A Simple Model for Short-Term Memory: Graded Persistent Firing in an Autapse

May 16, 2011

MAT 351 - Final Project

1 Introduction

Recordings have shown that many neurons in multiple regions in the brain of a diverse number of species, including humans, are characterized by their ability to "remember" a particular signal for many seconds of time. This *persistent firing* activity of neurons has been theorized to play an important foundational role in short-term memory (Goldman-Rakic 1995).

An example of the important role of persistent neural activity in short-term memory is evident when monkeys are subject to a delayed oculomotor task, in which they fixate at a target reticle and a visual cue is momentarily displayed a small distance away from the central point in some direction. Then after a small period of time (3-6 seconds), another cue will direct the monkey to saccade to the location of the initial visual cue. This task requires briefly storing the memory of the position of the visual cue. Single-neuron recordings are taken from neurons in the frontal eye fields and other prefrontal areas of the monkeys as they performed the task (Funahashi et al. 1989). These recordings show that neurons in prefrontal cortex are tuned to the specific directions in which the visual cue is given relative to the center point. When the visual cue disappears, the neurons tuned to the corresponding direction become active and maintain persistent firing until the memory is recalled and the saccade is made.

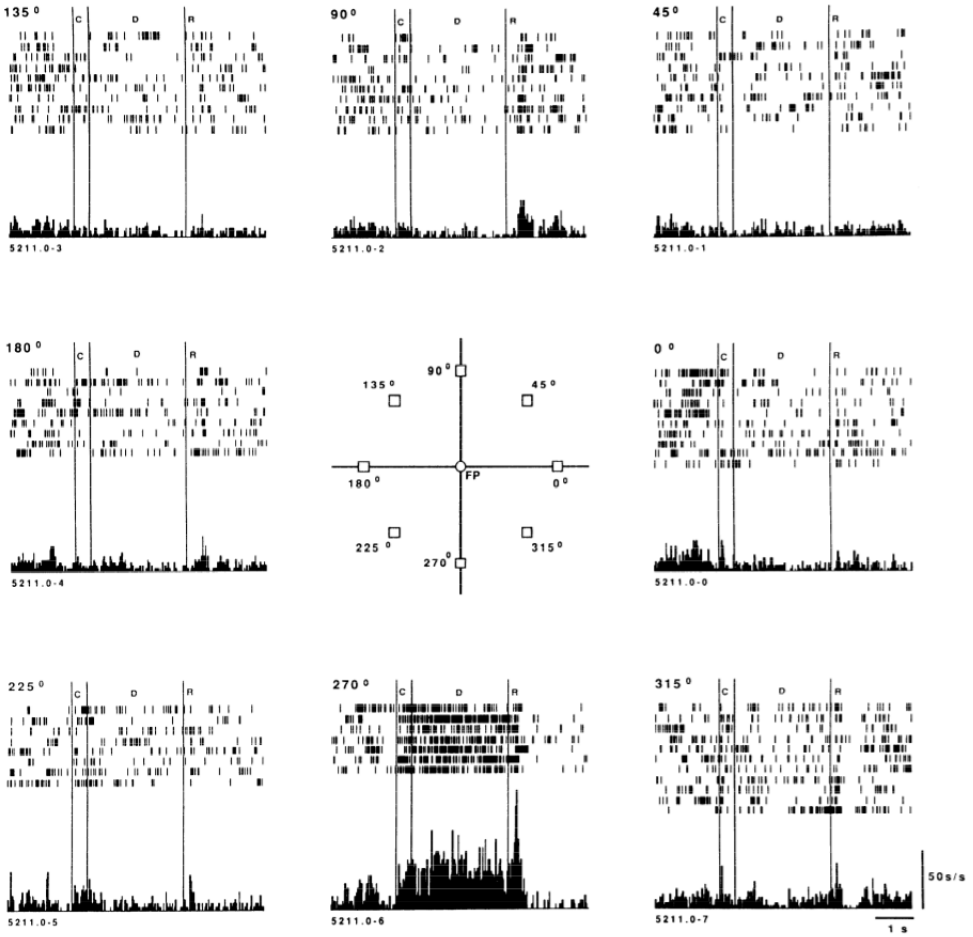


Figure 1: From Funahashi et al. 1989

The rate at which a neuron fires during this memory trace can signify how close the true direction of the stimulus is to encoded direction of the neuron, so we see a "bump" of activity, centered at the true direction of the stimulus.

Neurons that exhibit *graded persistent activity* have a continuum of stable firing rates. Depending on input, the stable firing rate of the neuron can be changed. There are neurons whose fixed firing rate can be increased with a long enough pulse of excitatory input, and similarly their firing rate can be decreased with a inhibitory pulse. With this kind of behavior, this type of neuron can act as an integrator of previous inputs. Evidence suggests that the neurons that control eye position receive input from neural integrators that exhibit this behavior (Goldman et al. 2002). We can imagine that when the subject rotates their head, input from the semicircular fluid cavities will reach these oculomotor integrators and cause the eye position to compensate and allow the position of fixation to stay contant.

A prime example of neurons that have been recorded that exhibit graded persistent activity are layer V neurons in entorhinal cortex (Egorov et al. 2002). In the figures below, we can see that the neurons behave like integrators. With every pulse of 0.3 nA depolarizing input (for 4 seconds), the stable firing rate of the neuron increased to a higher rate and persisted for about a minute before the next depolarizing pulse was given. Similarly, with every 0.4 nA hyperpolarizing pulse, the stable firing rate of the neuron stepped down to a lower rate.

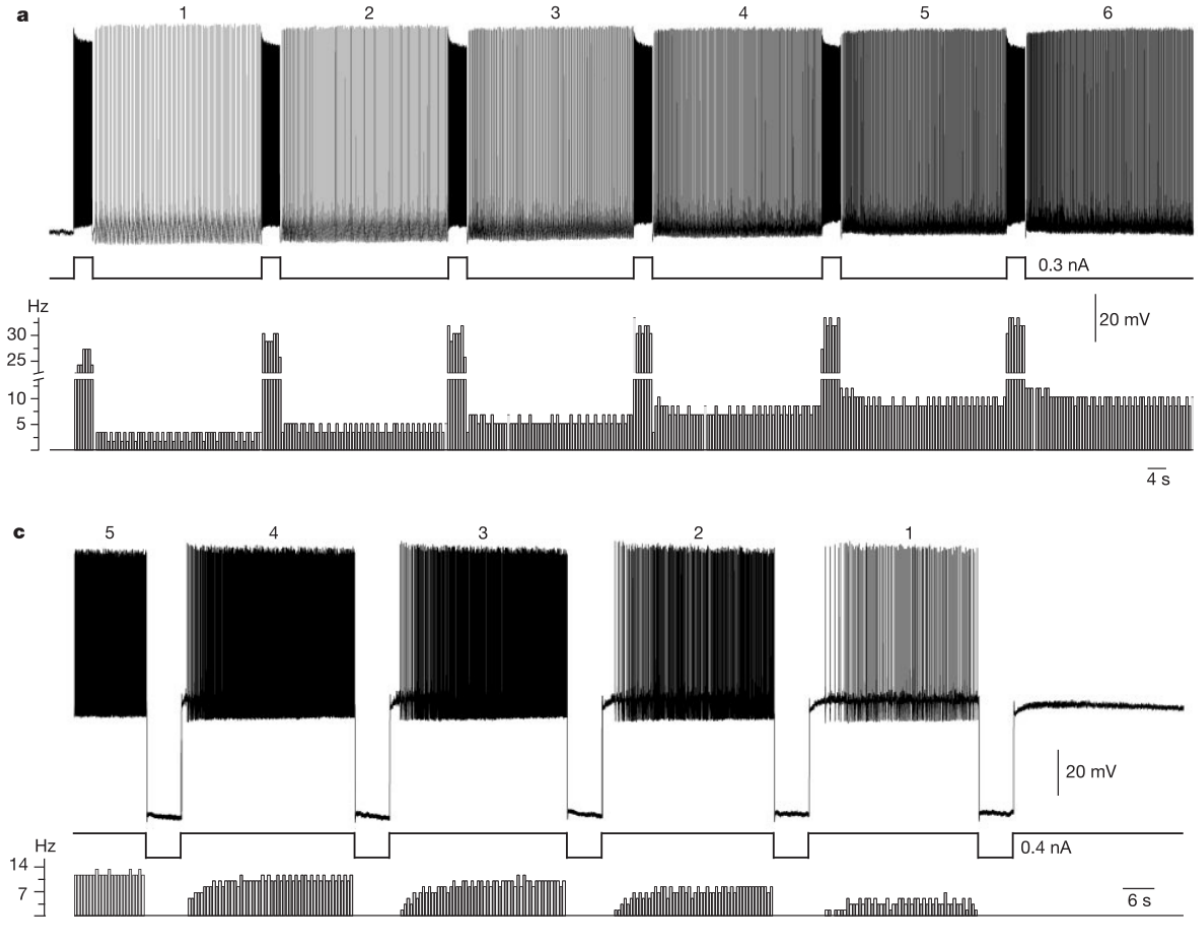


Figure 2: From Egorov et al. 2002

We will present a simple model, first devised in Seung et al. 2000, of a neuron that can exhibit graded persistent activity. A defining feature of this neuron in the model is its synaptic connection to itself: an autapse. In addition to presenting an implementation of this model, we will explore its robustness to changes in parameters, such as synaptic strength. We will present a modification of the model that is more robust to changes in these parameters.

2 The Autapse

2.1 Conductance-based model

To ensure that the model is biologically-realistic, Seung utilized a conductance-based model, closely related to the Hodgkin-Huxley equations, that accurately describes the dynamics of a firing neuron. The membrane potential V_i of neuron i evolves according to:

$$C_m \frac{dV_i}{dt} = -I_{int}(V_i, h_i, n_i, b_i) - g_{E,i}(V_i - V_E) - g_{I,i}(V_i - V_I)$$

$I_{int,i}$ is a measure of the current due to leak, sodium, delayed-rectifier potassium, and A-type potassium dynamics in neuron i. The channel variables are: h_i for sodium, n_i for delayed-rectifier potassium, and b_i for A-type potassium. g_E is a measure of conductance to excitatory inputs, and g_I is the conductance to inhibitory inputs. Thus, the term $g_E(V - V_E)$ is the current into the neuron due to synaptic input from other neurons, and similarly for the term $g_I(V - V_I)$. g_E and g_I change depending on how "activated" the synapses connecting to this neuron are. We can think of g_E as the autapse's "input." In the conductance-based model, synapses are modeled with the variable s_i, where every neuron i has one output synapse, according to the equation:

$$\tau \frac{ds_i}{dt} = -s_i + \alpha(1 - s_i)\sigma(V_i)$$

In this equation, α is a constant, and $\sigma(V)$ is the sigmoid function, defined by:

$$\sigma(V) = (1 + e^{-(V-\theta_s)/\sigma_s})^{-1}$$

s_i is the fraction of open channels at the output synapse of neuron i. Thus for an autapse, the input and output neurons are the same, and so we have that g_E relates to s_i linearly with the equation:

$$g_E = Ws + B$$

where W and B are constants determined by the strength of the synaptic inputs into the autapse.

2.2 Reduced model

To better understand the dynamics of this system, Seung used a reduced rate-model that could accurately describe the conductance-based model given a number of assumptions. The first assumption is that the

effect of the spikes on the variable s_i is not significant given a large enough time window. This is true if the synaptic time constant τ is long. In the model, τ is set to 100ms, which is similar to the time constant of the NMDA receptor. Thus, Seung defines a function $f(g_E)$ as the time-average of $\sigma(V_i)$ over an interspike interval $T(g_E)$. The voltage in this interval is a function of g_E. So we have the definition:

$$f(g_E) = \frac{1}{T(g_E)} \int_0^{T(g_E)} dt\, \sigma(V(t; g_E))$$

So given this smoothed version of $\sigma(V_i)$, we obtain an equation for s_i that is not dependent on V_i.

$$\tau \frac{ds_i}{dt} = -s_i + \alpha(1 - s_i) f(g_{E,i})$$

Thus, we have constructed a simplified model that is not dependent on the spiking of the neuron. The goal is to produce a neuron that can fire at a constant rate. If $\frac{ds_i}{dt} = 0$, then s and g_E will stay fixed, which in the reduced model, ensures that the activity of the neuron does not change. So we solve for the steady states of the system. Since g_E depends linearly on s_i, we can write it as $g_E(s_i)$. The following relation must hold in our reduced model for our system to be in a stable state:

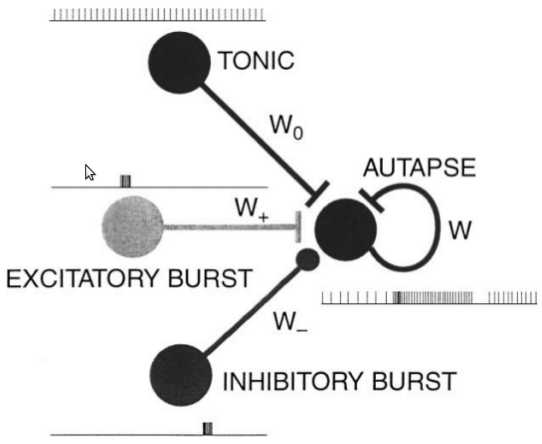
$$s_i = \frac{\alpha f(g_E(s_i))}{1 + \alpha f(g_E(s_i))} \equiv F(g_E(s))$$

This equation allows us to think of autapse as having output that depends on its input, and then the output becomes its input. We can conceptualize this by discretizing time and imagining the value of s_i at time t being input into the neuron. If our neuron is in steady state, then the value of s_i at time $t + 1$ is given by $F(g_E(s))$. Thus, solving $s = F(g_E(s))$ for the parameters W and B such that for some region of values of s, all values of s provide a solution to the equation. If we set s to a value not in the solution (with W and B fixed), for example, by adding excitatory input, then the system should move toward the steady state, and so the region should be attracting. To find W and B, the assumption that F is linear is made. Since we know that $g_E = Ws + B$, we find that $F = F_1(Ws + B) + F_0$. Thus, if we set $W = 1/F_1$ and $B = -F_0/F_1$, our linear reduced neuron will exhibit graded persistent firing at all s. We can use these values to explore what happens in the conductance-based model, as a test of using the reduced model as a tuning method.

2.3 Results

I implemented the model in Matlab. As in the Seung paper, I modeled four neurons using the conductance-based model. One is a tonic neuron that is given a constant input current of $3\,\mu A/cm^2$. Its synapse connects to the autapse (also called the memory neuron) with weight W_0 (the tuning determines this parameter from B). The second neuron is the excitatory burst neuron, that is usually silent, with no input current, is connected to the autapse, and has a excitatory connection weight W_+. The third is the inhibitory burst neuron, also usually silent, is

Figure 3: Circuit diagram of the system. From Seung et al. 2000.



also connected to the autapse with connection weight W_-. The fourth neuron is the autapse and is connected to itself with weight W. The figure to the right provides a circuit diagram.

Seung used a fourth-order Runge-Kutta method with step size 0.01ms to do the integration. To allow the simulation to run at interactive speeds, but not tradeoff too much accuracy, I used the built-in Matlab solver *ode*23 with initial step size 0.01ms and maximum step size 20ms. I also increased the error tolerances to tradeoff additional accuracy for speed. The results are still quite accurate, not differing much from that of a much slower and more accurate simulation. Firing rate was computed by counting the time steps in which the membrane voltage changed from a non-negative value to negative. A window of 500 time steps was used to compute the rate at each time step.

The initial conditions were determined by running the simulation with no synaptic activity. The system converged to a fixed point, and we used that point as the set of initial conditions for all neurons in subsequent simulations.

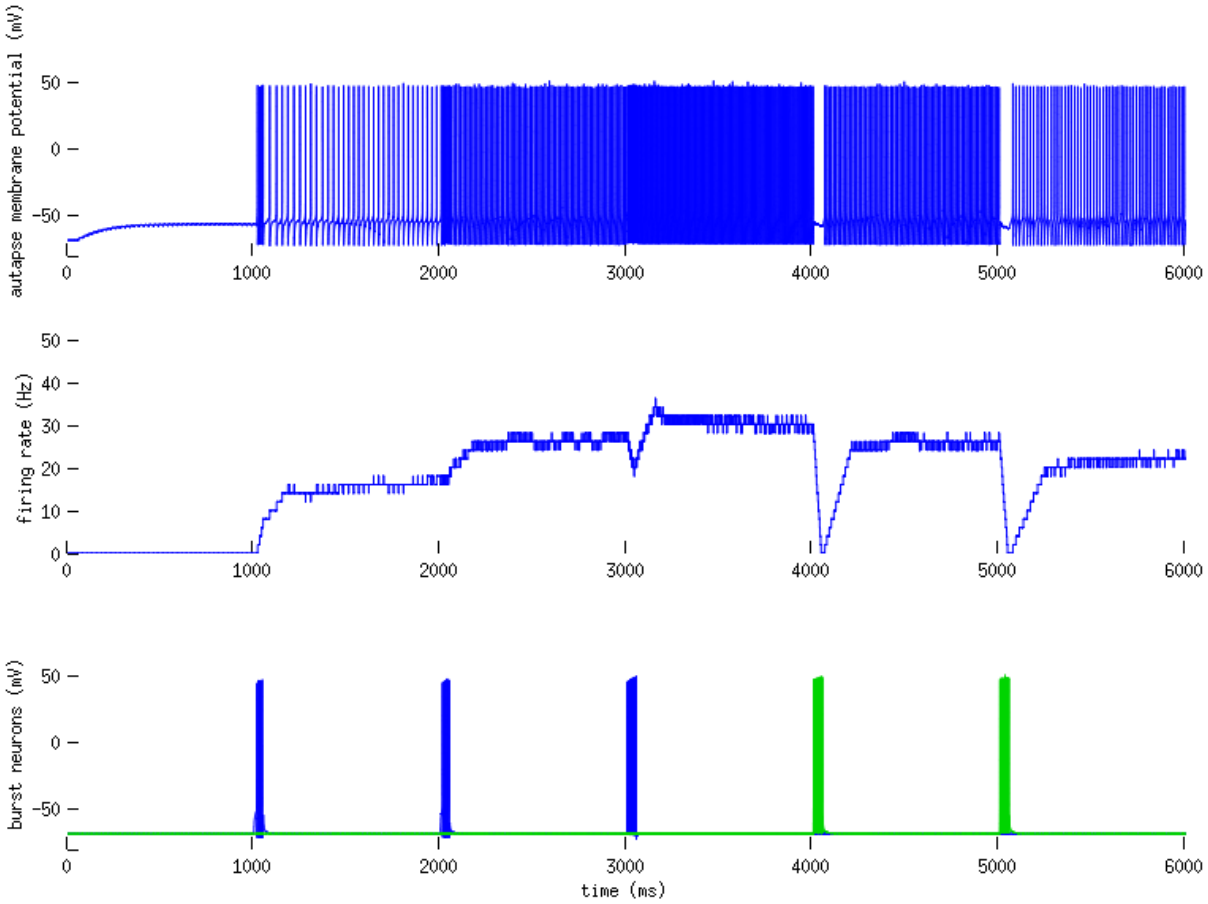


Figure 4: Six second simulation of the autapse with $W = 1.9$, $W_0 = 3.9$, $W^+ = 1.0$, $W^- = 1.6$. The top plot shows the membrane potential of the autapse in mV versus time in ms. The center plot shows the firing rate of the same neuron. The bottom plot depicts the membrane potential of the burst excitatory (in blue) and inhibitory (in green) neurons in mV.

In the simulations, the burst neurons were given 50ms bursts of current. The first burst at 1s was given only to the excitatory neuron at 6 $\mu A/cm^2$. The second burst at 2s was given to the excitatory neuron at 7 $\mu A/cm^2$. The third at 3s was given to the excitatory neuron at 15 $\mu A/cm^2$. These values were determined so that the levels of stable activity would look qualitatively distinct. A 50ms burst of current was given only to the inhibitory neuron at 4s and 5s. Both of these were at 25 $\mu A/cm^2$. The greater connection weight from the inhibitory neuron and the increased current necessary to decrease the level of firing rate is due to the fact that the synaptic activation variable s takes longer to decay than it does to increase by the same amount. We can see, between bursts of current, the memory neuron maintains a level of stable activity.

2.4 Stability Analysis

However, since our tuning procedure made assumptions that the neuron did not exactly match, we can see artifacts from the nonlinearities, such as a slow drift toward a fixed point somewhere around 30 Hz. In Figure 5, we can see a run of the simulation with the same parameters as in Figure 4, but with only a single excitatory burst. The simulation runs for 40 seconds and suggests that an attracting point exists at around 30 Hz. The persistent firing decays after about 5 seconds to this stable point.

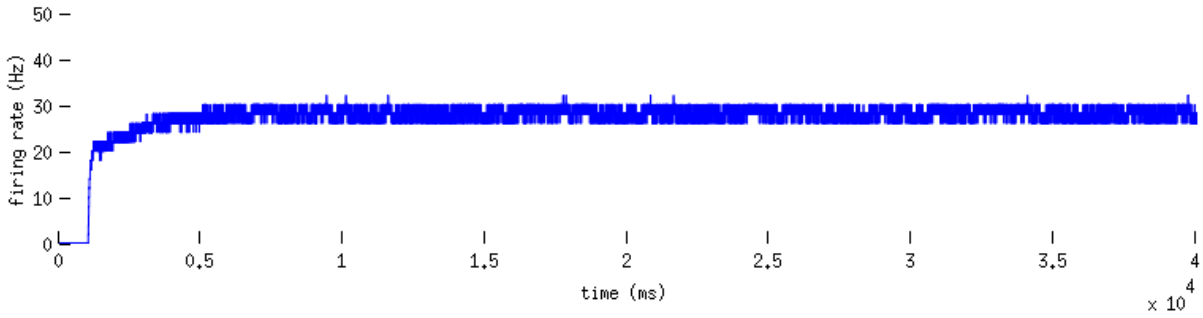


Figure 5: 40 second simulation of the autapse with $W = 1.9$, $W_0 = 3.9$, $W^+ = 1.0$, $W^- = 1.6$. At 1 second (0.1×10^4ms in the plot), the excitatory burst neuron was given a 50ms pulse of 8 $\mu A/cm^2$.

In addition, we can play with the synaptic connection strengths to measure how robust the system is. In our linear reduced model, increasing the bias would cause the firing rate to increase indefinitely with a rate proportional to the increase in bias if the system were tuned perfectly. Otherwise, the fixed point would simply move. The bias corresponds to our W_0 term, or the synaptic connection strength between the tonic neuron and the autapse. We can see in Figure 6 the effect of increasing the bias by 0.2 and decreasing it by 0.2, on the top and bottom plots respectively. We can see the neuron beginning to fire before the excitatory burst at 1s. However, the increase in firing rate seems to taper off as the system approaches the fixed point at 30 Hz. Once the neuron is given the excitatory burst at 3s, the firing rate decreases back to its stable point. The attempts to inhibit it seem futile as the neuron simply returns to its fixed point. In the second simulation, the firing of the autapse has lost its persistence. Only the excitatory burst at 3000 is enough to move the system close to the fixed point. The slowdown in the decay at this point suggests that the fixed point is still attracting. The inhibitory burst allows for the synaptic activation s to decrease, but not to the point where the neuron is silent. This causes the neuron to fire a bit more before becoming silent.

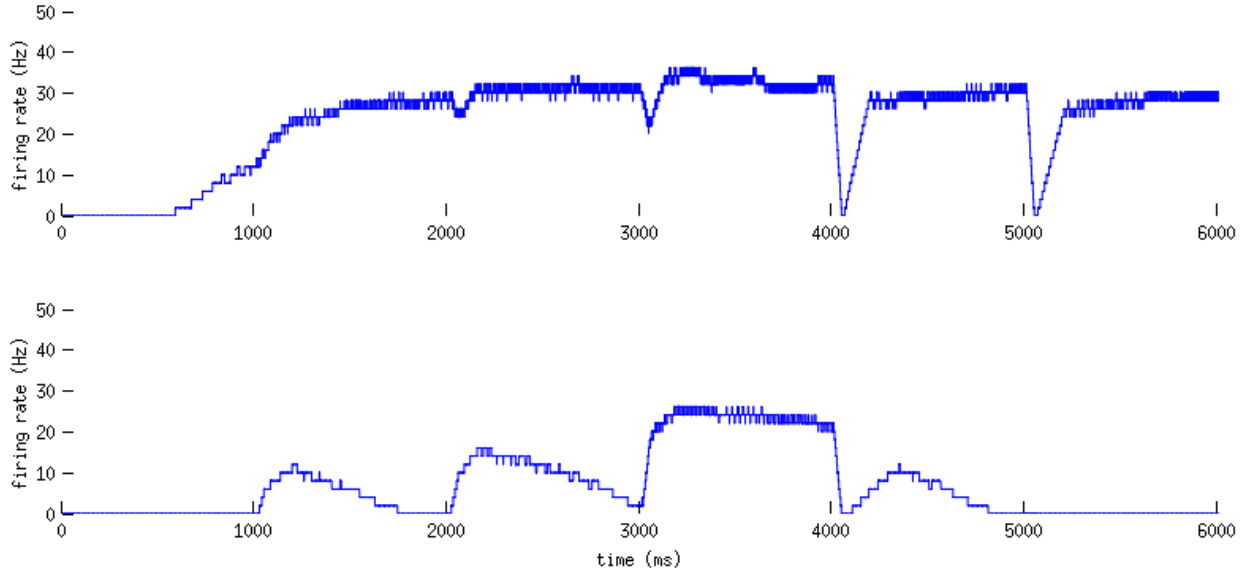


Figure 6: Two simulations of 6 seconds of the system. In both simulations, all parameters are identical to those in Figure 4, except for W_0. The upper plot depicts the case where W_0 is larger than its tuned value, $W_0 = 4.1$. In the lower plot, $W_0 = 3.7$.

The system is quite sensitive to changes in the synaptic connection strength between the tonic neuron and the autapse. Interestingly, increasing or decreasing the input current into the tonic neuron by 0.5 $\mu A/cm^2$ has a similar effect, but at a much greater extent. This is likely due to the fact that the synapse "filters" the input from the neuron into the autapse, and so changing the firing of the autapse directly will have a more significant effect on the synaptic conductance of the autapse than increasing the synaptic connection strength. Figure 7 shows the result of these tests.

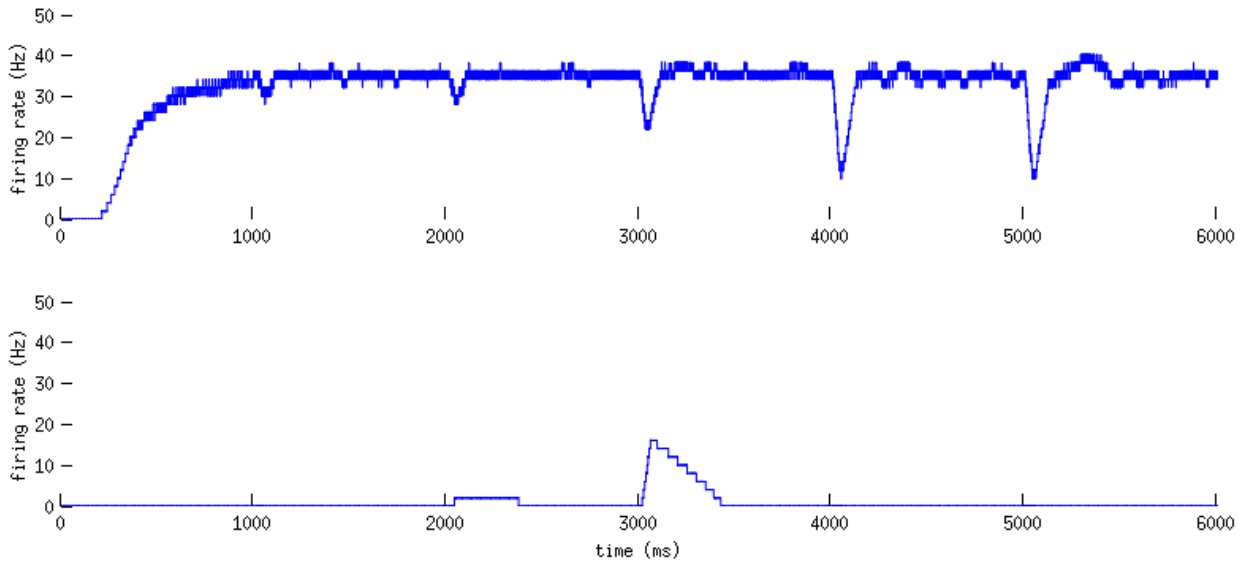


Figure 7: Two simulations of 6 seconds of the system. In both simulations, all parameters are identical to those in Figure 4, except for the input current into the tonic neuron. In the upper plot, $I_{appl} = 3.5\,\mu A/cm^2$. In the lower plot, $I_{appl} = 2.5\,\mu A/cm^2$.

Our final test of the robustness of the system perturbed the synaptic connection strength of the autapse with itself. In our linear reduced model, this would change W, and so it would either create a stable fixed point or an unstable fixed point. Figure 8 shows the simulation with an increased value of W and a decreased value in the top and bottom plots, respectively. In the top plot, we see that the fixed point has become stable (or at least *more* stable). The system is much quicker to fire at ~33 Hz. In the bottom plot, the fixed point has become unstable. Even with the third excitatory burst at 3s, the system decays more quickly, unlike the case with the decreased value of W_0. However, the system still exhibits some persistence at firing rates that are far enough away from the fixed point, interestingly.

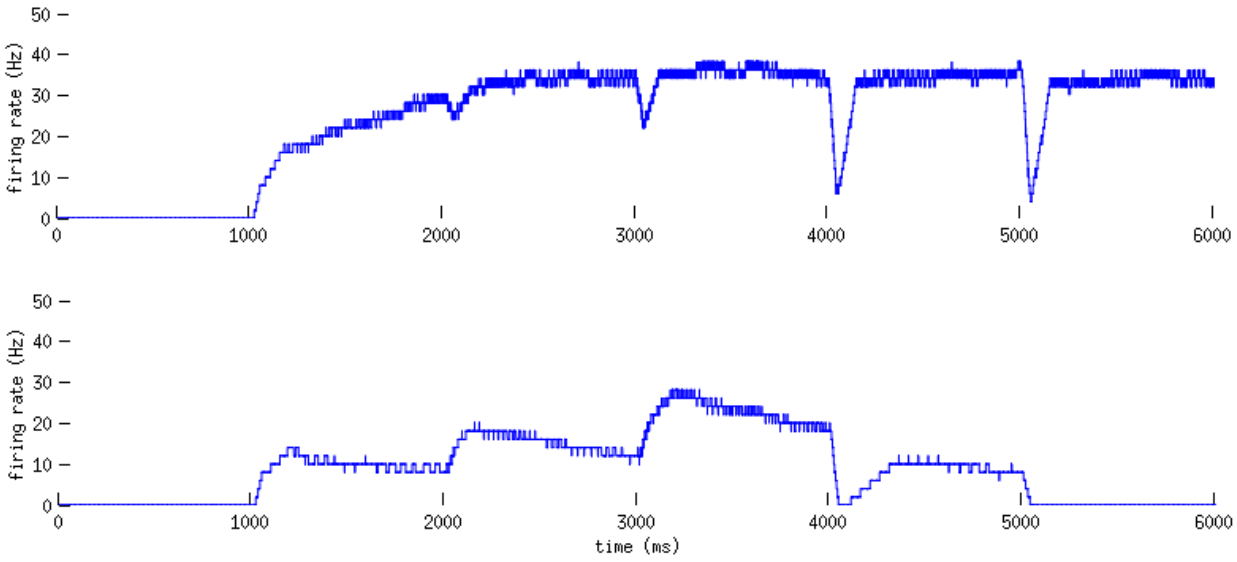


Figure 8: Two simulations of 6 seconds of the system. In both simulations, all parameters are identical to those in Figure 4, except for W. In the upper plot, $W = 2.1$. In the lower plot, $W = 1.7$.

3 Improved Autapse

3.1 New Variables

One of the problems of the model we observed was that the persistent firing behavior was very sensitive to any change in the weights. On a high level, the reason for this can be said to be a design issue. Creating a line attractor with a continuum of fixed stable points is very difficult. It is also not very biologically realistic to require hand-tuning of synaptic strengths in a model. One approach to fixing this problem is to have a stable point of activity that can move depending on input into the neuron (Teramae et al. 2005, Koulakov et al. 2002). We can accomplish this simply by making W large enough in our model. To have the position of the stable point move in the system, we need to introduce a new variable b that is directly proportional to the position of the stable point of activity. We can do this by adding a term to the synaptic conductance g_E. In the previous model, we had the relation for g_E of the autapse:

$$g_E = W s_{autapse} + W_0 s_{tonic} + W_+ s_{excitatory}$$

In our model, we simply add the term dependent on b, scaled by 0.01:

$$g_E = W s_{autapse} + W_0 s_{tonic} + W_+ s_{excitatory} + 0.01b$$

Physically, we can interpret this as a modulating factor in the self-connection of the autapse, or as a separate channel entirely. The reversal potential is the same as that of the excitatory synaptic current, which is 0 mV. We want b to change, depending on firing rate. With a long and large enough increase in firing rate, we want b to increase, and similarly for a decrease in firing rate. Thus, we created a new parameter, a, that is related to V with the equation:

$$\frac{da}{dt} = -a + exp((V_{autapse} + 20)/160)/20$$

a normally decays to 0 with time constant 20ms if the neuron is not firing. However, with every spike, there is an increase in the compound. At any constant firing rate, a also approaches a stable point at approximately $a = 0.8$ at which only the small fluctuations from the spikes are discernable. Large increases in firing rate will cause a to gravitate toward a higher point, and similarly for large decreases in the firing rate. Thus, we can relate b to a by causing b to increase if a is greater than a certain threshold. And if a is smaller than a lower threshold, then b will decrease. We have the equation:

$$\frac{db}{dt} = \begin{cases} 8(tanh(b) + \frac{1}{2})(a - a_{max\,thresh})^2 & a > a_{max\,thresh} \\ -40(tanh(b) + \frac{1}{2})(a - a_{min\,thresh})^2 & a < a_{min\,thresh} \\ 0 & otherwise \end{cases}$$

The dependence on $tanh(b)$ was added to limit the rate of change of b.

3.2 Results

With $a_{max\,thresh}$ set to 0.83 and $a_{min\,thresh}$ set to 0.78, there is a region the two thresholds in which transient bursts of input are ignored, adding an element of stability and robustness to the system. Small amounts of noise will not affect stability. We only added these two compounds to the autapse neuron, leaving other parts of the circuit unchanged. However, we did need to change the weights to ensure our fixed point would not easily turn into a line attractor or an unstable point. Thus, our input current into the tonic neuron was increased to $5\,\mu A/cm^2$. The activity of the autapse took too long to approach the stable fixed point if we kept the time constant of its synapse at 100ms. Instead, we decreased it to 5ms. The following are the synaptic connection weights: $W = 0.150$, $W_0 = 2.200$, $W_+ = W_- = 7.000$. In addition, our pulses of excitatory or inhibitory input were too small and too short. Instead, we kept the input at 40 $\mu A/cm^2$ for all bursts, and increased the duration to 100ms. With everything else being kept the same from our simulations in Figure 4, Figure 9 shows the resulting simulation.

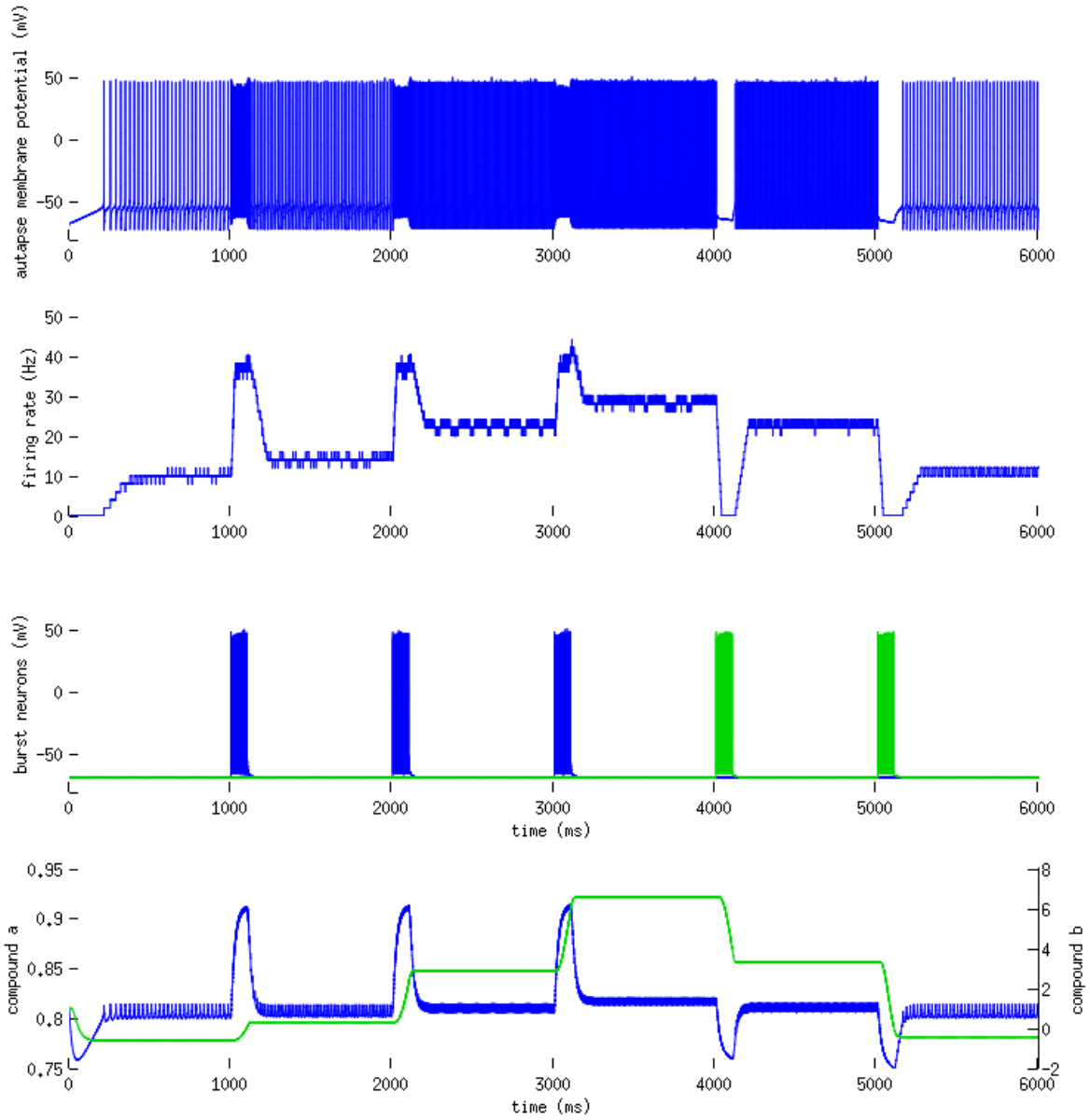


Figure 9: A six second simulation of the improved autapse model. The plots are idential to those in Figure 4. The fourth plot shows how our new parameters a and b evolved throughout the simulation. The blue curve indicates the value of a, whose scale is shown in the left axis. The green curve indicates the value of b, whose scale is shown in the right axis. With every increase of a beyond 0.83, b increases to a new fixed point.

As the simulation starts, the system quickly moves to its stable fixed point at around 10 Hz. Then with each excitatory burst, b increases to a new level, which simply moves the stable fixed point up by some amount.

3.3 Stability Analysis

To test how persistent the activity of this neuron is, we run a similar experiment to Figure 5. Our neuron requires a 150ms pulse of 40 $\mu A/cm^2$ in the excitatory burst neuron to bring its firing rate to around 20Hz, which was the same initial firing rate after the excitatory burst in Figure 5. We can see the result in Figure 10. After 40 seconds of firing, the neuron's activity has remained persistent.

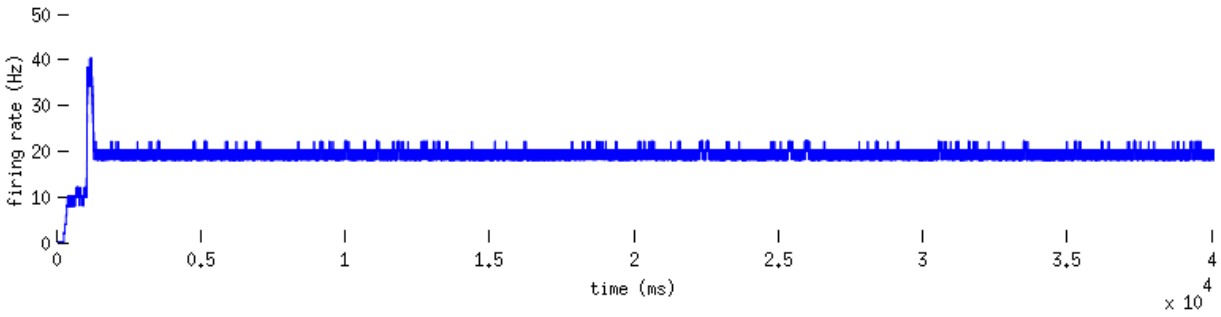


Figure 10: 40 second simulation of the improved autapse. At 1 second (0.1×10^4ms in the plot), the excitatory burst neuron was given a 150ms pulse of 40 $\mu A/cm^2$.

For our final stability test, we tweak the same variables we perturbed in the analysis for the original model. We can see in Figure 11, in the top plot, that when we increase the strength of the autapse's connection to itself, the stable levels of firing rate increased slightly, with the neuron's "default" firing rate at around 12 Hz. However, this did not affect the persistent graded firing behavior. In the second plot, we effective removed the autapse entirely, setting W to 0. Although the levels of firing decreased slightly, the overall persistent graded activity did not change. Next, in the third plot, we increased the strength of the connection from the tonic neuron to the autapse to 2.4. This had the same effect as increasing the value of W. Although we can see at the highest level of activity (3 - 4 seconds), the neuron's activity fluctuates. This is due to the fact that the variable a hovers very near the upper threshold. However, with the hyperpolarizing burst, the persistent graded activity is restored. In the fourth plot, we see the effect of reducing the value of W_0 to 1.650. At first, it seems that the neuron has gone silent. But a and b are still changing with every excitatory burst. After the second and third bursts, we see that graded persistent activity did not disappear. Instead, the levels of activity were reduced to the point where the first burst was not enough to get the neuron firing. The neuron seems to be more sensitive to the inhibitory bursts.

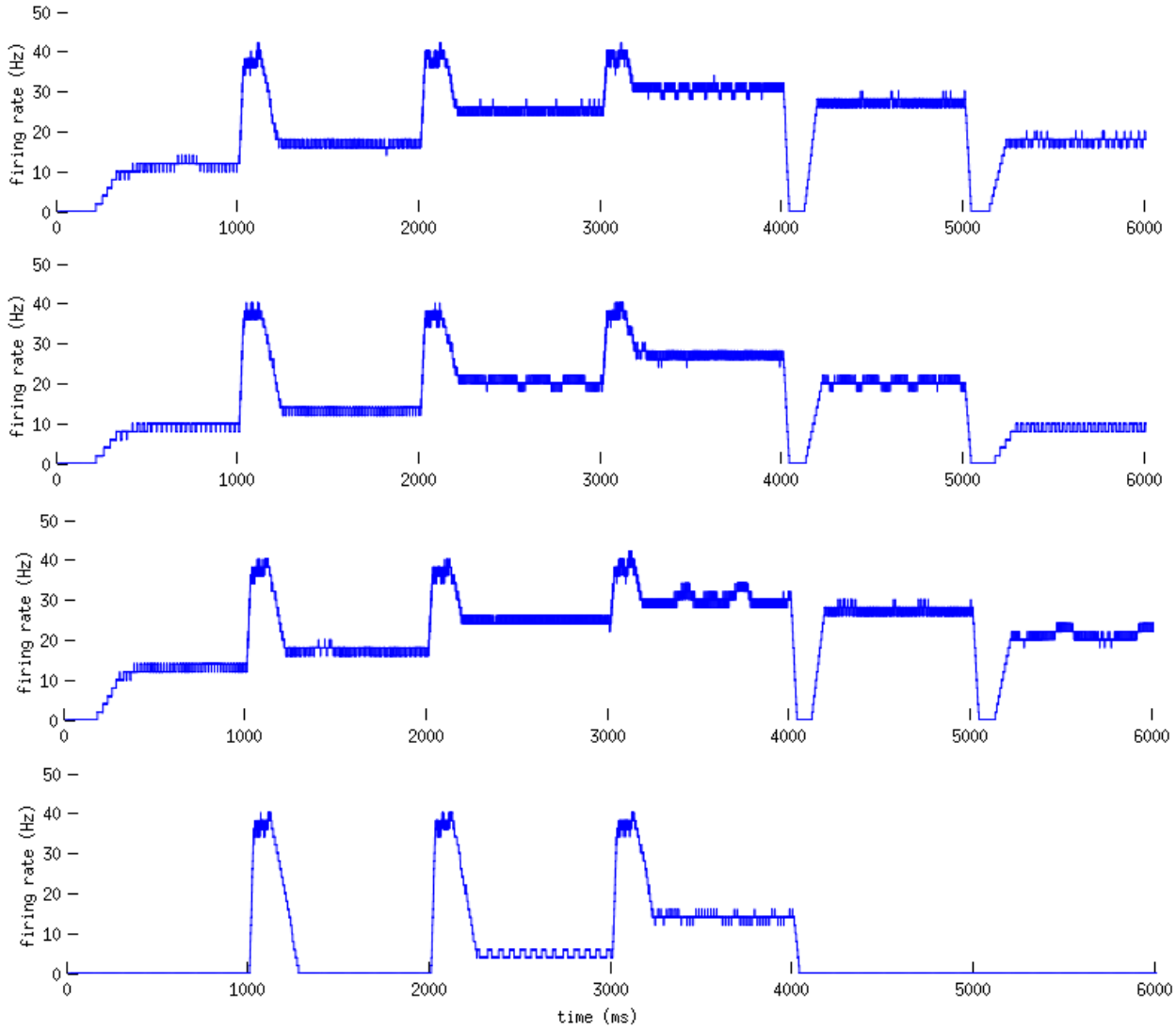


Figure 11: Four simulations with identical parameters to the simulation in Figure 9, except: In the top plot, W was increased to 0.35. In the second plot, W was set to 0, effectively removing the self-connection. In the third plot, W_0 was set to 2.4. In the last plot, W_0 was set to 1.650.

4 Conclusion

We have shown that it is possible to implement a neural system that exhibits graded persistent activity in a single autapse neuron using the model devised by Seung. However, this model is not very robust to changes in parameters, and it is very easy to destroy the graded persistent behavior of the neuron with very small changes to the weights. By introducing two variables that effectively determine a stable fixed point of firing activity for the neuron, we can significantly improve the robustness of the model. Interestingly, the self-connection is no longer necessary in this improved model. There is evidence (Koulakov et al. 2002) that a significant proportion of ion channels on the surface of neural integrators in brain areas such as

entorhinal cortex are calcium-dependent. Calcium would fit the role of our variable a perfectly. Pumps would normally cause the level of internal calcium to decay, but calcium channels allows an influx of calcium with every action potential. There is evidence that biological systems actually implement a more complex system of careful intracellular storage and release of a compound known to be fundamentally involved in graded persistent activity in entorhinal neurons (Teramae et al. 2005). Nevertheless, our model, with its dramatically improved stability and robustness, presents a biologically-plausible solution to the problem of tuning.

5 References

Egorov, A.V., Hamam, B.N., Fransen, E., Hasselmo, M.E., and Alonso, A.A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature* 420, 173–178.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the primate dorsolateral prefrontal cortex. *Journal of Neurophysiology*. 61, 331-349.

Goldman MS, Kaneko CR, Major G, Aksay E, Tank DW, Seung HS (2002) Linear regression of eye velocity on eye position and head velocity suggests a common oculomotor neural integrator. *Journal of Neurophysiology* 88, 659-665.

Goldman-Rakic, P.S. (1995). Cellular basis of working memory. *Neuron* 14, 477–485.

Koulakov, A. A., Raghavachari, S., Kepecs, A., & Lisman, J. E. (2002). Model for a robust neural integrator. *Nature Neuroscience* 5, 775–782.

Seung, H. S., Lee, D. D., Reis, B. Y., and Tank, D. W. (2000). The autapse: A simple illustration of short-term analog memory storage by tuned synaptic feedback. *Journal of Computational Neuroscience* 9, 171–185.

Teramae, J.-N., and Fukai, T. (2005). A cellular mechanism for graded persistent activity in a model neuron and its implications in working memory. *Journal of Computational Neuroscience* 18, 105–121.