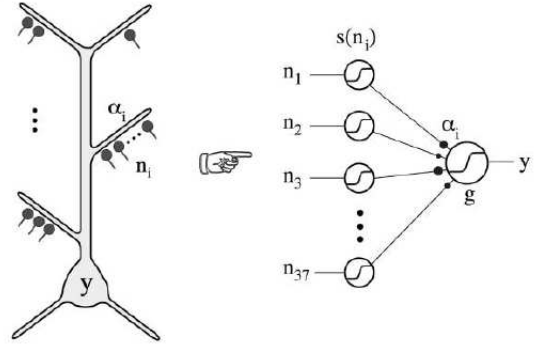


Fig. 6. Single Neuron Computation [8]

- 1) To make an argument for the STDP-covariance relationship using simple communication theory techniques (maximization of Shannon AWGN rate)
- 2) To seek a link between observed synaptic weight distributions, STDP, and input covariance

Additionally, we also briefly consider a coherent selection-amplify-and-forward relaying scheme and it's relationship with dendritic computation. This model is of interest for future work.

#### A. Modified Multiple-Access Channel Model

The biological STDP mechanism depends upon correlation between the input and output to determine synaptic weight modification. When output follows input, synaptic gain is increased, and vice-versa for input-follows-output. Since the output is dependent (non-linearly) on the set of inputs, we know that the synaptic weight is ultimately dependent upon covariance between the inputs. Many neural networks follow simple Hebbian rules based upon these concepts of input covariance [27]. Here, we develop a basic communication model for the perceptron-neuron model which supports this idea.

We begin by defining the perceptron equation as

$$Y = \Theta \left( \sum_{i=1}^n w_i X_i - \theta \right) = \Theta (w^T x - \theta) \quad (6)$$

where  $\mathbf{x} = [X_1 \dots X_n]^T$  is a vector of correlated bernoulli random variables with an  $n \times 1$  mean vector  $\mu_X$  and an  $n \times n$  covariance matrix  $\Sigma_X$ , and  $\mathbf{w} = [w_1 \dots w_n]^T$  is the vector of synaptic weights. The heaviside step function is  $\Theta$ .

The set of input patterns (i.e. all possible binary combinations of the Bernoulli input vector  $\mathbf{x}$ ) are not equally likely to occur. We treat the set of input patterns for an  $n$ -input perceptron as a codebook  $\mathcal{C}$  which takes values between decimal 0 and  $2^n - 1$ . The probability of a given code  $\mathcal{C}_i$  is a function of both the mean and covariance matrices. For these systems, the Bernoulli mean  $\mu$  for a given input dominates the correlation between two inputs, so to a first approximation we estimate that the probability of a given code is the product of the probabilities for each individual binary value in that code. Mathematically this is defined as

$$P(c = C_i) = \prod_{k=1}^n P(x_k = C_i(k)) \quad (7)$$

where  $C_i(k)$  is the  $k^{\text{th}}$  binary element of the code  $C_i$ . For instance, for a 4-input perceptron with pattern 0110, the likelihood of that code is  $P(X_4 = 0) * P(X_3 = 1) * P(X_2 = 1) * P(X_1 = 0)$ .

There has been some work in defining probability distributions for sums of correlated bernoulli random variables [28], but these distributions require complex evaluation and do not necessarily represent the actual distribution well. In numerical simulation, we do not find large deviations in codebook probability from the approximation suggested here. We considered numerical monte-carlo simulations for all modeling, but determined that significant, time-consuming runs would be required to properly characterize the input code probability distribution.

We treat the axon hillock as a decoder which compares the synaptic summation in the soma against the threshold  $\theta$ . For a single input pattern  $C_i$ , only the 1-valued inputs contribute to the output signal. Thus, the signal power in a given code  $C_i$  is then

$$S_{C_i} = \sum_{x \in C_i} (x_i w_i)^2 = \sum_{x_i=1} w_i^2 \quad (8)$$

Additionally, we define the constant average power in the soma as

$$S_I = E \left[ \sum_{i=1}^n (w_i x_i)^2 \right] = w^T \Sigma_x w + w^T \text{diag}(\mu_x^2) w \quad (9)$$

We utilize the definition of the expected value of a sum of random variables to expand it into covariance and mean components. Thus, for a given code, we can define its signal-to-interference ratio (SIR) as

$$\gamma_i \approx \frac{\sum_{x \in C_n} (x_i w_i)^2}{w^T (\Sigma_x + \text{diag}(\mu_x^2)) w} \quad (10)$$

It is clear that this under-estimates the SIR  $\gamma_i$  for a given code, because the interference term includes the average contribution of the signal from that code. But as we will see, this greatly simplifies analysis of our system model because the denominator in the SIR ratio is constant across all input codes. We justify this approximation by noting that the probability of a non-sparse code is low, so the impact on the SNR for the most probable (sparse) codes is not significant (i.e.  $a/(a+b) \approx a/b$  when  $b \gg a$ ).

We utilize the Shannon-information rate equation for an AWGN channel, which is a commonly utilized equation in many of the aforementioned neuron/information theoretic research efforts. It is defined as follows:

$$R(\gamma) = \log_2(1 + \gamma) \quad (11)$$

and for  $\gamma < 1$ , we make the approximation that

$$R(\gamma) \approx \gamma, \quad 0 < \gamma < 1 \quad (12)$$

We treat each code as a user of this dendritic channel, and propose that at a given time, a single code is attempting to reach the output. At all times, the underlying expected firing rates of every input contribute to stochastic interference in the channel. We ignore noise from other sources, at this point. With this simplification, we can say that the total sum-rate across all codes for the perceptron is given as

$$R_{sum} = \sum_{i=1}^n Pr(C_i) \gamma_i \approx \frac{w^T \text{diag}(\mu_x) w}{w^T (\Sigma_x + \text{diag}(\mu_x^2)) w} \quad (13)$$

The numerator in the RHS can be explained as follows: Since we have estimated the probability of a given code to be the product of it's bernoulli RV constituents,  $Pr(x_i = 1 \in C) = Pr(x_i = 1) = \mu_{x_i}$ . The expression in 13 simplifies the RHS after we sum over all code probabilities. We theorize that the neuron seeks to maximize this measure of information rate. Thus, we can pose the synaptic-weight selection as an optimization routine. We are interested in

$$\max_w R_{sum}(w) \quad (14)$$

subject to no constraints, at this point.

The equation in 13 is a ratio of two quadratic forms. The matrix in the denominator is the sum of two PSD matrices, so it is invertible and eigen-decomposable. We expand the denominator quadratic as

$$w^T (\Sigma_x + \text{diag}(\mu_x^2)) w = w^T \bar{\Sigma}_x w = w^T Q_{\bar{\Sigma}} \Lambda_{\bar{\Sigma}} Q_{\bar{\Sigma}}^T w \quad (15)$$

Next, we define an intermediate variable  $z$ , such that

$$z = \Lambda_{\bar{\Sigma}}^{\frac{1}{2}} Q_{\bar{\Sigma}}^T w \quad (16)$$

Thus, we can reformulate the ratio of two quadratic forms in eq. 13 as

$$R_{sum} = \frac{w^T \text{diag}(\mu_x) w}{w^T (\Sigma_x + \text{diag}(\mu_x^2)) w} = \frac{z^T \Lambda^{\frac{-1}{2}} Q^T \text{diag}(\mu_x) Q \Lambda^{\frac{-1}{2}} z}{z^T z} \quad (17)$$

This problem is a readily solvable quadratic form. The vector  $z_{max}$  which maximizes the above form is the first eigenvector of  $\Lambda^{\frac{-1}{2}} Q^T \text{diag}(\mu_x) Q \Lambda^{\frac{-1}{2}}$ , and thus, the maximized  $w$  is

$$w_{max} = Q \Lambda^{\frac{-1}{2}} z_{max} \quad (18)$$

It is clear that this maximizing vector is closely related to the maximal eigenvector of the correlation matrix. Additionally, if the matrix  $\Lambda^{\frac{-1}{2}} Q^T \text{diag}(\mu_x) Q \Lambda^{\frac{-1}{2}}$  has strictly positive entries, the Perron-Frobenius theorem states that the maximum eigenvector should have only positive components. We will see that in many cases, this is true. Since this is often the case, we do not impose non-negativity bounds on the components of the optimizing weight vector. It would also be useful to impose a maximal bound on the weight vector components, but this requires significant complication of the solution since it is generally not a convex problem.

There are a number of biological constraints which we might place on this maximization problem. One discussed constraint is non-negativity and maximum values on the components of the weight vector  $w$ . Another biologically plausible constraint is that either the total synaptic weight or the total power in synaptic weight is constrained to a maximum. In [29], the authors find that STDP should lead to a conservation of the total synaptic weight.

### B. Dendritic Selection Amplify-and-Forward

Recent experimental results have shown that dendritic morphology contributes to the I/O function of a neuron, and that this greatly enhances the computational abilities of the neuron. Some of the research indicates that within-branch summation is nearly linear, and then subject to a sigmoidal amplification function which coherently adds gain to the signal if it exceeds some threshold. Other research indicates that the sigmoidal amplification function itself is subject to STDP-like rules, where the threshold for a given branch is decreased when that branch contributes to an AP.

Researchers propose that the branch and/or synaptic inputs become aware of their contribution via back-propagating action potentials (BPAPs), which arise because the AP generated at the soma propagates both down the axon and back up into the dendritic tree. We compare this to a signal ACKnowledgement mechanism. It is believed that there are dendritic sites which contribute to amplifying this acknowledgement signal as well. Also, research and analysis shows that the communication channel is asymmetric, such that signals generated in the soma and propagated to the dendrites are less-attenuated than signals heading in the opposite direction. This is due to properties of the channel and the effect of branching in the dendritic tree.

We treat these interactions as a distributed sensor network, which contains selection AF relaying nodes that sum signals from a variety of constituent nodes. If the signals generated by the constituents exceed some threshold, the selection AF node relays the information to a central node. When the central node deems that the relayed information was important, it provides an acknowledgement signal to the selection AF node, which decreases its threshold for relaying.

We leave further modeling and simulation of this setup for future work.

## V. RESULTS

In simulating this model, we aim to show the dependence of the maximizing weight vector upon the correlation matrix  $\Sigma$ . Additionally, we aim to be able to duplicate some of the results discussed above, with regards to optimal and observed weight distributions as well as STDP simulations and experiments.

### A. Experiment 1: Selection of Correlated Inputs

One of the interesting results predicted by STDP in the perceptron model is that un-correlated synaptic inputs will be treated as noise. Over time, these inputs will be stochastically driven to zero weight since they're unlikely to result in the

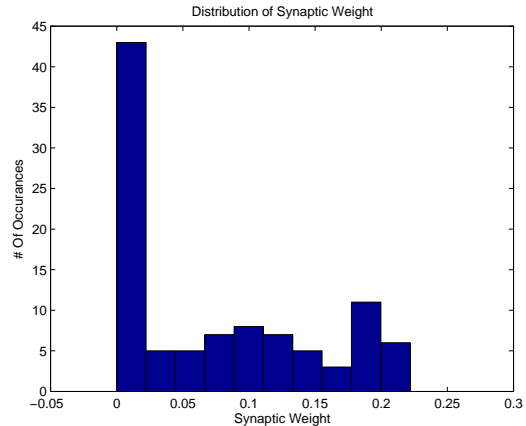


Fig. 7. Experiment 1: Synaptic Weight Distribution

generation of an AP. The first test of our system model is to observe the effects of correlation on the maximal weight vector.

For a system with 100 synaptic inputs, we choose  $E[X_i] = 0.2$ , and randomly choose a subset approximately 50% of the inputs to be correlated with each other. We begin by utilizing the methodology in [15], whereby each of the correlated inputs is assigned a correlation factor  $c_i$ , and the correlation between two inputs  $X_i$  and  $X_j$  is  $c_i c_j$ . This effectively correlates a single family of inputs with each other, since for any input with  $c_i = 0$ , its correlation with every other input is zero.

After random selection of a correlated input sub-family, we select correlation factors  $c$  randomly from a random uniform distribution  $U[0,0.2]$ . We plot a histogram of weights in figure 7. There were 44 zero-valued correlation factors in this run, which is matched by the histogram above. The synaptic weight for every input with a zero-valued correlation factor is driven to zero, which supports basic assumptions of STDP.

Note that this distribution bears close resemblance some of the predicted, STDP-simulation based, and experimentally measured synaptic weight distributions.

### B. Experiment 2: STDP

The STDP experiment in [15] was set up as follows:

- $N = 1000$  synapses
- correlation 'factor'  $c_a = 0.2 \frac{(a-1)}{(N-1)}$
- Average firing rate per synapse constant across all synapses
- $w_i$  constrained to be  $0 \leq w_i \leq w_{max}$

We utilize our function maximization stated in equation 17, noting that a maximum weight constraint is not possible with simple evaluation of this quadratic form, and define the off-diagonal elements of  $\Sigma = cc^T$ , and the diagonal elements as  $\mu(1 - \mu)$ . Our problem is formed slightly differently, since we've treated each input as a bernoulli variable with no concept of time. None the less, the basic idea of the maximization is the same, and the correlation is constant across both problems.

In figure 8, we plot the correlation factor vs. the optimal synaptic weight and find that it resembles that which was

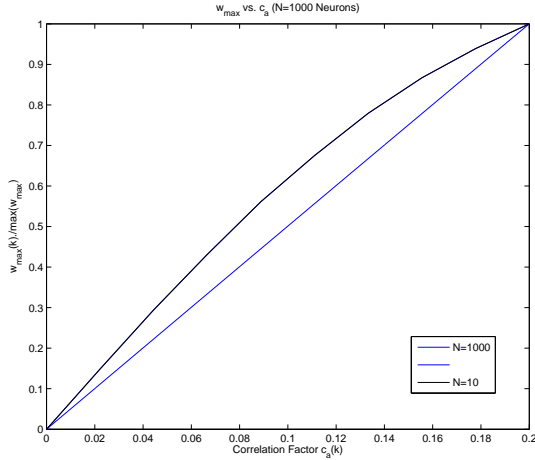


Fig. 8. Experiment 2: Synaptic Weight vs. Correlation Factor, Linear (N=10,1000)

found in the STDP simulation shown in figure 4. For small number of neurons ( $N=10$ ), we see a similar curvature to the weight distribution. At large  $N$  ( $N=1000$ ), this curvature is no longer present. We find a roughly linear dependence upon the synaptic weight as opposed to their model for large  $N$ , which saturates values of  $w_i$  at some maximum and minimum limit. It is possible that with a maximal synaptic value constraint imposed on our model, the maximizing weight vector would more closely resemble that which was found in the STDP simulation.

Unfortunately, this ratio of quadratics is not necessarily convex and requires complex optimization routines. An approach for this is detailed in [30], where a branch-and-bound optimization method is used to find the optimum value of the quadratic ratio under non-negativity and linear constraints. Utilizing this extension is left for future work. None the less, these initial results are encouraging.

### C. Experiment 3: Matching the Optimal Distribution

We note that there is nearly a linear dependance between  $c_a$  and  $w_{max}$ , which is expected due to the linked dependence between the cross-correlation matrix  $\Sigma$ , the pairwise correlation factor vector  $c$ , and the resultant weight vector  $w_{max}$ . As such, experimentation led to the choice of a logarithmic distribution of correlation factors for the vector  $c$ . Additionally, some questionable research indicates that cross-correlation between inputs is log-normally distributed by nature. We plot the results of a run with this input in figure 9. Note that this logarithmic distribution closely matches the optimal and experimental weight distributions discussed in previous sections.

Our model suggests that a lognormally distributed correlation factor would result in an optimal weight vector which was also lognormally distributed. This would account for the significant number of 'silent' or very-small valued synapses that are present both in optimality models and experimentally measured data. Better experimental results measuring correlation between inputs would be of use to further this idea.

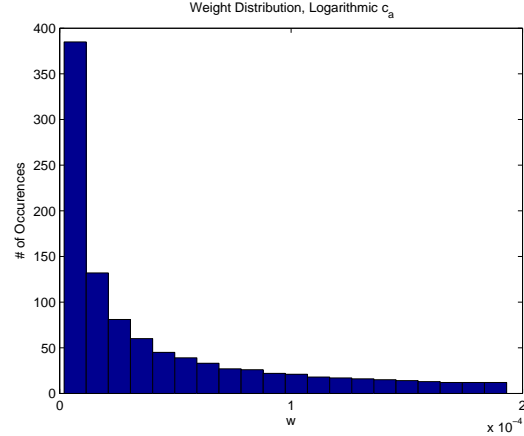


Fig. 9. Experiment 3: Synaptic Weight Distribution with lognormal correlation factors ( $N=1000$ )

## VI. FUTURE WORK

The first step in future work would be to rigorously mathematically derive a tight-knit link between expected cross-correlation between inputs and the optimal weight distribution models which have already been derived. It is possible that our model could serve as that link with a more rigorous mathematical treatment. We discuss an alternative approach which could also possibly allow for the same link.

Our basic system models neuron functionality as a communications system with many users (codes) and a single destination (the soma). There are alternative system formulations which might be of interest for future work. We briefly explored the idea of a MISO channel model, where each synaptic input is part of the gain vector  $h$  and the input covariance  $Q$  is treated as above. Normally, when treating MISO channel models, the gain is assumed to take a certain distribution, and the input covariance matrix is setup as a gaussian codebook.

For our problem, this formulation would be slightly different. We would consider the case where the gain matrix distribution is known (as suggested by the optimal distribution discussed herein), and then attempt to determine if an optimal distribution for the covariance matrix could be found to maximize capacity.

In [31], aspects of correlated channels, interference, and signals are considered for the MISO channel. It is plausible that some form of waterfilling is taking place, which would support some of the experimental and theoretical findings with regards to the relationship between input correlation and synaptic weight allocation. Experimental support would be required to evaluate if the capacity-maximizing distribution for input covariance matches that which is found biologically. This is a difficult experimental task due to the small physical size of dendritic trees, challenges of dendritic imaging in vitro, and the small EPSP signals.

We would also like to be able to incorporate effects of dendritic morphology in future system models. It is unclear how this would take place with a MISO model, or system model derived in this paper.

Another interesting model, as discussed above, yields a

sensor-network with selection-amplify and forward nodes following. Each dendritic branch functions as a selection amplify-and-forward (AF) node, by taking the input signals and then deciding if they exceed a given threshold. If they do exceed the threshold, the signal is amplified and forwarded to the collection node (the soma). Including the effects of potentiation and depression [9] leads to a changing threshold for selection relaying. Modeling this system and considering its practical use is left for future work.

## VII. CONCLUSIONS

In this paper, we have provided a brief overview of neuronal function. We have reviewed in detail some of the more interesting, recent findings with respect to communications theory. It seems that the neuronal communication system does employ ideas that are related to those utilized in current communications systems.

We have successfully demonstrated that a simple model for a neurons functionality can be captured through maximization of the Shannon information rate. This model captures the importance of cross-correlation between synaptic inputs, as well as the effect of mean firing rate. Input correlation being a key contributor to synaptic weight distributions is a widely held idea. In future work, it is clear that this model would have to be extended to include different plausible maximization constraints on the total power, total synaptic weight, and/or non-negativity of the synaptic weights.

We have introduced some additional system modeling concepts, but not treated them in detail. We plan to continue investigating these topics.

## REFERENCES

- [1] A. Hodgkin and A. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve (reprint)," *Bulletin of Mathematical Biology*, vol. 52, no. 1-2, pp. 25 – 71, 1990. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WC7-4GP2506-4/2/8a68ac710fc6dd933c684747f30d96f3>
- [2] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain (Book with CD-ROM)*. Lippincott Williams & Wilkins, March 2002.
- [3] D. Attwell and S. B. Laughlin, "An energy budget for signaling in the grey matter of the brain," *J Cereb Blood Flow Metab*, vol. 21, pp. 1133–1145. [Online]. Available: <http://dx.doi.org/10.1097/00004647-200110000-00001>
- [4] L. F. Abbott, J. A. Varela, K. Sen, and S. B. Nelson, "Synaptic Depression and Cortical Gain Control," *Science*, vol. 275, no. 5297, pp. 221–224, 1997. [Online]. Available: <http://www.sciencemag.org/cgi/content/abstract/275/5297/221>
- [5] A. A. Faisal, J. A. White, and S. B. Laughlin, "Ion-channel noise places limits on the miniaturization of the brain's wiring," *Current Biology*, vol. 15, no. 12, pp. 1143 – 1149, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/B6VRT-4GFD19F-13/2/83968e524bdcf04b3d1c82c0c7711187>
- [6] P. Poirazi, T. Brannon, and B. W. Mel, "Arithmetic of subthreshold synaptic summation in a model cal pyramidal cell," *Neuron*, vol. 37, no. 6, pp. 977 – 987, 2003. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WSS-488M518-C/2/4d8b0a26fb22cc6d9828bfa0cbe57ae7>
- [7] J. Haoa, X. dong Wanga, Y. Danb, M. ming Poo, and X. hui Zhanga, "An arithmetic rule for spatial summation of excitatory and inhibitory inputs in pyramidal neurons," *PNAS*, 2009.
- [8] P. Poirazi, T. Brannon, and B. W. Mel, "Pyramidal neuron as two-layer neural network," *Neuron*, vol. 37, no. 6, pp. 989 – 999, 2003. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WSS-4CVXF15-1/2/64d4ca3eaa0bc4f91086aa23ebd40a60>
- [9] Z. Wang, N. long Xu, C. ping Wu, S. Duan, and M. ming Poo, "Bidirectional changes in spatial dendritic integration accompanying long-term synaptic modifications," *Neuron*, vol. 37, no. 3, pp. 463 – 472, 2003.
- [10] N. Spruston, "Pyramidal neurons: dendritic structure and synaptic integration," *Nature reviews. Neuroscience*, vol. 9, no. 3, pp. 206–221, March 2008. [Online]. Available: <http://dx.doi.org/10.1038/nrn2286>
- [11] A. Polsky, B. W. Mel, and J. Schiller, "Computational subunits in thin dendrites of pyramidal cells," *Nat Neurosci*, vol. 7, no. 6, pp. 621–627, June 2004. [Online]. Available: <http://dx.doi.org/10.1038/nn1253>
- [12] D. Hebb, *The organization of behavior*. New York: Wiley, 1949.
- [13] L. F. Abbott and W. G. Regehr, "Synaptic computation," *Nature*, vol. 431, pp. 796–803. [Online]. Available: <http://dx.doi.org/10.1038/nature03010>
- [14] J. A. Varela, K. Sen, J. Gibson, J. Fost, L. F. Abbott, and S. B. Nelson, "A Quantitative Description of Short-Term Plasticity at Excitatory Synapses in Layer 2/3 of Rat Primary Visual Cortex," *J. Neurosci.*, vol. 17, no. 20, pp. 7926–7940, 1997. [Online]. Available: <http://www.jneurosci.org/cgi/content/abstract/17/20/7926>
- [15] S. Song, K. D. Miller, and L. F. Abbott, "Competitive hebbian learning through spike-timing-dependent synaptic plasticity," *Nature*, 2000.
- [16] B. Barbour, N. Brunel, V. Hakim, and J. P. Nadal, "What can we learn from synaptic weight distributions?" *Trends in neurosciences*, vol. 30, no. 12, pp. 622–629, Dec 2007, jID: 7808616; RF: 64; 2007/07/25 [received]; 2007/09/25 [revised]; 2007/09/26 [accepted]; 2007/11/05 [aheadofprint]; ppublish.
- [17] N. Brunel, V. Hakim, P. Isope, J. P. Nadal, and B. Barbour, "Optimal information storage and the distribution of synaptic weights: perceptron versus purkinje cell," *Neuron*, vol. 43, no. 5, pp. 745–757, Sep 2 2004, IR: 20061115; JID: 8809320; 2004/02/03 [received]; 2004/07/21 [revised]; 2004/08/16 [accepted]; ppublish.
- [18] N. Brunel, J. Nadal, and P. Toulouse, "Information capacity of a perceptron," *J. Phys. A: Math. Gen.*, vol. 25, 1992.
- [19] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *Electronic Computers, IEEE Transactions on*, vol. EC-14, no. 3, pp. 326–334, 1965. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4038448](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4038448)
- [20] E. Gardner, H. Gutfreund, and I. Yekutieli, "The phase space of interactions in neural network models," *Journal of Physics*, vol. 21, pp. 257–270, 1988.
- [21] A. B. Barrett and M. C. W. van Rossum, "Optimal learning rules for discrete synapses," *PLoS Comput Biol*, vol. 4, no. 11, p. e1000230, 11 2008. [Online]. Available: <http://dx.doi.org/10.1371/journal.pcbi.1000230>
- [22] J. Porrill and P. Dean, "Silent synapses, ltp, and the indirect parallel-fibre pathway: computational consequences of optimal cerebellar noise-processing," *PLoS computational biology*, vol. 4, no. 5, p. e1000085, May 23 2008.
- [23] T. Toyozumi, J. P. Pfister, K. Aihara, and W. Gerstner, "Optimality model of unsupervised spike-timing-dependent plasticity: synaptic memory and weight distribution," *Neural computation*, vol. 19, no. 3, pp. 639–671, Mar 2007, jID: 9426182; ppublish.
- [24] L. R. Varshney, P. J. Sjöström, and D. B. Chklovskii, "Optimal information storage in noisy synapses under resource constraints," *Neuron*, vol. 52, no. 3, pp. 409 – 423, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WSS-4M99JHX-5/2/d4778722069f9f4932827382b01c123f>
- [25] M. London and M. Hausser, "Dendritic computation," *Annu. Rev. Neurosci.*, vol. 28, pp. 503–532, 2005.
- [26] C. Bădărd and A. Destexhe, "A modified cable formalism for modeling neuronal membranes at high frequencies," *Biophysical Journal*, vol. 94, no. 4, pp. 1133 – 1143, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/B94RW-4V0KNMT-3/2/a9bc86901f36457e7cff29b969641190>
- [27] K. D. Miller and D. J. MacKay, "The role of constraints in hebbian learning," *Neural Computation*, vol. 6, pp. 100–126, 1994.
- [28] D. H. M. II, "Evaluation of five discrimination procedures for binary variables," *Journal of the American Statistical Association*, vol. 68, no. 342, pp. 399–404, 1973. [Online]. Available: <http://www.jstor.org/stable/2284083>
- [29] S. Royer and D. Pare, "Conservation of total synaptic weight through balanced synaptic depression and potentiation," *Nature*, vol. 422, pp. 518–522. [Online]. Available: <http://dx.doi.org/10.1038/nature01530>
- [30] J.-Y. Gotoh and H. Konno, "Maximization of the ratio of two convex quadratic functions over a polytope," *Computational*

*Optimization and Applications*, vol. 20. [Online]. Available: <http://dx.doi.org/10.1023/A:1011219422283>

- [31] D.-W. Yue, Q. T. Zhang, and S. H. Song, "Capacity of correlated miso channels with correlated co-channel interference and noise," *Wireless Personal Communications*. [Online]. Available: <http://www.springerlink.com/content/lu211g6452h53425>