Journal of Computational Neuroscience 9, 171–185, 2000

The Autapse: A Simple Illustration of Short-Term Analog Memory Storage by Tuned Synaptic Feedback

H. SEBASTIAN SEUNG
Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, Cambridge, MA 02139; Lucent Technologies, Bell Laboratories, Murray Hill, NJ 07974
seung@mit.edu

DANIEL D. LEE, BEN Y. REIS AND DAVID W. TANK
Lucent Technologies, Bell Laboratories, Murray Hill, NJ 07974

Received May 20, 1999; Revised November 30, 1999; Accepted December 2, 1999

Action Editor: John Rinzel

Abstract. According to a popular hypothesis, short-term memories are stored as persistent neural activity maintained by synaptic feedback loops. This hypothesis has been formulated mathematically in a number of recurrent network models. Here we study an abstraction of these models, a single neuron with a synapse onto itself, or autapse. This abstraction cannot simulate the way in which persistent activity patterns are distributed over neural populations in the brain. However, with proper tuning of parameters, it does reproduce the continuously graded, or analog, nature of many examples of persistent activity. The conditions for tuning are derived for the dynamics of a conductance-based model neuron with a slow excitatory autapse. The derivation uses the method of averaging to approximate the spiking model with a nonspiking, reduced model. Short-term analog memory storage is possible if the reduced model is approximately linear and if its feedforward bias and autapse strength are precisely tuned.

Keywords: short-term memory, persistent neural activity, synaptic feedback, reverberating circuit

1. Introduction

In parts of the central nervous system ranging from the spinal cord (Prut and Fetz, 1999) to the neocortex (Fuster, 1995), transient inputs have been observed to cause persistent changes in the rate of action potential firing. By now, there is no doubt that this general phenomenon is closely related to short-term memory, but its physiological mechanisms remain unknown. According to one long-standing hypothesis, persistent neural activity is maintained by synaptic feedback loops (Lorente de No, 1933; Hebb, 1949; Amit, 1995). This hypothesis has found precise mathematical formulation in a number of recurrent network models (Cannon et al., 1983; Seung, 1996; Georgopoulos et al., 1993; Zipser et al., 1993; Griniasty et al., 1993; Amit et al., 1994; Zhang, 1996; Camperi and Wang, 1998). Not only do these models maintain persistent activity patterns, but they also reproduce the experimentally observed ways in which neural firing rates encode computational variables.

In this article we analyze an abstraction of these recurrent network models: a single neuron that makes a synapse onto itself, or autapse. In this model, feedback is localized to a single loop, rather than distributed over a complex web of connections. This simplification

handicaps the model because it is unable to capture the distributed nature of the persistent activity patterns that are observed in the brain.

However, the autapse model can reproduce a key property of persistent activity—the fact that it is often observed to be continuously graded (Seung, 1996; Muller et al., 1996; Romo et al., 1999). As we shall see, such analog persistence requires precise tuning of model parameters, a requirement that is shared by more complex network models. Consequently, the autapse model is valuable as a particular example of the general idea that analog short-term memory through feedback requires precise tuning (Seung, 1996).

We begin with a general discussion of the dynamics of a conductance-based model neuron with a slow excitatory autapse (Ermentrout, 1998b). Our analysis utilizes the method of averaging to eliminate the dynamics of the intrinsic conductances, leaving a nonspiking, reduced model (Rinzel and Frankel, 1992; Ermentrout, 1994; Ermentrout, 1998b). We establish the conditions for analog short-term memory in the reduced model: it must be linear, and the strength of the autapse and the feedforward bias must be precisely tuned.

This is followed by a demonstration of analog short-term memory in numerical simulations of the spiking, conductance-based model. We use a particular set of intrinsic conductances for which the reduced model is approximately linear (Shriki et al., 1999). The autapse strength and feedforward bias are tuned using the reduced model. When these tuned parameters are placed into the conductance-based model, numerical simulations show that transient inputs cause persistent changes in neural activity.

Mistuning the parameters in the simulations causes a loss of persistence. According to linear feedback theory, which is of limited applicability here, neural activity should behave roughly exponentially, with a time constant that depends on the deviation of the autapse strength from its tuned value. To achieve good persistence, the autapse strength must be tuned to a precision equal to the ratio of the synaptic time constant to the persistence time of activity. In other words, the autapse model is more robust to changes in parameters when it has a longer synaptic time constant.

Even when the autapse is precisely tuned, small amounts of drift remain in the numerical simulations. If the autapse is viewed as a system for analog memory storage, this means that the stored variable is gradually corrupted, so that the memory is only short-term. The drift is quantitatively related to nonlinearities in the reduced model.

On the whole, the reduced model is a very good approximation to the spiking model. However, there are some inconsistencies between them. The spiking model has more null points (stable levels of activity at which drift completely vanishes) than the reduced model. Furthermore, the spiking model exhibits firing rate oscillations that are not present in the reduced model. These discrepancies are due to a breakdown of the method of averaging due to a general phenomenon known as resonance.

Given that the autapse model is a simple and natural abstraction of recurrent networks, it is not surprising that there have been previous studies of it. These studies differ from the present model either in their lack of biophysical realism (Kamath and Keller, 1976; Cannon et al., 1983; Nakahara and Doya, 1998) or by not considering analog memory storage (Ermentrout, 1998b).

2. The Conductance-Based Model

We begin by describing an autapse[1] model based on the dynamics of intrinsic and synaptic conductances (Ermentrout, 1998b). The membrane potential V of the model neuron obeys the current balance equation

$$C_m \frac{dV}{dt} = -I_{int}(V, c_1, \ldots, c_n) - g_E(V - V_E), \quad (1)$$

where C_m is the membrane capacitance. The intrinsic currents I_{int} depend on voltage and a set of channel variables $c_1, \ldots, c_n$. The synaptic current is produced by the synaptic conductance g_E, with excitatory reversal potential V_E. For now, we will not specify the functional form of I_{int} or the dynamics of $c_1, \ldots, c_n$ to keep the discussion general.

The current balance equation describes how synaptic input is converted into action potentials. Action potentials lead to synaptic transmission and the opening of postsynaptic receptors. The kinetics of these receptors are given by the simple two-state model

$$\tau \frac{ds}{dt} + s = \alpha(1 - s)\sigma(V), \quad (2)$$

which is illustrated in Fig. 1. The synaptic activation s is the fraction of open channels at the autapse and is a dimensionless variable taking values in the range from zero to one. The presynaptic voltage V enters through the sigmoid function (Wang and Rinzel, 1992)

$$\sigma(V) = \frac{1}{1 + \exp[-(V - \theta_s)/\sigma_s]}. \quad (3)$$

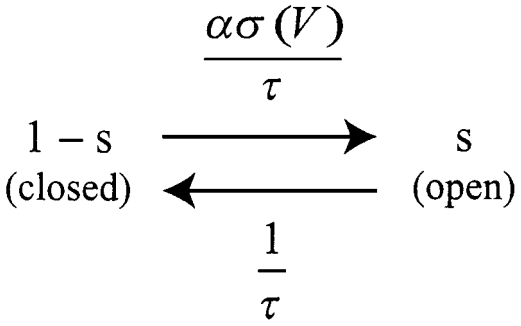

Figure 1. Two-state kinetic model of synapse, with forward and backward rate constants of $\alpha\sigma(V)/\tau$ and $1/\tau$, respectively. The forward reaction is gated on and off by the sigmoid function $\sigma(V)$, which is vanishingly small except during action potentials.

The threshold θ_s and width σ_s of the sigmoid function are chosen so that it is normally close to zero but approaches unity during an action potential, causing a rapid increment in the synaptic activation s. The magnitude of the increment is controlled by the parameter α and saturates as s approaches unity, due to the factor of $1 - s$ on the right-hand side of Eq. (2). During the intervals between action potentials, s decays towards zero with time constant τ.

When activated, the autapse provides recurrent excitation back to the neuron through the equation

$$g_E = Ws + B. \tag{4}$$

The synaptic weight W is the maximal conductance of the autapse if all of its channels are open ($s = 1$). There is also a bias B, which could come from a variety of sources. In our numerical simulations to be described later, the bias will arise from tonic feedforward input. Alternatively, it could correspond to a nonzero resting level of open synaptic channels or an intrinsic conductance with the same reversal potential as the synaptic conductance.

3. The Method of Averaging

We have assumed that the dynamics of action potential generation involve the membrane potential V and a set of channel variables $c_1, \ldots, c_n$. If the model neuron is firing action potentials, these variables change rapidly. In contrast, the synaptic activation s changes relatively slowly, provided that the synaptic time constant τ is long. In such a situation, where there are distinct subsets of fast and slow dynamical variables, it is possible to approximately eliminate the fast variables using the method of averaging.

This is done by averaging Eq. (2) with respect to V while holding g_E constant (Rinzel and Frankel, 1992; Ermentrout, 1994; Ermentrout, 1998b). We will assume that there is a threshold value of g_E above which the model neuron converges to repetitive firing and below which it converges to a quiescent state. Let the periodic membrane potential be denoted by $V(t; g_E)$, and replace $\sigma(V(t; g_E))$ in Eq. (2) by its time average

$$f(g_E) = \frac{1}{T(g_E)} \int_0^{T(g_E)} dt\, \sigma(V(t; g_E)) \tag{5}$$

over an interspike interval of length $T(g_E)$. The substitution yields

$$\tau \frac{ds}{dt} + s = \alpha(1 - s) f(g_E), \tag{6}$$

which approximately describes the way in which synaptic input g_E to the model neuron leads to changes in the autapse activation s. The autapse in turn determines the synaptic input via $g_E = Ws + B$. Substituting this in Eq. (6) yields a dynamics in the single variable s, a description of the autapse from which the dynamics of action potential generation have been eliminated.[2]

The relevant properties of the intrinsic currents have been encapsulated in the function $f(g_E)$. This function is the time average of the sigmoid function $\sigma(V)$, which detects the occurrence of action potentials. Therefore, f is generally an increasing function of frequency of action potentials above threshold and vanishes below threshold. For the specific set of intrinsic currents to be numerically simulated later, we will see that f is almost exactly proportional to firing rate.

4. Conditions for Analog Memory Storage

We now determine the conditions under which the reduced model is able to store a memory of an analog variable. Here it is convenient to regard $g_E(s)$ as a function of s defined by Eq. (4). Then the reduced model (6) can be rewritten as

$$\tau \frac{ds}{dt} = -s + \alpha f(g_E(s))(1 - s) \tag{7}$$

$$= [\alpha f(g_E(s)) + 1]\left[\frac{\alpha f(g_E(s))}{\alpha f(g_E(s)) + 1} - s\right]. \tag{8}$$

The steady states of this dynamics satisfy

$$s = \frac{\alpha f(g_E(s))}{1 + \alpha f(g_E(s))} \equiv F(g_E(s)). \tag{9}$$

Analog memory storage is possible if every value of s is a steady state, at least over some range of s. Equivalently, F must be the inverse of the function $g_E(s)$ over some range of s.

This is possible, provided that the following conditions are satisfied. First, F must be linear,

$$F(g_E) = F_1 g_E + F_0, \tag{10}$$

since $g_E(s)$ is linear. Second, the autapse strength W and bias B must be tuned so that

$$W = 1/F_1, \tag{11}$$

$$B = -F_0/F_1. \tag{12}$$

If these conditions are fulfilled, Eq. (9) is satisfied for all s, which can be verified by substituting $Ws + B$ for $g_E(s)$.

5. Tuning the Reduced Model

We have arrived at a set of conditions for analog memory storage by the reduced model. The synaptic time constant τ should be long, so that the method of averaging is applicable. The function F should be linear, and B and W should be tuned according to Eqs. (11) and (12).

The shape of the function F depends on the particular intrinsic conductances of the model neuron and the parameters of the synaptic dynamics. It is unrealistic to expect the function F to be perfectly linear. The best that we can hope for is that F will be approximately linear over some range. This means that ds/dt will be small but nonzero after tuning: memory storage will be imperfect, due to drift.

To design an approximately linear F, we use a model neuron introduced by Shriki, Sompolinsky, and Hansel (Shriki et al., 1999). The neuron has a single compartment with a leak current I_L, a fast sodium current I_{Na}, a delayed-rectifier potassium current I_K, and an A-type potassium current I_A. The dynamics of these currents are described completely in the appendix. For our model synapse, we choose $\tau = 100$ ms in Eq. (2). We expect that the reduced model should be accurate when this time constant is slower than the dynamics of intrinsic conductances. We also choose $\alpha = 1$ to make the effects of synaptic saturation weak.

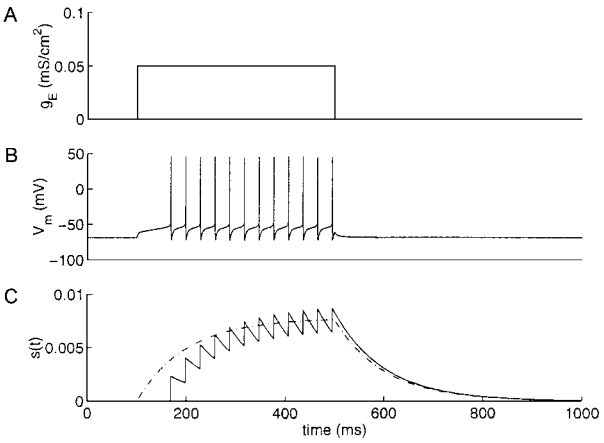

Figure 2. Dynamics of membrane voltage V and synaptic activation s in response to synaptic input. **A:** The excitatory synaptic conductance g_E steps from 0 to 0.05 mS/cm^2, starting at 100 ms and ending at 500 ms. **B:** The resulting train of action potentials starts after a latency of about 68 ms. Convergence to periodic behavior is rapid, with little spike frequency adaptation. **C:** Synaptic summation for $\alpha_s = 1$ and $\tau = 100$ ms. Each action potential causes a jump in the synaptic activation s, with rise time equal to the width of the action potential. Between action potentials, s decays exponentially with time constant τ. The uninterrupted decay after the last action potential exhibits this exponential behavior clearly. Since the synapse is far below saturation, all the jumps are of roughly equal amplitude. The broken line shows the behavior of the reduced model Eq. (6) for comparison.

Numerical simulations of the model neuron and model synapse are shown in Fig. 2. If the synaptic conductance g_E is held constant in time at a value above threshold, the model neuron converges rapidly to repetitive firing at constant rate, as shown in Figs. 2A and B. Each action potential causes a fast jump in the synaptic activation, shown in Fig. 2C. Every jump is of approximately the same magnitude because there is little saturation, due to our choice of a small α. During the intervals between action potentials, the synaptic activation decays with time constant $\tau = 100$ ms. The exponential decay is most evident in the uninterrupted period after the last action potential.

The shape of $f(g_E)$ is found by numerically simulating repetitive firing like that of Fig. 2B for various values of g_E and computing the time average in Eq. (5). We chose the parameters of the sigmoid function (3) to be $\theta_s = -20$ mV and $\sigma_s = 2$ mV. Figure 3A shows that the function $f(g_E)$ is zero for values of g_E below threshold, and roughly linear in g_E for values above threshold. It turns out that $f(g_E)$ is almost exactly proportional to firing rate ν.[3]

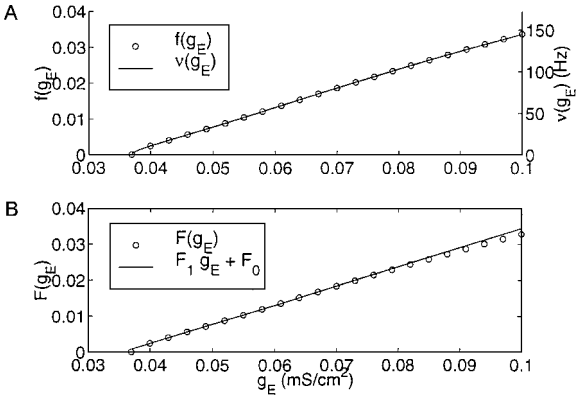

Figure 3. Repetitive firing properties of the model neuron. If the excitatory conductance is held constant, then both the membrane potential V and the synaptic activation s converge to periodic behavior. **A:** Transfer function f and firing rate ν versus excitatory synaptic conductance g_E. The two functions are almost exactly proportional to each other, $f(g_E) \approx 0.2328\ \text{kHz}^{-1}\nu(g_E)$. Except for some rounding of the curves near threshold, both functions are roughly linear in g_E over the range in the graph. **B:** Transfer function F versus g_E. Unlike f, this function includes the effect of synaptic saturation. The graph is close to linear, as saturation is weak for the parameter values chosen. The best linear fit (see Eq. (13)) to points evenly spaced from 0.038 to 0.070 in steps of 0.0005 is also shown.

The function F is calculated from f by using Eq. (9). According to the reduced model (6), the synaptic activation approaches $F(g_E)$ exponentially. As shown in Fig. 2, this exponential behavior is a good approximation, except that it neglects the sudden jumps due to action potentials. This is a reasonable approximation when the firing rate is greater than $1/\tau = 10$ Hz, so that low-pass filtering by the synapse makes the synaptic activation behave more smoothly than the membrane potential.

Our choices of parameters result in a very linear F, as evident in Fig. 3B. This is because F has the same shape as f for small α, according to Eq. (9). A least squares fit yields the linear approximation

$$F(g_E) \approx 0.5314 g_E - 0.01878. \qquad (13)$$

As shown in Fig. 3B, this is a good approximation over a range of values of g_E, although some signs of saturation are visible in F at larger values. The values $F_1 = 0.5314$ and $F_0 = -0.01878$ can be substituted into the formulas (11) and (12) to obtain the tuned values $W = 1.882$ and $B = 0.03534$. These and other parameters are listed in Table 1.

Table 1. Parameters used in the autapse simulations of Figs. 5 and 9. The synaptic weights are tabulated according to the identity of the presynaptic neuron. The two tuned parameters are the strengths of the autapse and tonic synapse. The current applied to the burst neurons is normally zero, except during bursts, when it is 5 μA/cm^2.

	Memory	Tonic	Excitatory Burst	Inhibitory Burst
Synaptic weight (mS/cm^2)	1.882	3.800	1	-4
I_{app} (μA/cm^2)	0	3	0/5	0/5
τ_{syn} (ms)	100	100	5	5

6. Numerical Simulations of Persistent Activity

The tuning procedure outlined above was based on the reduced model. Now that we have arrived at tuned values for W and B, it is important to verify that these values indeed endow the capability of analog memory storage to the original spiking model.

For our numerical simulations, it is helpful to use the circuit illustrated in Fig. 4 and described mathematically in the appendix. The neuron with the autapse, which we call the *memory neuron*, is the only one that is essential for memory storage. However, there are also three input neurons, which are useful for illustrative purposes. The bias input B to the memory neuron is provided by a *tonic neuron*, which fires repetitively at a constant firing rate of roughly 40 Hz. The bias has been adjusted to its tuned value given in Table 1 by setting the strength of the tonic synapse to $W_0 = B/\langle s_0 \rangle$, where $\langle s_0 \rangle = 0.00930$ is the mean value of the activation of the tonic synapse. In addition to the tonic neuron, there are also two *burst neurons*. As will be seen shortly, transient excitatory and inhibitory inputs from these neurons cause persistent changes in the activity of the memory neuron.

All input neurons have the same intrinsic currents as the memory neuron, but they do not receive synaptic inputs. Instead, their activities are produced by applied currents. The time constant of the tonic synapse is the same as that of the autapse. The burst synapses have been made faster (5 ms) to better illustrate persistence in the activity of the memory neuron, but this choice is not important for the model. Equations describing all neurons are given in the appendix.

The membrane potentials of all four neurons in Fig. 4 are graphed as functions of time in Fig. 5A. The tonic neuron fires at a constant rate of roughly 40 Hz. The two burst neurons are silent, except for brief bursts of a

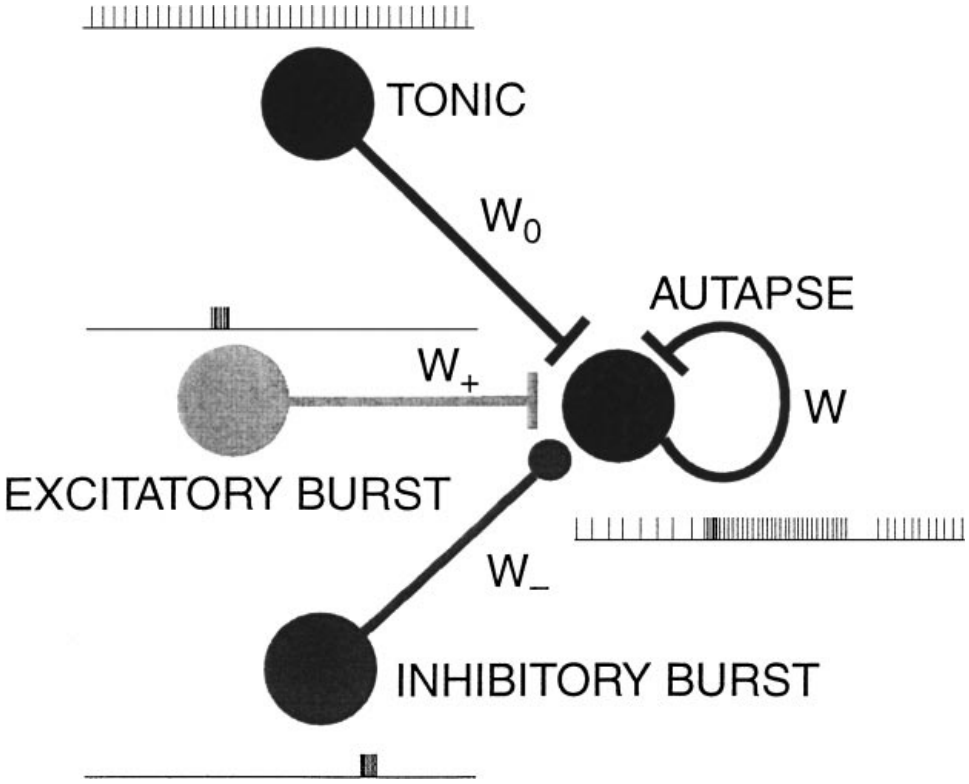

Figure 4. Circuit diagram of the autapse model, with three input neurons in addition to the memory neuron, which has an autapse to itself. The tonic neuron provides an excitatory feedforward bias to the memory neuron. The burst neurons change the activation of the autapse by providing excitatory and inhibitory burst inputs.

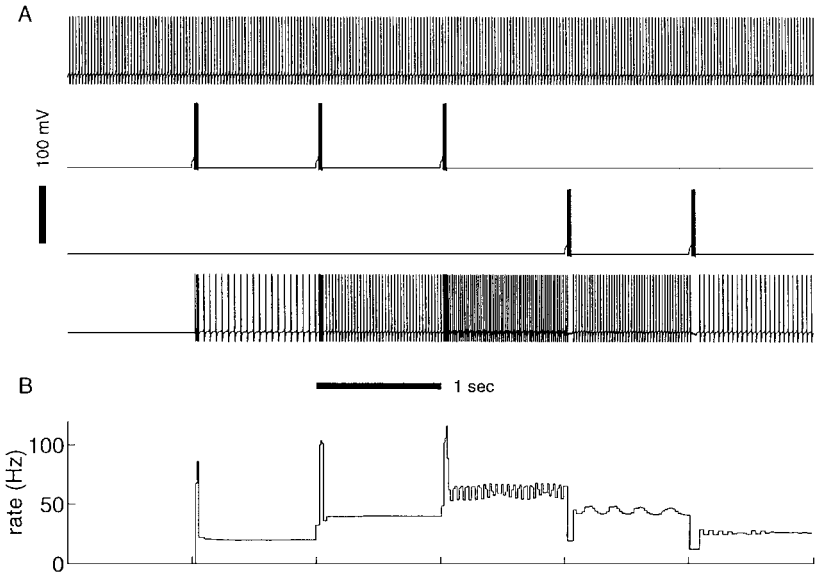

Figure 5. Analog memory storage by the tuned autapse of Fig. 4. **A:** Membrane potential as a function of time. At the top, the tonic neuron fires at a steady rate of 40 Hz. The second and third traces are of the excitatory and inhibitory burst neurons. The fourth trace is of the memory neuron, which fires at a rate that is roughly constant over an interburst interval, but varies from interval to interval. **B:** Instantaneous firing rate of the memory neuron, computed as the reciprocal of the interspike interval. Each excitatory burst input causes a transient pulse and a persistent upward step in rate. Each inhibitory burst input causes a transient pause and a persistent downward step.

few action potentials at a rate exceeding 100 Hz. Each excitatory burst evokes a burst in the memory neuron, and each inhibitory burst causes the memory neuron to pause briefly from firing. These transient effects of transient stimulation are not surprising. More remarkably, the transient stimulation has a persistent effect: the firing rate immediately after a burst is different than it was before the burst, and this new rate persists until the next burst a second later.

This persistence is not caused by feedforward synaptic input. Since the burst synapses have time constants of 5 ms, burst synaptic input quickly decays back to zero after the burst. Furthermore, the tonic input never changes: it is the same both before and after the burst. Only the autapse has different activation before and after the burst. After an excitatory burst, the autapse is more activated and therefore stimulates the memory neuron to a higher firing rate. In turn, the higher firing rate causes the autapse to have a higher level of activation. This circular interaction is a positive feedback loop.

The action potential times in Fig. 5A can be converted to instantaneous firing rates, as shown in Fig. 5B. The firing rate of the memory neuron is roughly constant in time during each interburst interval, though there is some systematic drift visible in the plot.

This simple circuit illustrates a general principle—that positive feedback can lead to persistent neural activity. In Fig. 5B, the memory neuron is active at five different nonzero rates, although its feedforward input during each interval is the same. In other words, the memory neuron is able to store information in the form of its activity. During each interburst interval, its activity is determined by the past, not by the feedforward input it is currently receiving. The persistence of these activity states will be discussed in more detail below.

7. Mistuning and Linear Feedback Theory

The consequences of mistuning W and B can be predicted by approximating the nonlinear reduced model (7) with the linear reduced model

$$\tau \frac{ds}{dt} \approx (WF_1 - 1)s + F_1 B + F_0. \qquad (14)$$

In addition to approximating F as linear (Eq. 10), we have also neglected a factor of $\alpha f(g_E) + 1$, which is valid when the effects of synaptic saturation are weak.

The linear reduced model has a unique fixed point at $s = (F_1 B + F_0)/(1 - WF_1)$. The fixed point is stable if $WF_1 < 1$ and unstable if $WF_1 > 1$. In the stable case, s approaches the fixed point exponentially with time constant $\tau/(1 - WF_1)$. In the unstable case, s diverges exponentially from the fixed point with time constant $\tau/|1 - WF_1|$.

The predictions of the linear reduced model are confirmed by the numerical simulations of Figs. 6, 7 and 8. In Fig. 6, the autapse strength W is reduced to 3/4 of its tuned value, and the strength of the tonic synapse is increased to $W_0 = 4.4$. For these parameters, the linear reduced model (14) predicts a stable fixed point at $s = 0.0118$ and a time constant of 400 ms. Indeed, the behavior of Fig. 6C approximately matches this prediction.

Unstable behavior can be produced if W is changed to 5/4 of its tuned value, and $W_0 = 3.2$. The linear reduced model (14) predicts an unstable fixed point at $s = 0.0119$ and a time constant of 400 ms. Again, this predicted behavior matches the behavior of the conductance-based model, as shown in Fig. 7C.

If W is held at its tuned value, and only the tonic synapse is detuned to $W_0 = 3.98$, then the autapse shows imbalanced behavior. The linear reduced model (14) predicts $ds/dt = (F_1 B + F_0)/\tau_{syn} = 8.9 \times 10^{-3}$/sec, independent of s. The linearly increasing traces in Figs. 8B and C bear this prediction out.

A useful way of visualizing the nonlinear reduced model (7) is to make a graph of the drift velocity ds/dt as a function of s. Such graphs are shown in the insets of Figs. 6C, 7C, and 8C. The solid line in the inset is from the nonlinear reduced model, while the points are from numerical simulations of the conductance-based model. The graphs are nearly straight lines, as to be expected from the linear reduced model (14). The case of $WF_1 < 1$ is a line with negative slope, indicating a stable fixed point in Fig. 6C. The slope is positive for $WF_1 > 1$, so that the fixed point is unstable in Fig. 7C. In Fig. 8C, the slope is zero, but the intercept is positive, resulting in imbalanced behavior.

8. Drift and Nonlinearities

Our tuning of the autapse was based on a linear approximation (13) to the function $F(g_E)$ over a finite range of its argument g_E. Outside this finite range, the threshold and saturation nonlinearities of F compromise the accuracy of the linear reduced model. Because of the threshold nonlinearity, s cannot become negative in the nonlinear reduced model (7), while there is no such restriction in the linear reduced model. Furthermore,

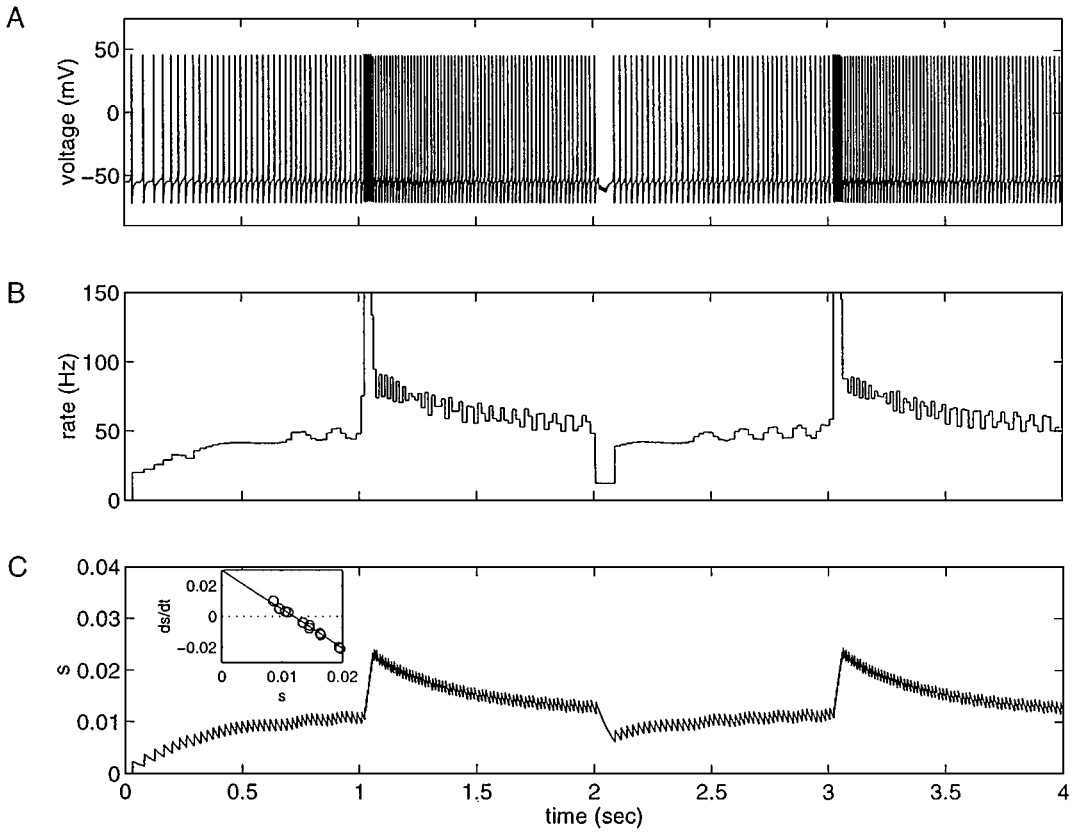

Figure 6. Activity of a leaky autapse, with strength 3/4 of its tuned value. The strength of the tonic synapse is $W_0 = 4.4$. The burst synapses have strengths $W_+ = 3$ and $W_- = 10$. There are excitatory bursts at 1 and 3 seconds, and an inhibitory burst at 2 seconds. **A:** Membrane potential versus time for the memory neuron. During the interburst intervals, the firing rate is not constant but either speeds up or slows down. **B:** Instantaneous firing rate versus time for the memory neuron. Excitatory and inhibitory bursts drive the memory neuron to high and low activity levels, but the firing rate always converges to a null point of around 50 Hz. **C:** The synaptic activation s behaves similarly to the firing rate, as they are roughly linearly related for $\alpha_s = 1$. The inset shows the relationship of ds/dt to s. The line is from the reduced model of Eq. (7). The points are from linear fits of 200 ms segments of $s(t)$.

the saturation nonlinearity is weak in F but eventually starts to take effect for large s, making the linear reduced model very inaccurate.

Even within the finite range over which the linear approximation (13) was constructed to be valid, the linear reduced model is not perfect. To reveal some of its imperfections, we examine the nonlinear reduced model (7) more closely. The solid line in Fig. 9C is a graph of the drift ds/dt as a function of s, using Eq. (7). Because the autapse has been tuned, the drift ds/dt is small everywhere. However, it does not vanish, except at a few discrete values of s. And there is no way of improving the tuning of W and W_0 so that ds/dt vanishes for all s. This is because changing these parameters cannot change the fact that the graph is nonlinear and hence cannot vanish everywhere. This is a direct consequence of the nonlinearity of the function F, which determines ds/dt through Eq. (7). In contrast, in the perfectly tuned linear reduced model (14), the drift ds/dt would vanish for all s.

The fixed points of the nonlinear reduced model are the values of s where the solid line in Fig. 9C intersects the horizontal dotted line. There are two stable fixed points with nonzero s, identifiable by their negative slope, and one stable fixed point at $s = 0$. At long times, the nonlinear reduced model converges to one of these fixed points. Therefore, the autapse can only store a *short-term* memory of a continuous variable. At long times, the stored memory converges to one of three discrete values. When mistuned, a linear reduced model also shares the limitation that memory of a continuous variable is short-term. But there is the

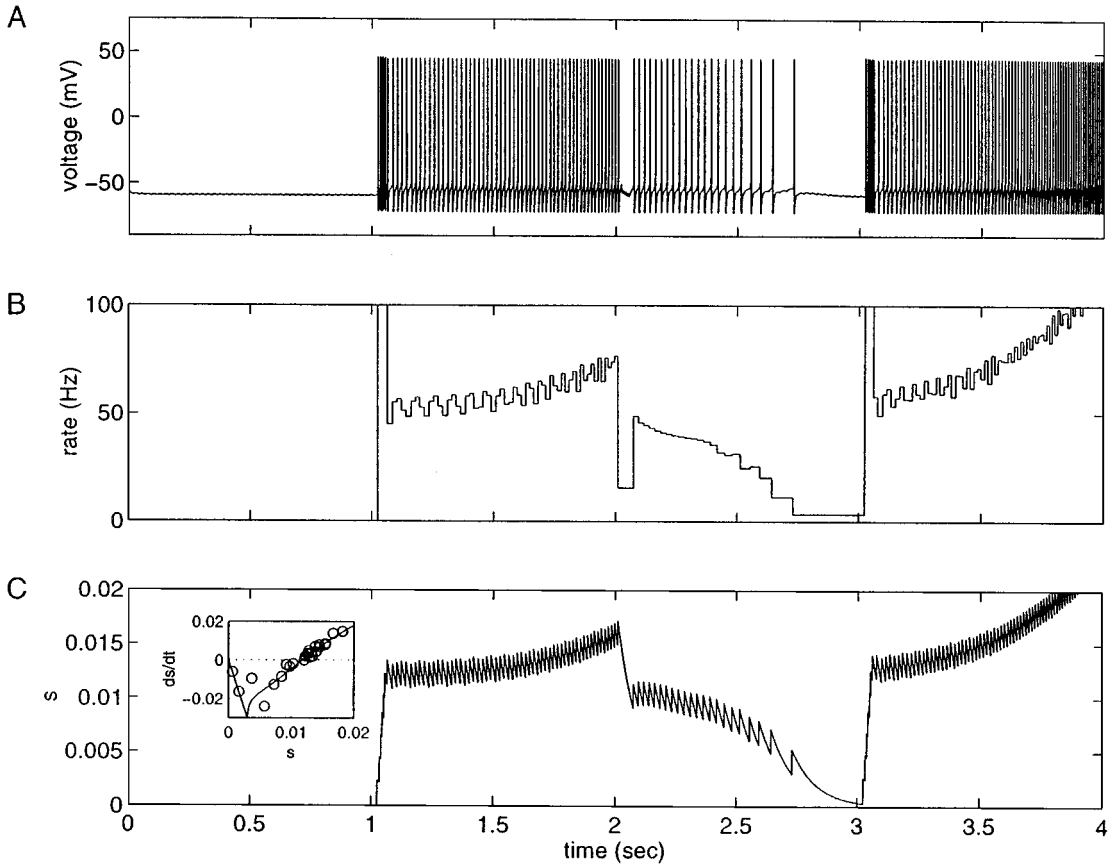

Figure 7. An unstable autapse, with strength 5/4 of its tuned value. The tonic synapse is decreased to $W_0 = 3.2$ from its tuned value. The burst synapses have strengths $W_+ = 2.93$ and $W_- = 5.1$. There are excitatory bursts at 1 and 3 seconds, and an inhibitory burst at 2 seconds. **A:** Membrane potential versus time for the memory neuron. **B:** Instantaneous firing rate versus time for the memory neuron. The firing rate repeatedly diverges from 50 Hz, after being reset by excitatory and inhibitory bursts. **C:** The synaptic activation s behaves similarly to the firing rate, as they are roughly linearly related for $\alpha_s = 1$. The inset shows the relationship of ds/dt to s. The line is from the reduced model of Eq. (7). The points are from linear fits of 200 ms segments of $s(t)$.

qualitative difference that a mistuned linear model has at most one stable fixed point; it cannot have three.

9. Resonances

We have seen that the linear reduced model (14) does not perfectly reproduce the properties of the nonlinear reduced model (7). In turn, the nonlinear reduced model (7) is not a perfect description of the original spiking model.

Figure 9A shows $s(t)$ for a tuned autapse driven by a randomized sequence of bursts, once per second for five minutes of simulated time. A 6-second portion of this time series is shown in Fig. 9B. During each interburst interval the drift velocity ds/dt and average value of s are calculated. The relationship between these two quantities is plotted in the points of Fig. 9C, which can be compared with the solid line from the reduced model (7). The correspondence with the points from the simulations is generally quite good.

But there are some discrepancies, such as the notches at $s = 0.0045$, 0.009, and 0.018. A continuous line drawn through one of these notches crosses zero with negative slope, indicating that value of s is an attractive state of the dynamics. In contrast, the nonlinear reduced model has only two stable fixed points with nonzero activity, and they are at different locations.

This failure of the nonlinear reduced model is due to the phenomenon of *resonance*, in which the method of averaging breaks down because the frequencies of a dynamical system are in whole number ratios to each

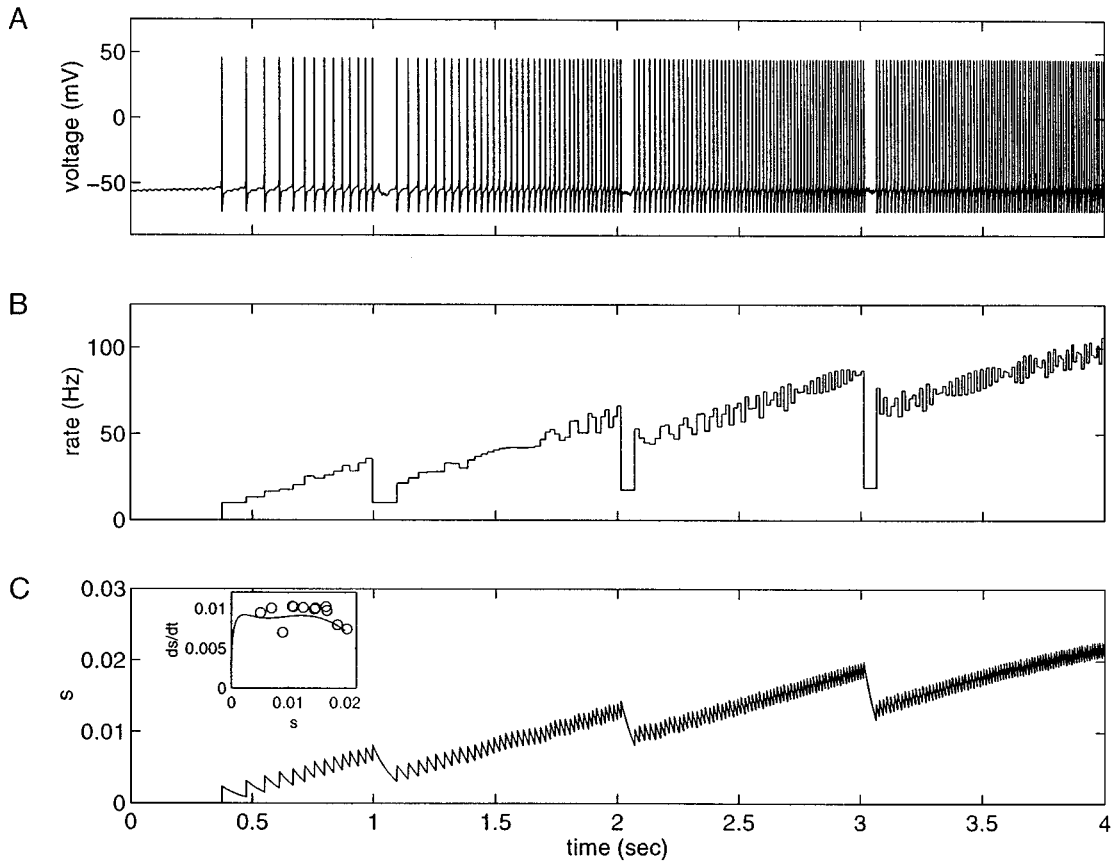

Figure 8. Imbalanced autapse due to increased bias from tonic input. The strength of the tonic synapse is $W_0 = 3.98$, and the inhibitory burst synapse has strength $W_- = 4$. There are inhibitory bursts at 1, 2, and 3 seconds. **A:** Membrane potential versus time for the memory neuron. **B:** Instantaneous firing rate versus time for the memory neuron. The firing rate of the memory neuron drifts higher during interburst intervals. **C:** The synaptic activation s behaves similarly to the firing rate, as they are roughly linearly related for $\alpha_s = 1$. The inset shows the relationship of ds/dt to s. The line is from the reduced model Eq. (7), and the points are from the simulations.

other (Sanders and Verhulst, 1985). Here there are two frequencies—the firing rates of the memory neuron and the tonic neuron. The tonic neuron has a tonic rate of 40 Hz. The three stable values of s mentioned above correspond to 20, 40, and 80 Hz firing rates of the memory neuron. In other words, the memory neuron is attracted to 1 : 2, 1 : 1, and 2 : 1 resonances with the tonic neuron. The 1 : 2 and 1 : 1 resonances can be seen in the graph of instantaneous firing rate shown in 5B. In the interburst intervals starting at 1 and 2 seconds, the autapse converges to repetitive firing at roughly 20 Hz and 40 Hz, respectively.

More complex resonance effects can be seen in the interburst intervals starting at 3, 4, and 5 seconds. During these intervals, the rate oscillates, a behavior not predicted by the nonlinear reduced model. If the periodic tonic input s_0 is replaced by its time average in the numerical simulations, the rate oscillations vanish, confirming that they are indeed due to resonance between the memory and tonic neurons. Similar oscillatory behaviors are seen in the instantaneous firing rates of mistuned autapses (see Figs. 6B, 7B, and 8B).

Discrepancies between the nonlinear reduced model and the conductance-based model can also be seen in Fig. 9C at small values of s corresponding to frequencies lower than 10 Hz. This is because the intrinsic currents that generate the action potential are changing on a time scale that is comparable to or longer than the synaptic time constant of 100 ms. The method of averaging is valid only in the opposite limit of intrinsic currents that change much faster than the synaptic variables.

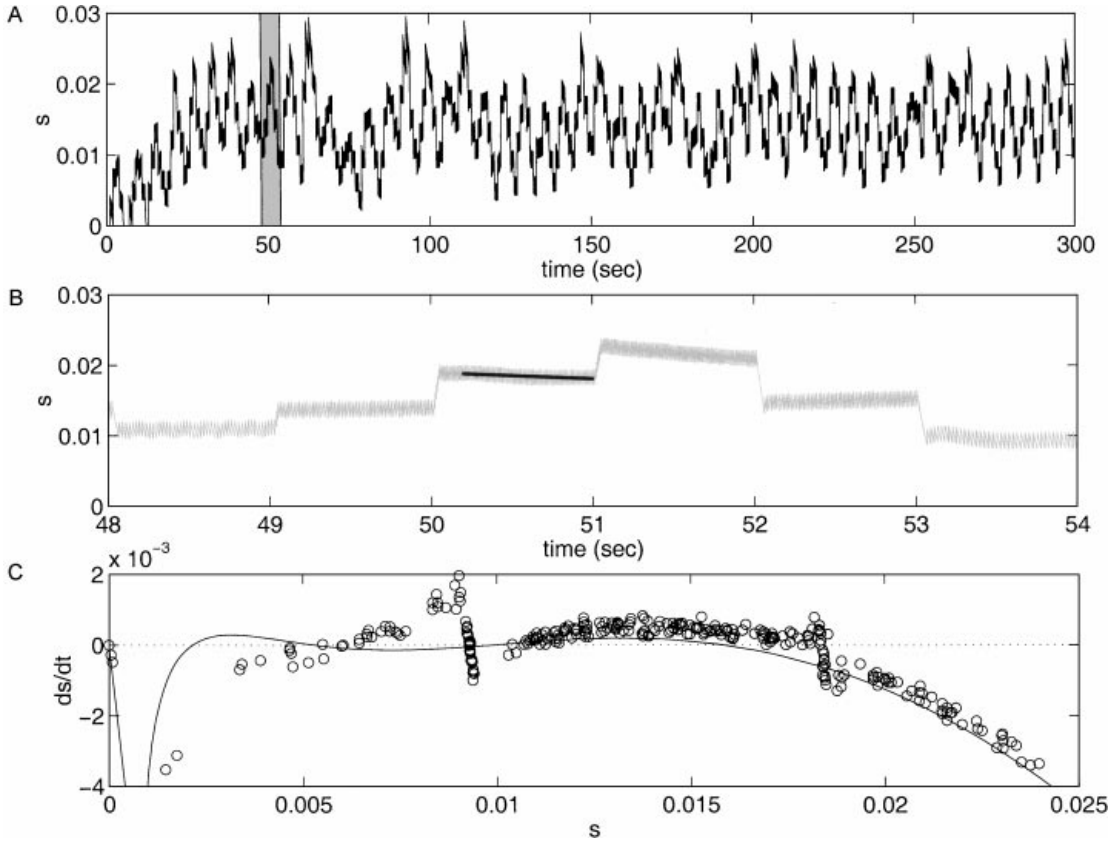

Figure 9. Tuned autapse driven by bursts at one second intervals for five minutes of simulated time. The bursts are stimulated by applied currents that are Gaussian random variables with mean 5 μA/cm^2 and standard deviation 1 μA/cm^2. **A:** The autapse activation s versus time. **B:** An expanded view of six seconds of the time series shown in (a), along with a sample linear fit to $s(t)$ during an interburst interval, excluding the first 200 ms after the burst. The slope of this line is a measure of ds/dt. **C:** Drift ds/dt as a function of s. The points, obtained from the numerical simulations of the conductance-based model, are in good agreement with the solid line, from the reduced model.

10. Discussion

10.1. Design of the Autapse Model

We have demonstrated that a conductance-based autapse model can perform analog memory storage with a persistence time that is much longer than its elementary biophysical time constants. This capability is not generic: we carefully designed our autapse model to have it. Our first design choice was to make the time constant of the autapse long. This enabled us to use the method of averaging to eliminate the dynamics of intrinsic conductances, yielding a nonlinear reduced model that approximates the original spiking model. Then we showed that analog memory storage is possible only when the function F of the reduced model (7) is linear.

To make the function F approximately linear, two more design choices were made. We chose a model neuron for which the relationship between firing rate and synaptic input is approximately linear above threshold. Our model neuron contained leak, sodium, delayed rectifier, and A-type potassium currents (Shriki et al., 1999), but there are a number of other model neurons that also possess the required linearity of response (Wang, 1998; Ermentrout, 1998a). Also, we chose the parameter α of the synaptic dynamics (2) so that the effects of saturation would be weak, making the relationship between synaptic activation and firing rate approximately linear.

Because of these design choices, it was possible to linearly approximate F by Eq. (13) in a range of synaptic conductances. Using this approximation, we tuned the autapse strength and feedforward bias based on

Eqs. (11) and (12). We then used these parameters in the original spiking model and verified that they resulted in analog persistence, as shown in Fig. 5. Some small amount of drift remained, even with the tuned parameters, because F was not perfectly linear. As shown in Fig. 9, this drift in the spiking model was consistent with the reduced model.

To understand the roles of these various design choices in producing persistence, it is helpful to work backwards and relax them one by one. If the autapse strength or feedforward bias are mistuned, then substantial drift in activity appears. This drift may be manifested as leakiness (Fig. 6), instability (Fig. 7), or imbalance (Fig. 8).

If F were more nonlinear (due to synaptic saturation, for example), no amount of tuning of the autapse strength of feedforward bias could produce analog memory storage, though digital memory storage would be possible. In a certain parameter range, the autapse would rapidly converge to one of two stable states, silent or firing repetitively at a fixed rate. This bistability could be used to store a single bit of information with an infinite persistence time.

If the time constant of the autapse were short, then the method of averaging would no longer be applicable, and the precise timing of spikes would be important. It is not clear whether a continuously variable firing rate could be maintained in this case. The maintenance of persistent activity in networks by "synfire" mechanisms involving precise spike timing has been studied by others (Abeles, 1991; Herrmann et al., 1995; Hertz and Prugel-Bennett, 1996; Maass and Natschlager, 1997) and is perhaps closer to the original concept of "reverberating activity" due to Lorente de No (1933) and Hebb (1949).

It is interesting to note that even in the case of a slow autapse, there are some dynamical behaviors dependent on spike timing. In particular, the resonance effects evident in the instantaneous firing rates of Figs. 5B, 6B, 7B, and 8B are not captured by the reduced model and show that the spiking nature of neural activity cannot be ignored completely.

10.2. Relation to Recurrent Network Models

The autapse model should be viewed as an abstraction of more complex network models in which synaptic feedback is distributed over many pathways, rather than localized to a single loop. Such recurrent network models of motor cortex (Georgopoulos et al., 1993), prefrontal cortex (Camperi and Wang, 1998), the head direction system (Zhang, 1996), and the oculomotor integrator (Cannon et al., 1983; Seung, 1996) are able to maintain persistent patterns of neural activity distributed over a population of neurons. In contrast, the autapse model does not reproduce the distributed nature of neural codes.

Another limitation of the autapse is its inability to perform analog memory storage when F has a strong saturation nonlinearity. In another article, we have shown that this is not a problem for recurrent networks, which can compensate for the sublinearity of saturation by recruiting neurons above threshold (Seung et al., 2000).

In spite of these differences, the recurrent network models and the autapse model are similar in one respect. They share the property of sensitivity to mistuning of synaptic weights and other parameters (Seung, 1996; Zhang, 1996). In the linear reduced model (14), the persistence time is equal to $\tau/(1-WF_1)$, where τ is the elementary time constant of the system, and WF_1 is the strength of feedback. Therefore, a small time constant τ can be boosted to a long persistence time if WF_1 is tuned to one. Significant increases in persistence require precise tuning of WF_1. For example, 1 percent accuracy is required to boost a τ of 100 ms to a 10-second persistence time. This result was derived for a linear system with a single feedback loop and is not directly applicable to the recurrent network models, which have nonlinearities and distributed feedback. However, the general message of linear feedback theory about the dependence of analog memory storage on precise tuning does seem to be relevant for these networks (Seung, 1996; Seung et al., 2000). In certain network models, the need for precise tuning is hidden behind an assumption of symmetry in the synaptic connections but it is still there (Zhang, 1996).

Therefore, it seems that these recurrent network models should be combined with some adaptive mechanism for tuning the synaptic weights and other parameters to sustain persistent neural activity. Possible synaptic learning rules have been studied in a number of network models with nonspiking neurons (Arnold and Robinson, 1992, 1997; Seung, 1997).

It should be mentioned that there are some recurrent networks in which persistent activity does not depend on precise tuning (Hopfield, 1982; Griniasty et al., 1993; Amit et al., 1994). The difference is that the persistent activity patterns of these networks are discrete,

while those of the networks mentioned previously are continuously variable. The memory storage capabilities of networks that do not require precise tuning are more digital in character, rather than analog. As argued elsewhere, the differences between these two types of networks can be conceptualized in terms of discrete and continuous dynamical attractors (Seung, 1996, 1998).

Appendix: Model Equations and Parameters

Our simulations utilize a model neuron introduced by Shriki, Hansel, and Sompolinsky (1999). The only modification we have made is to increase the threshold by using a higher leak conductance. Unless otherwise noted, the measurement units are voltage (mV), conductance (mS/cm^2), current (μA/cm^2), and capacitance (μF/cm^2). The specific membrane capacitance for all neurons is $C_m = 1\ \mu$F/cm^2.

Intrinsic Conductances

The intrinsic currents are given by the sum

$$I_{int}(V,h,n,b) = I_L(V) + I_{Na}(V,h) + I_K(V,n) + I_A(V,b), \tag{15}$$

where V is the membrane potential, and h, n, and b are channel variables.

1. Leak Current I_L

$$I_L(V) = g_L(V - V_L), \tag{16}$$

where $g_L = 0.2$, $V_L = -65$.

2. Sodium Current I_{Na}

$$I_{Na}(V,h) = g_{Na} m_\infty^3(V) h (V - V_{Na}) \tag{17}$$

$$m_\infty(V) = \frac{\alpha_m(V)}{\alpha_m(V) + \beta_m(V)} \tag{18}$$

$$\alpha_m(V) = \frac{(V+30)/10}{1 - \exp[-(V+30)/10]} \tag{19}$$

$$\beta_m(V) = 4\exp[-(V+55)/18] \tag{20}$$

$$\phi_h^{-1}\frac{dh}{dt} = \alpha_h(V)(1-h) - \beta_h(V)h \tag{21}$$

$$\alpha_h(V) = 0.07\exp[-(V+44)/20] \tag{22}$$

$$\beta_h(V) = \frac{1}{\exp[-(V+14)/10] + 1}, \tag{23}$$

where $g_{Na} = 100$, $V_{Na} = 55$, $\phi_h = 10$.

3. Delayed Rectifier Potassium Current I_K

$$I_K(V) = g_K n^4 (V - V_K) \tag{24}$$

$$\phi_n^{-1}\frac{dn}{dt} = \alpha_n(V)(1-n) - \beta_n(V)n \tag{25}$$

$$\alpha_n(V) = \frac{(V+34)/100}{1 - \exp[-(V+34)/10]} \tag{26}$$

$$\beta_n(V) = \frac{1}{8}\exp[-(V+44)/80], \tag{27}$$

where $g_K = 40$, $V_K = -80$, $\phi_n = 10$.

4. A-Type Potassium Current I_A

The activation variable a is instantaneous, while the inactivation variable b has a relaxation time τ_b that is independent of voltage:

$$I_A(V) = g_A a_\infty^3 b (V - V_K) \tag{28}$$

$$a_\infty(V) = \frac{1}{\exp[-(V+50)/20] + 1} \tag{29}$$

$$b_\infty(V) = \frac{1}{\exp[(V+80)/6] + 1} \tag{30}$$

$$\frac{db}{dt} = \frac{b_\infty(V) - b}{\tau_b}, \tag{31}$$

where $g_A = 20$, $\tau_b = 20$.

Synaptic Conductances

As shown in Fig. 4, the memory neuron (neuron with the autapse) is driven by three input neurons. The synaptic conductances of the memory neuron are given by

$$g_E = Ws + W_0 s_0 + W_+ s_+, \tag{32}$$

$$g_I = W_- s_-, \tag{33}$$

where s is the autapse activation, and s_0, s_+, and s_- are the activations tonic, excitatory burst, and inhibitory burst synapses, respectively. The input neurons do not have synaptic conductances; their firing is driven by applied currents.

For the sake of brevity, only the dynamical equations for the membrane potentials and synaptic activations are given. They must be supplemented by the equations for the channel variables of the intrinsic conductances, as described above.

1. Tonic Neuron

$$C_m \frac{dV_0}{dt} = -I_{int}(V_0, h_0, n_0, b_0) + I_{app,0}, \quad (34)$$

$$\tau_0 \frac{ds_0}{dt} + s_0 = \alpha(1 - s_0)\sigma(V_0). \quad (35)$$

Tonic activity in this neuron is generated by an applied current of $I_{app,0} = 3\ \mu\text{A/cm}^2$. The resulting behavior of s_0 shows periodic fluctuations about a mean value of 0.00930. The synaptic time constant τ_0 is 100 ms, as with the memory neuron.

2. Excitatory Burst Neuron

$$C_m \frac{dV_+}{dt} = -I_{int}(V_+, h_+, n_+, b_+) + I_{app,+}, \quad (36)$$

$$\tau_+ \frac{ds_+}{dt} + s_+ = \alpha(1 - s_+)\sigma(V_+). \quad (37)$$

This is like the tonic neuron but with a shorter synaptic time constant τ_+ of 5 ms, and an applied current $I_{app,+}$ that consists of 50 ms current pulses with amplitudes described in the figure captions. These pulses cause brief bursts of activity.

3. Inhibitory Burst Neuron

$$C_m \frac{dV_-}{dt} = -I_{int}(V_-, h_-, n_-, b_-) + I_{app,-}, \quad (38)$$

$$\tau_- \frac{ds_-}{dt} + s_- = \alpha(1 - s_-)\sigma(V_-). \quad (39)$$

As with the excitatory burst neuron, burst activity is generated by 50 ms current pulses with amplitudes described in the figure captions. The synaptic time constant τ_- is 5 ms.

4. Memory Neuron

$$C_m \frac{dV}{dt} = -I_{int}(V, h, n, b) - g_E(V - V_E) - g_I(V - V_I), \quad (40)$$

$$\tau \frac{ds}{dt} + s = \alpha(1 - s)\sigma(V). \quad (41)$$

The voltages $V_E = 0$ and $V_I = -70$ are the reversal potentials of excitatory and inhibitory synapses, respectively. The strength of the autapse is tuned to $W = 1.882$ mS/cm^2, and the strength of the tonic synapse to $W_0 = 3.800$, unless otherwise noted in the figure captions. These values are determined from the conditions (11) and (12), along with the linear approximation (13). The strengths of the excitatory and inhibitory burst synapses are $W_+ = 1$ and $W_- = 4$ mS/cm^2, unless otherwise noted in the figure captions. Each synaptic strength is the maximal conductance of the synapse, attainable only when all of its receptors are open.

Numerical Methods

We used the fourth-order Runge-Kutta method with step size 0.01 ms to integrate these equations, except in Fig. 3, where a step size of 0.002 ms was used. With no synaptic or applied current, the dynamical variables converge to a fixed point at $V = -68.3737$, $h = 0.9820$, $n = 0.0631$, and $b = 0.1259$.

Instantaneous rate functions were calculated from spike times defined as the downward zero crossings of the membrane potential. The rate between successive spikes at times t_a and t_{a+1} was defined as $1/(t_{a+1} - t_a)$. In other words, the rate function was piecewise constant in each interspike interval.

Acknowledgments

We are grateful to O. Shriki, H. Sompolinsky, and D. Hansel for providing us with their model neuron. This work was supported by Lucent Technologies and MIT.

Notes

1. Autapses are seen frequently in cultured neurons (Bekkers and Stevens, 1991; Segal, 1991; 1994). Some have argued that they are artifacts of the low cell densities in culture and do not occur normally in the intact brain. But recent studies have found that

autapses are common in both excitatory (Lubke et al., 1996) and inhibitory (Tamas et al., 1997) neurons of the neocortex.
2. When α is large and saturation is strong, this approximation is less accurate. A modified form of the method of averaging for saturating synapses is described elsewhere (Seung et al., 2000).
3. This relationship holds because the shape of the action potential, and hence the integral in Eq. (5), are roughly independent of frequency, as is typical of Class I neurons (Ermentrout, 1994, 1998b). Consequently, f is proportional to the prefactor $1/T(g_E)$, which is just the frequency ν.

References

Abeles M (1991) Corticonics: Neural Circuits of the Cerebral Cortex. Cambridge University, Cambridge.

Amit DJ (1995) The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behav. Brain Sci.* 18:617–626.

Amit DJ, Brunel N, Tsodyks MV (1994) Correlations of cortical Hebbian reverberations: Theory versus experiment. *J. Neurosci.* 14:6435–6445.

Arnold DB, Robinson DA (1992) A neural network model of the vestibulo-ocular reflex using a local synaptic learning rule. *Phil. Trans. R. Soc. Lond.* B337:327–330.

Arnold DB, Robinson DA (1997) The oculomotor integrator: Testing of a neural network model. *Exp. Brain Res.* 113:57–74.

Bekkers JM, Stevens CF (1991) Excitatory and inhibitory autaptic currents in isolated hippocampal neurons maintained in cell culture. *Proc. Natl. Acad. Sci. USA* 88(17):7834–7838.

Camperi M, Wang XJ (1998) A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability [in process citation]. *J. Comput. Neurosci.* 5(4):383–405.

Cannon SC, Robinson DA, Shamma S (1983) A proposed neural network for the integrator of the oculomotor system. *Biol. Cybern.* 49:127–136.

Ermentrout B (1994) Reduction of conductance-based models with slow synapses to neural nets. *Neural Comput.* 6:679–695.

Ermentrout B (1998a) Linearization of f-i curves by adaptation. *Neural Comput.* 10:1721–1729.

Ermentrout B (1998b) Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys.* 61:353–430.

Fluster JM (1995) Memory in the Cerebral Cortex. MIT Press, Cambridge, MA.

Georgopoulos AP, Taira M, Lukashin A (1993) Cognitive neurophysiology of the motor cortex. *Science* 260:47–52.

Griniasty M, Tsodyks MV, Amit DJ (1993) Conversion of temporal correlations between stimuli to spatial correlations between attractors. *Neural Comput.* 5:1–17.

Hebb DO (1949) Organizational Behavior. Wiley, New York.

Herrmann M, Hertz JA, Prugel-Bennett A (1995) Analysis of synfire chains. *Network* 6:403–414.

Hertz J, Prugel-Bennett A (1996) Learning short synfire chains by self-organization. *Network* 7:357–363.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79:2554–2558.

Kamath BY, Keller EL (1976) A neurological integrator for the oculomotor control system. *Math. Biosci.* 30:341–352.

Lorente de No R (1933) Vestibulo-ocular reflex arc. *Arch. Neurol. Psych.* 30:245–291.

Lubke J, Markram H, Frotscher M, Sakmann B (1996) Frequency and dendritic distribution of autapses established by layer 5 pyramidal neurons in the developing rat neocortex: Comparison with synaptic innervation of adjacent neurons of the same class. *J. Neurosci.* 16(10):3209–3218.

Maass W, Natschlager T (1997) Networks of spiking neurons can emulate arbitrary Hopfield nets in temporal coding. *Network* 8:355–371.

Muller RU, Ranck Jr JB, Taube JS (1996) Head direction cells: Properties and functional significance. *Curr. Opin. Neurobiol.* 6:196–206.

Nakahara H, Doya K (1998) Near-saddle-node bifurcation behavior as dynamics in working memory for goal-directed behavior. *Neural Comput.* 10(1):113–132.

Prut Y, Fetz EE (1999) Primate spinal interneurons show premovement instructed delay activity. *Nature* 401(6753):590–594.

Rinzel J, Frankel P (1992) Activity patterns of a slow synapse network predicted by explicitly averaging spike dynamics. *Neural Comput.* 4:534–545.

Romo R, Brody CD, Hernandez A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399(6735):470–473.

Sanders JA, Verhulst F (1985) Averaging Methods in Nonlinear Dynamical Systems. Applied mathematical sciences. Springer-Verlag, New York.

Segal MM (1991) Epileptiform activity in microcultures containing one excitatory hippocampul neuron. *J. Neurophysiol.* 65(4):761–770.

Segal MM (1994) Endogenous bursts underlie seizurelike activity in solitary excitatory hippocampal neurons in microcultures. *J. Neurophysiol.* 72(4):1874–1884.

Seung HS (1996) How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. USA* 93:13339–13344.

Seung HS (1997) Learning to integrate without visual feedback. *Soc. Neurosci. Abstr.* 23(1):8.

Seung HS (1998) Learning continuous attractors in recurrent networks. *Adv. Neural Info. Proc. Sys.* 10:654–660.

Seung HS, Lee DD, Reis BY, Tank DW (2000) Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron.* 26:259–271.

Shriki O, Sompolinksy H, Hansel D (1999) Rate models for conductance based cortical neural networks. Unpublished.

Tamas G, Buhl EH, Somogyi P (1997) Massive autaptic self-innervation of gabaergic neurons in cat visual cortex. *J. Neurosci.* 17(16):6352–6364.

Wang XJ (1998) Calcium coding and adaptive temporal computation in cortical pyramidal neurons. *J. Neurophysiol.* 79:1549–1566.

Wang XJ, Rinzel J (1992) Alternating and synchronous rhythms in reciprocally coupled inhibitory model neurons. *Neural Comput.* 1992:534–545.

Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *J. Neurosci.* 16:2112–2126.

Zisper D, Kehoe B, Littlewort G, Fuster J (1993) A spiking network model of short-term active memory. *J. Neurosci.* 13:3406–3420.