ARTICLE Communicated by Mark Goldman

Neural Integrator: A Sandpile Model

Maxim Nikitchenko
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, U.S.A., and Center for Neurobiology and Behavior, Columbia University, New York, NY 10032, U.S.A.

Alexei Koulakov
koulakov@cshl.edu
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, U.S.A.

We investigated a model for the neural integrator based on hysteretic units connected by positive feedback. Hysteresis is assumed to emerge from the intrinsic properties of the cells. We consider the recurrent networks containing either bistable or multistable neurons. We apply our analysis to the oculomotor velocity-to-position neural integrator that calculates eye positions using the inputs that carry information about eye angular velocity. By analyzing this system in the parameter space, we show the following. The direction of hysteresis in the neuronal response may be reversed for the system with recurrent connections compared to the case of unconnected neurons. Thus, for the NMDA receptor-based bistability, the firing rates after ON saccades may be higher than after OFF saccades for the same eye position. The reversal of hysteresis occurs in this model only when the size of hysteresis differs from neuron to neuron. We also relate the macroscopic leak time constant of the integrator to the rate of microscopic spontaneous noise-driven transitions in the hysteretic units. Finally, we investigate the conditions under which the hysteretic integrator may have no threshold for integration.

1 Introduction

Persistent neuronal firing is a likely correlate of short-term memory (Fuster, 1995; Goldman-Rakic, 1995). In some cases, the variables stored in memory are continuous in nature (Romo, Brody, Hernandez, & Lemus, 1999). Examples of such quantities include continuous sensory inputs (Romo et al., 1999; Miller, Brody, Romo, & Wang, 2003; Machens, Romo, & Brody, 2005), tension of a muscle, or variables representing accumulated sensory evidence (Shadlen & Newsome, 2001; Mazurek, Roitman, Ditterich, & Shadlen, 2003). The continuously varying parameters are encoded in neuronal firing, which has a graded set of values. The components of the nervous system

that encode the graded values of parameters are called parametric memory systems (Romo et al., 1999; Miller et al., 2003; Machens et al., 2005).

Perhaps the best-studied system of this type is the oculomotor neural integrator (Robinson, 1974, 1989; Fukushima & Kaneko, 1995; Major & Tank, 2004). Graded persistent activity in this system represents continuously varying eye position, which depends on the prior inputs carrying information about eye angular velocity. Since the transformation from velocity to position involves temporal integration, this system is also sometimes called a velocity-to-position neural integrator (VPNI). The graded persistent activity in VPNI is likely to be maintained by positive feedback (Rosen, 1972; Robinson, 1989; Seung, Lee, Reis, & Tank, 2000a; Major & Tank, 2004). The presence of positive feedback poses a problem of robustness (Robinson, 1989; Aksay, Gamkrelidze, Seung, Baker, & Tank, 2001). This is because mistuning of the feedback leads to instabilities, which are hard to avoid in realistic systems. Previous researchers proposed that robustness with respect to parameter mistuning could stem from hysteresis in the neuronal responses (Koulakov, Raghavachari, Kepecs, & Lisman, 2002; Goldman, Levine, Major, Tank, & Seung, 2003). In this approach, the robustness is similar to the stability of digital electronic systems to mistuning of parameters and noise. A similar argument has been made for robustness of persistent activity with respect to distracters and noise (Camperi & Wang, 1998).

A recent study in the goldfish oculomotor integrator (Aksay et al., 2003) has tested the history dependence in the responses of VPNI neurons. This study makes the following observations. First, the firing rate of a single neuron as a function of eye position exhibits hysteresis (see Figure 1B). Second, the firing rates during fixations are typically higher after the ON saccades than after the OFF saccades (see Figure 1B, by ON or OFF saccades, it is customary to understand the eye movements in the direction of increasing or decreasing firing rates of neurons). This finding implies that the hysteresis has an inverted direction compared to a typical positive feedback system, such as that due to nonlinear conductance of the NMDA receptor current shown in Figure 1C (Wang, 1999; Koulakov et al., 2002). Third, the firing rate of one cell versus the other also displays history dependence. Fourth, the hysteresis width varies from cell to cell, with some cells showing no statistically substantial history dependence (see Figure 1A).

Our study addresses these experimental observations. We developed a simple model for VPNI that can be solved exactly without the use of a computer. We considered two related versions of this model involving bistable and multistable neurons. The bistability is attributed to the bistable compartments within a single neuron (Lisman, Fellous, & Wang, 1998; Koulakov et al., 2002), while the multistability is formed by many bistable dendritic compartments. Although the specific mechanism is proposed, the properties of neurons in this model could be understood phenomenologically and could be generated by many intracellular or

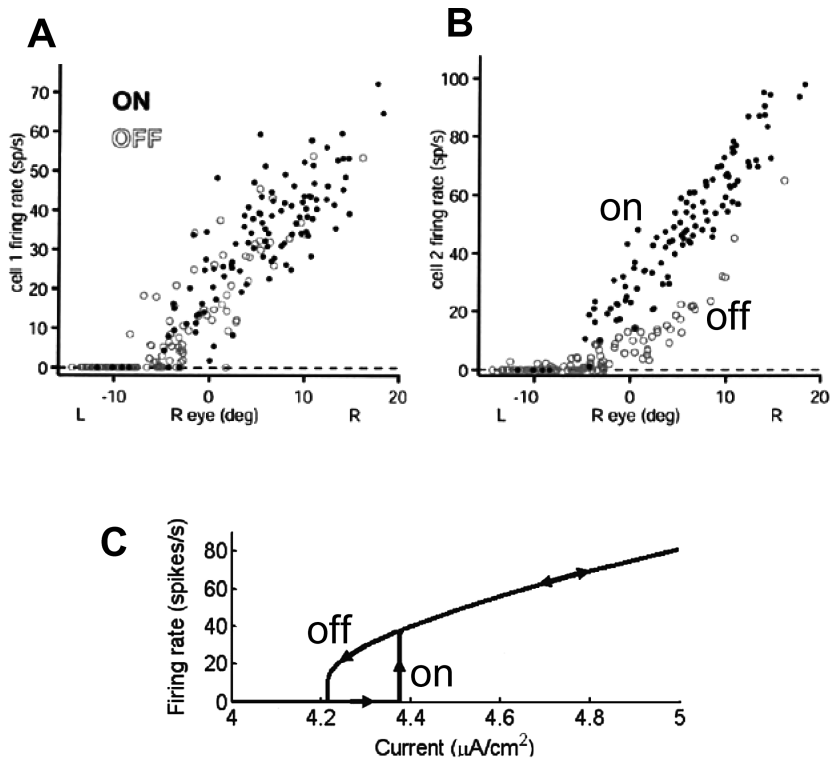

Figure 1: Although responses of some goldfish medulla area I neurons display little or no hysteresis (A), other neurons are clearly hysteretic (B) (Aksay et al., 2003). For the hysteretic responses, the firing rates after ON saccades are above those after the OFF saccades. Here and throughout the rest of the article, the ON/OFF saccades are defined as abrupt eye movements in the direction of increasing or decreasing the firing rate of the neuron. (C) Firing rate as a function of input current for a model neuron. The hysteretic response in the neuron has origins in the nonlinearity of NMDA receptor conductance (adopted from Koulakov et al., 2002). The ON response is below the OFF response in contrast to experiments in *B*. The discrepancy between *B* and *C* will be explained below in this article.

network mechanisms. We also present the results for a more biologically plausible computational model, which are consistent with the simpler model.

The main results of our study are as follows. First, we show that if the neurons in the absence of recurrent connections have hysteresis of regular sign (counterclockwise) (Lisman et al., 1998; Koulakov et al., 2002), adding global recurrent feedback produces the reversed hysteresis (clockwise)

that is consistent with the higher firing rates seen after the ON saccades as described above. Thus, the direction of hysteresis observed experimentally could be attributed to the global recurrent connections between cells. Second, the phenomenon of the reversal of the sign of hysteresis occurs only if different neurons have different widths of hysteresis. Thus, the experimental observation number two, that the firing rates are higher after the ON saccades, may follow from observation number four, that the hysteresis width varies from cell to cell. Finally, we studied the temporal properties of VPNI using a kinetic equation formulated in the parameter space of the system. We show that the rate of integration is controlled by the synaptic time constant τ_s, which, in the case of the NMDA receptor, is about 0.1 sec. The integrator leak time constant τ_{leak} is determined by the rate of spontaneous transitions in the bistable neurons denoted here τ_h. The expression for the integrator leak is of the form

$$\tau_{\text{leak}} = \tau_h/\varepsilon. \tag{1.1}$$

The parameter $\varepsilon \ll 1$ defines the precision with which the integrator is tuned. This expression is valid for small values of tuning parameter ε as discussed in section 5. For the VPNI without hysteresis, the leak is given by the same expression with τ_h replaced by τ_s (Robinson, 1989; Seung et al., 2000a). Because the time constant of spontaneous transitions τ_h is usually much larger than the synaptic time constant $\tau_s \sim 0.1$ sec (Bialek, 2000; Koulakov, 2001), the use of hysteretic neurons allows stabilization of the integrator at a much larger value of the precision of tuning ε, which provides another argument for the robustness of the hysteretic system.

2 Computational Model

The network model that we used in the computational part of this study is similar to the previously described NMDA-based models (Lisman et al., 1998; Koulakov et al., 2002). The network included 40 two-compartmental neurons. Each neuron contained the somatic and dendritic compartments. The somatic compartment included sodium and potassium currents, making it capable of generating action potentials. The dendritic compartments received feedforward NMDA current, feedback NMDA current, and an offset current needed to distribute the thresholds for activation. The NMDA-based bistability was produced by the feedforward NMDA currents into the dendritic compartments. This current was due to feedforward inputs from 100 neurons discharging at 30 Hz. The NMDA conductance for feedforward inputs was equal to 0.7, 1, 1.1, and 1.2 mS/cm^2 for neurons from groups 1 through 4 in Figure 11A. Different values of NMDA conductance in these

neurons resulted in differing hysteresis. These four groups of neurons also received feedforward input currents of 1.98, 1.20, 0.93, and 0.68 $\mu A/cm^2$ to equate their mean thresholds for activation (parameter θ) as shown in Figure 11A. To produce a difference in the mean thresholds, another offset current was added to the quadruples of neurons. Each quadruple was separated by 0.05 $\mu A/cm^2$ (see Figure 11A) in θ-space from its nearest neighbor. The offset feedforward current needed to satisfy these assumptions could be of AMPA origin. However, no specific implementation for the synaptic current was introduced into the model to simplify the numeric algorithm. The feedback connections between neurons contained NMDA conductance of 4 $\mu S/cm^2$. For the surface area of the dendrite of about $1.2 \cdot 10^5 \mu m^2$ estimated from images obtained in Aksay, Baker, Seung, and Tank (2000) and the single-channel NMDA conductance of 50 pS (Lin, Skeberdis, Francesconi, Bennett, & Zukin, 2004), the value of NMDA conductance given above yields about 100 NMDA channels per recurrent connection. Since there are 40 recurrent connections per cell, we estimate about $4 \cdot 10^3$ NMDA channels per neuron localized in the recurrent synapses. For simplicity, our model did not include AMPA current in the recurrent synapses. The effects of the possible presence of AMPA currents in the recurrent synapses are discussed in the online supplement 1.[1] The somatic, dendritic capacitances and leak current (both somatic and dendritic) were taken to be 1 $\mu F/cm^2$, 0.5 $\mu F/cm^2$, and 0.1 mS/cm^2, respectively (Lisman et al., 1998; Koulakov et al., 2002). The details of the numerical implementation of our model are described in the supplementary material 3.

3 Results

Our model for the neural integrator is based on recurrent positive feedback (Rosen, 1972; Robinson, 1989; Seung et al., 2000a; Koulakov et al., 2002; Goldman et al., 2003). First, we present the results obtained for a simplified model, which can be solved exactly without the use of a computer. To make the exact solution possible, some assumptions have to be made about the recurrent network connectivity. One of the assumptions that we will make throughout this study is that the neurons are connected in the all-to-all fashion with equal weights (see Figure 2). In this case, all neurons receive the same input current, which greatly simplifies the analysis. This assumption about network connectivity is in contrast to the one made by Goldman et al. (2003), who considered feedback connectivity targeting specific dendritic compartments.

[1] Supplemental materials available online at http://www.mitpressjournals.org/doi/suppl/10.1162/neco.2008.12-06-416.

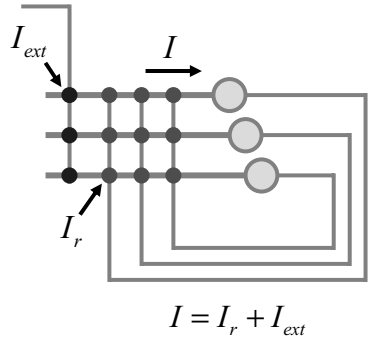

Figure 2: The recurrent feedback model, which is thought to underlie the neural integrator. Synapses and somata are shown by full and empty circles, respectively. In the fully connected network considered here, all the neurons receive the same input (I). The input is a sum of external and recurrent currents (I_{ext} and I_r, respectively).

3.1 Bistable Neurons.

3.1.1 The Case of No Recurrent Feedback. In this section, we consider the properties of integrator neurons without recurrent connections. The recurrent connections are included in section 3.2.

Figure 3A shows the response of the neuron that is used in this simplified model. The dark gray and light gray curves show the firing rate dependencies for increasing and decreasing inputs, respectively. If the input is in the range marked by the tailed arrow (region II in Figure 3A), the firing rate of the neuron may have two values depending on the prior history. The neuron is therefore bistable for this range of inputs.

The response of a neuron as a function of input current exhibits three regimes shown in Figure 3A. For the values of input current below the bistable regime (region I in Figure 3A), the neuron is deactivated (OFF) unconditionally. For the inputs above the bistable range (region III), the neuron is always active (ON). If the inputs are in region II, the neuron can be either ON or OFF depending on prior inputs.

We consider an ensemble of units that differ in two respects: hysteresis width and position. The former parameter is described by the half-width of hysteresis $\Delta = (\theta^{\uparrow} - \theta^{\downarrow})/2$, where $\theta^{\uparrow}$ and $\theta^{\downarrow}$ are thresholds for activation and deactivation, respectively. The position of hysteresis is described by the average of the two thresholds: $\theta = (\theta^{\uparrow} + \theta^{\downarrow})/2$, as illustrated in Figure 3A. The ensemble of many of such neurons is distributed in the 2D parameter space (Δ, θ) as shown in Figure 3B. The density of neurons in the parameter plane is assumed to be

$$\rho(\Delta, \theta) = C \exp(-\Delta/\overline{\Delta}). \tag{3.1}$$

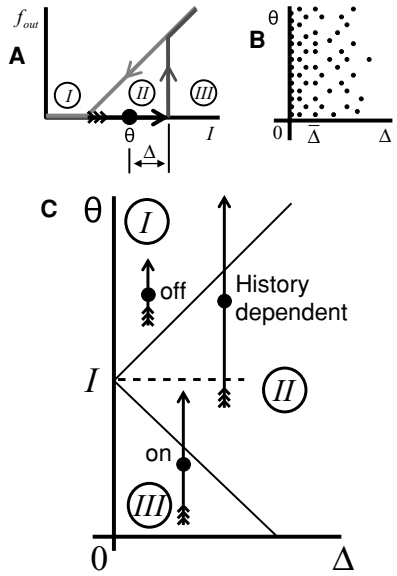

Figure 3: Ensemble of bistable neurons with differential parameters. (A) Example of the input to firing rate relationship for a bistable neuron. The activation (dark gray) and deactivation curves (light gray) do not coincide in the bistable range (tailed arrow). If the input current (I) to the neuron is in the bistable range, the firing rate can have two values depending on the previous history. The bistable range is shown by the arrow with a tail, with the threshold for activation denoted by the arrowhead. The threshold for deactivation coincides with the arrow's tail. Such neurons can be described by two parameters, θ and Δ, representing the mean position of the threshold (black dot) and the half-width of hysteresis. The unconditionally OFF, bistable, and unconditionally ON ranges of input are labeled by I, II, and III, respectively. (B) We considered an ensemble of many bistable neurons. Each neuron is represented by a point in the parameter space (Δ, θ). The density of neurons is uniform along the vertical θ-axis, while the density decays for larger values of the hysteresis half-width. (C) For a given value of the synaptic current on the inputs of all neurons (I, dotted line), the parameter space is divided into three regions. In the top region, the neurons are unconditionally OFF. This region corresponds to the range of inputs labeled by I in *A*. Thus for one of such neurons, whose bistable range is shown by the tailed arrow, the input level (dotted line) is below both thresholds for activation and deactivation, which implies that the neuron is unconditionally OFF. In the bottom region, labeled by III, all the neurons are unconditionally ON, since, as shown for another example neuron, the input is above the threshold for activation. In the middle region (labeled II), the input current is in the bistable range, as shown for one of the neurons. The firing state of these neurons is therefore history dependent, which implies that they can be either ON or OFF.

Here, C and $\overline{\Delta}$ are the maximal 2D density of neurons and the average half-width of hysteresis. We assume that $\Delta \geq 0$ throughout the study. Thus, only neurons with regular hysteresis are included in our model. The number of neurons in the square of parameter space with dimensions $d\Delta$ and $d\theta$ along the Δ- and the θ-axes, respectively, is given by $\rho(\Delta, \theta)d\Delta d\theta$ for a sufficiently small square. Thus, although ρ depends on only one coordinate, it is a 2D density of neurons. The 1D density along the θ-axis is constant $\rho(\theta) = C\bar{\Delta} =$ const. Although we adopted distribution (3.1) for concreteness, the analysis described below could be performed for an arbitrary distribution.

We now recall that in case of all-to-all connectivity, all neurons receive the same value of synaptic input current (see Figure 2). It is interesting therefore to consider properties of the neuronal ensemble when the same input is supplied to all of the neurons. In particular, it is of interest to determine what neurons are ON or OFF for a given value of input current. Clearly, an unambiguous answer to this question cannot be given. This is because for the given value of input current I, there are neurons that are in the bistable regime, that is, their state depends on their history.

For a given value of input current, all neurons can be divided into three groups: neurons that are unconditionally OFF, ON, and the history-dependent units. In Figure 3C, the areas occupied by these groups are marked by I, III, and II, respectively. For neurons in these areas, the input currents are in the ranges I, III, and II indicated in Figure 3A. For the neurons that are unconditionally ON (group III), the value of input current is above their threshold for activation, as follows from Figure 3A: $I \geq \theta^{\uparrow} = \theta + \Delta$. Therefore, such units are located in the region of the parameter space defined by the following condition:

$$\text{Group I} : \theta \leq I - \Delta. \tag{3.2}$$

The units that are unconditionally OFF (group I) receive input current, which is below their threshold for activation: $I \leq \theta^{\downarrow} = \theta - \Delta$. These units are therefore defined by another condition:

$$\text{Group II} : \theta \geq I + \Delta. \tag{3.3}$$

Finally, the neurons that are neither unconditionally ON nor OFF have a state that depends on history (group II). Their positions are defined by an alternative to conditions 3.2 and 3.3:

$$\text{Group III} : I - \Delta \leq \theta \leq I + \Delta. \tag{3.4}$$

Note that the positions of areas I to III depend on the value of input current I and therefore may change with time. By manipulating external current, one can form various patterns of activation and deactivation

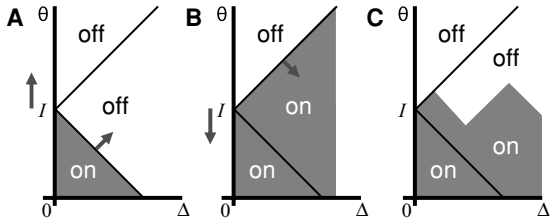

Figure 4: Possible configurations of active units (activation function $h(\Delta, \theta)$) in the absence of global recurrent feedback connections. The active areas are shaded. (A) If initially all neurons are OFF and the external current was increased from zero, only the units that are unconditionally ON (group I) are active. (B) If initially all units are ON and the external current sweeps the parameters space downward, there is a deactivation wave propagating with the boundary between regions II and III. The active region extends indefinitely in the direction of a large Δ and is truncated in this and following figures. (C) A more complex pattern of activation in the history-dependent region (III) can be produced by a complex pattern of inputs. For the profile shown, current was going up, down, up, and down.

in the parameter space, some of which are shown in Figure 4. These patterns do not depend on the distribution of units in the parameter space—function ρ in equation 3.2. The latter function defined the density of neurons independent of whether they are ON or OFF. It is therefore not history dependent. Another distribution has to be used to describe history dependence in the activation of hysteretic neurons.

We next define the activation function $h(\Delta, \theta)$. This function specifies whether a neuron at a point with coordinates (Δ, θ) is ON or OFF. It is equal to one in the areas occupied by active units and to zero in other areas. The total number of active neurons in the ensemble is determined by the sum of the product of the densities over the parameter space:

$$n(t) = \int h(\Delta, \theta, t)\rho(\Delta, \theta) d\Delta d\theta. \tag{3.5}$$

In this equation, the activation function h acts as a marker, which allows inclusion of only the areas occupied by the active neurons into the sum. This expression will allow us to calculate the recurrent current in section 3.1.2.

3.1.2 Recurrent Feedback and the Stability Condition. In the section 3.1.1, we studied the properties of a simplified model of hysteretic neurons with only feedforward connections present. Here we include the recurrent connections the model. To this end, we assume that the recurrent current I_r is

proportional to the total number of active neurons n given by equation 3.5. We therefore neglect the variations in the recurrent current due to changes in the neuronal firing rates assuming synaptic saturation (Seung, Lee, Reis, & Tank, 2000b; Koulakov et al., 2002). This approximation is valid if a receptor with a large time constant, such as an NMDA receptor, is responsible for neurotransmission in the recurrent synapses (Seung et al., 2000b; Koulakov et al., 2002). The long time constant of NMDA receptors leads to saturation of synaptic currents even at small firing rates (10–20 Hz), implying little dependence of the recurrent current on the firing rates. The saturation at low firing rates may also be of a presynaptic nature, arising from synaptic depression.

We now address the dynamics of our model in the case of recurrent connections that are present. We will assume here that each active neuron contributes I_0 to the recurrent current. Therefore, the total recurrent current into each neuron can be found as a product of the number of active neurons and parameter I_0:

$$I_r(t) = I_0 n(t - \tau_s). \tag{3.6}$$

Note that the recurrent current is related to the number of active neurons n with a synaptic delay τ_s. This statement is justified in appendix A. If the NMDA receptor is a primary neurotransmitter in the recurrent synapses, one should expect synaptic delay to be $\tau_s \simeq 100$ msec.

To complete the description of the simplified model, we introduce the total value of the input current for each neuron (see Figure 2):

$$I(t) = I_r(t) + I_{ext}(t). \tag{3.7}$$

Here I_{ext} is the external "command" input. The new value of input current $I(t)$ determines the neuronal activation function for the new time step $h(\Delta, \theta, t)$ through the set of inequalities 3.2 to 3.4. The activation function through equation 3.6 leads to a new value of recurrent current at the next time step $t + \tau_s$. Thus, the system of equations 3.2 to 3.7 allows accounting for iterative dynamics of the system of hysteretic neurons connected by recurrent synapses. This dynamic is illustrated below with a series of examples.

We first discuss the response to a tonic external input [$I_{ext}(t) = const$]. Let us assume that the current is positive ($I_{ext} > 0$). It is expected, then, that under certain conditions, which become evident below, the input is integrated temporarily, implying that the total current in the system increases with time. We then expect a wave of activation, similar to shown in Figure 4A, to propagate upward in the parameter space. We now discuss the equations governing the propagation of this wave and the conditions of its existence.

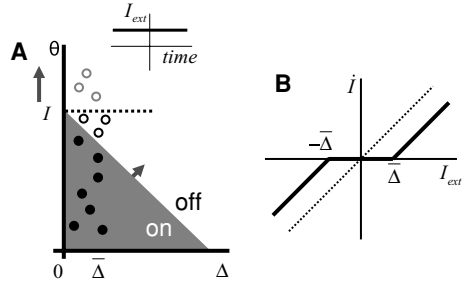

Figure 5: Response to tonic input. (A) For nonzero external input that is constant in time (inset), the current in the system (I) may increase as a function of time (see B for the condition of this). In this case, the number of active neurons (full circles) is given by equation 3.8. The inactive units are shown by open circles. (B) For stationary external input, the system's response increases with time if the value of input exceeds a threshold equal to $\overline{\Delta}$. The rate of increase ($\dot{I} \equiv dI/dt$) is proportional to the external current above threshold.

Given the value of total current I, the number of active units is approximately given by the product of the area occupied by neurons under the dotted line in Figure 5A, $I\overline{\Delta}$, and the concentration of neurons in the parameter space C (see equation 3.1): $n = CI\overline{\Delta}$. This is assuming that the value of current I exceeds substantially the average hysteresis width $\overline{\Delta}$. A small correction has to be made to subtract the neurons, represented by open circles in Figure 5A. The corrected expression for the number of neurons in the ON state is

$$n(t) = C\left[I(t) - \overline{\Delta}\right]\overline{\Delta}. \tag{3.8}$$

This equation can also be obtained directly from equation 3.5. As mentioned before, in deriving equation 3.8, we assumed that $I(t) >> \overline{\Delta}$, which allowed us to neglect contributions to equation 3.8 proportional to $\exp(-I(t)/\overline{\Delta})$. We use this approximation in all subsequent derivations.

The new value of current at the next time step is, according to equations 3.6 to 3.8,

$$I(t + \tau_s) = \alpha I(t) + I_{ext} - \alpha\overline{\Delta}, \tag{3.9}$$

where we introduced the tuning parameter $\alpha = I_0 C\overline{\Delta}$. The "perfect integrator" condition corresponds to the value of parameter $\alpha = 1$. In this case, the current is accumulated according to equation 3.9 without a loss:

$$I(t) = \left(I_{ext} - \overline{\Delta}\right)t/\tau_s + I(t = 0). \tag{3.10}$$

The quantity being accumulated is actually $I_{ext} - \overline{\Delta}$. The system is therefore capable of acting as a temporal integrator. Note that equations described above are valid only on the temporal scales exceeding τ_s. To describe the behavior of the system on shorter timescales requires a more precise model of synaptic dynamics.

The sustained integration is possible only if external current exceeds the average value of hysteresis: $I_{ext} > \overline{\Delta}$ for $\alpha = 1$. For $I_{ext} < \overline{\Delta}$, the current would have to decrease according to equation 3.9, which is not valid in this case. Indeed, in deriving equation 3.10, we assumed that there is a wave of activation propagating up in the parameter space (see Figure 5A). Because we found that the current would have to decrease for the case $I_{ext} < \overline{\Delta}$, the wave that persistently propagates in the positive direction is not self-consistent in this regime. Therefore, sustained increase in the input synaptic current I is not possible if $I_{ext} < \overline{\Delta}$. Consequently, the value of external input current of $\overline{\Delta}$ represents a threshold for integration. Similarly, only the negative inputs below $-\overline{\Delta}$ can be integrated in a sustained manner, which results in a negative threshold for integration. The rate of change in the synaptic current $\dot{I} \equiv dI/dt$ as a function of external input is summarized in Figure 5B. This figure shows that if the external input is between $-\overline{\Delta}$ and $\overline{\Delta}$, it is not integrated in a sustained manner. This statement is valid for an arbitrary distribution of active units in the parameter space ρ. That is, if, instead of an exponential distribution given by equation 3.1, any other distribution of hysteresis widths is found, the threshold for sustained integration is equal to the average value of hysteresis $\overline{\Delta}$. The issue of threshold for integration is discussed in sections 3.4 and 3.5.

We now discuss the case of zero external input, which for the oculomotor VPNI corresponds to eye fixations. It turns out that our model displays more sophisticated behaviors in this stationary condition rather than in the nonstationary one. The possible distributions of active units are shown in Figure 6. We start from the simplest case of the activation function shown in Figure 6A. In this case, all the neurons that have the medial threshold θ below the present value of input current I are active. The interface between the ON and OFF neurons is a straight line parallel to the Δ-axis. It is not so difficult to see that this distribution function cannot be realized using just the network architecture, with all-to-all recurrent connections and the same synaptic input to all neurons. This is because in this architecture, the boundaries separating ON and OFF neurons form 45 degree angles with the parameter axes (see Figures 4 and 5). Nevertheless, we consider the activation function in Figure 6A because it gives insights into more complex cases (see Figures 6B and 6C). Assume that by manipulating each neuron individually we set up the activation function in Figure 6A. What is the condition needed for it to remain stable as a function of time?

To answer this question we need to repeat the calculations, which led us to equation 3.9. The number of active units as a function of total synaptic

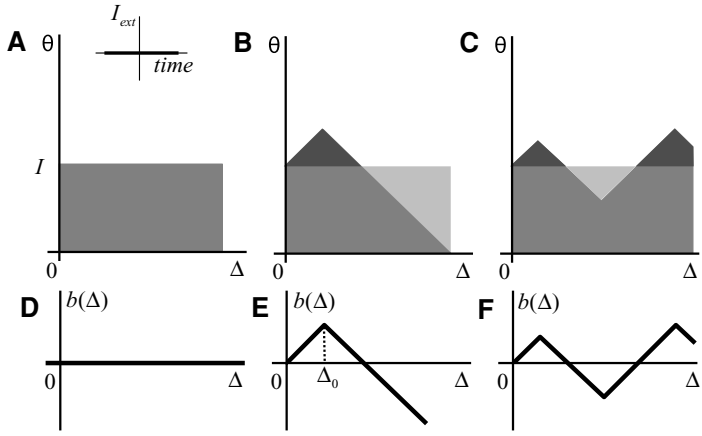

Figure 6: The case of zero external input (inset). (A–C) Possible stationary configurations of the activation function. The areas shaded by dark(light) gray lead to an increase(decrease) in the number of active neurons with respect to A. Stable configurations are achieved when the dark and light gray areas have the same number of neurons. (D–F) The boundary functions $b(\Delta)$ corresponding to the activation profiles in A–C.

current is $n(t) = C\,I(t)\overline{\Delta}$ (see the text preceding equation 3.8). The value of the total synaptic current I at the next time step $t + \tau_s$ is equal to the total recurrent current since $I_{ext} = 0$:

$$I(t + \tau_s) = \alpha I(t). \tag{3.11}$$

For the configuration in Figure 6A to remain stable, one has to satisfy the condition of stationarity of the synaptic current: $I(t + \tau_s) = I(t)$. In view of equation 3.11, this is equivalent to setting $\alpha = 1$, that is, having a perfectly tuned integrator. We conclude that the activation function in Figure 6A is stable for a perfectly tuned integrator ($\alpha = 1$). We assume the perfect integration for the remainder of this section. Some solutions for a nonperfect integrator will be given in the next section.

What other stable activation functions are possible? One can generate stable activation functions from the previous example (see Figure 6A) using the following prescription. First, one modifies the activation function in Figure 6A in such a way that the number of active units is unchanged. Thus, Figure 6B shows the activation function for which the number of newly recruited neurons (dark gray) is compensated by the deactivated ones (light gray), so that the total number of active neurons is the same as in Figure 6A. Second, one ensures that the intersection of the boundary between ON and OFF neurons with the vertical axis θ should be the same. Near this

intersection, neurons have very small hysteresis ($\Delta = 0$), and therefore the position of this intersection corresponds to the total synaptic current I. The total current is not perturbed according to the first requirement because the number of active neurons is unchanged. The first requirement (the change in the number of ON neurons with respect to Figure 6A Δn is zero) amounts to

$$\Delta n = \int_0^\infty b(\Delta)\rho(\Delta)d\Delta = 0. \tag{3.12}$$

The boundary function $b(\Delta)$ is the shape of the interface between the active and inactive neurons, which is set to be zero at $\Delta = 0$ (see Figures 6D–6F). This function describes the deviation of the shape of the interface from the case shown in Figure 6A. The boundary function is positive in the dark gray areas in Figure 6 and is negative in the light gray areas. Equation 3.12 implies that an increase in the number of active units in the dark gray areas in Figures 6B and 6C, where $b(\Delta) > 0$, is compensated by the decrease in the light gray areas ($b(\Delta) < 0$), thus leading to no overall change in the total number of active units. If $\Delta n = 0$, the recurrent current is the same as in Figure 6A, leading to the stable configuration. We refer to condition 3.12 as the *stability condition*. We will show how this condition leads to the reversal of the sign of hysteresis in the firing rate as a function of eye-position dependence.

The parameter Δ_0 introduced in Figure 6B determines the position of the maximum of the function in Figure 6B (the wedge in the activation function). This parameter can be found from equation 3.12. In appendix B we calculate Δ_0 for the exponential distribution of cellular hystereses, equation 3.1:

$$\Delta_0 = \overline{\Delta} \ln 2. \tag{3.13}$$

This relationship is valid only for the exponential distribution of the hysteresis widths. For a distribution different from exponential, a different coefficient of proportionality between Δ_0 and $\overline{\Delta}$ is expected. Thus, previously we considered the ensemble of neurons with the same values of hysteresis width (Koulakov et al., 2002). This ensemble is defined by the distribution that replaces equation 3.1:

$$\rho(\Delta, \theta) = C\overline{\Delta}\delta\big(\Delta - \overline{\Delta}\big), \tag{3.14}$$

where δ is the Dirac delta function. The stability condition 3.12 can also be used to calculate the parameter Δ_0 for this distribution (see appendix B):

$$\Delta_0 = \overline{\Delta}/2. \tag{3.15}$$

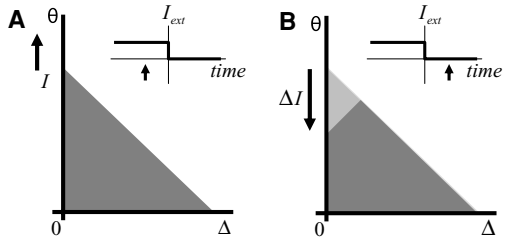

Figure 7: Response to input current that suddenly drops to zero (insets). The black arrows in the insets indicate the moment of time when this activation function is expected.

Thus, although the numerical coefficients for exponential and delta function distributions differ between equations 3.13 and 3.15, the value of parameter Δ_0 is determined by the average hysteresis width, $\overline{\Delta}$.

Finally, stability condition 3.12 allows finding stable parameters of more complex configurations, such as the one shown in Figure 6C.

We will now describe the transitional regime between integration ($I_{ext} > \Delta_0$, eye movement) and the stable configuration described in the previous paragraph ($I_{ext} = 0$, eye fixation). With the external current present and larger than the integration threshold $\overline{\Delta}$, the activation function is described by the wave of activation propagating in the parameter space (see Figure 7A, which replicates Figure 5A), which means that the eye position is increasing. Assume that the external current is suddenly removed (see Figure 7A, inset). The activation function in Figure 7A is formed during eye movements and does not satisfy the stability condition. As such, it cannot exist during eye fixations. The activation function has to evolve to one of the stable configurations, such as that shown in Figure 6. In the simplest case, the activation function evolves to the configuration in Figure 6B, also shown in Figure 7B. As a result, the recurrent current I_r drops after the removal of external current in the direction opposite to the previous eye movement. The amount of such a recurrent current drop ΔI is

$$\Delta I = 2\Delta_0 = 2\overline{\Delta} \ln 2. \tag{3.16}$$

This drop is given here with a positive sign despite the fact that the current was decreasing after the disappearance of the external input.

To calculate the firing rate as a function of eye position during eye fixations, we will assume here that the eye position is proportional to the recurrent current. During eye fixations, the external input is absent: $I_{ext} = 0$. The eye position is therefore equal to the total synaptic current I

(see equation 3.8). Ignoring the proportionality constant, we assume that during fixations, the eye position is equal to the input synaptic current:

$$E = I. \tag{3.17}$$

The problem of finding the firing rate as a function of eye position is therefore seemingly simple: it is to determine the response of a neuron as a function of input current. It may appear that this problem is already solved in Figure 3A, which postulates neuronal response to the input current as one of the assumptions of our model. Indeed, in regions I and III (unconditionally OFF/ON), nothing can change the response postulated in Figure 3A. However, in the bistable region II, the firing rate can follow one of the dependencies—either ON or OFF. The problem is therefore to determine what branch of the firing rate dependence is followed after ON and OFF saccades.

Figure 8 demonstrates qualitatively the reversal of the hysteresis sign in the case of neurons connected by recurrent feedback. Figures 8A and 8C show the activation function for a neuron whose location in the parameter space is illustrated by the black dot. For the same value of the input current and, consequently, the same eye position (see equation 3.17), this neuron will be in the active state after the ON saccade (see Figure 8A). After the OFF saccade, the same neuron is expected to be OFF (see Figure 8C). The firing rate of this neuron is therefore expected to be higher after the ON saccade, as illustrated in Figures 8B and 8D. This behavior is due to the drop of the recurrent current after the eye movement in the ON direction (increase after the OFF saccade), as discussed in the previous paragraph. The presence of the global positive feedback therefore leads to the reversal of the sign of hysteresis for some neurons in the network in agreement with experimental findings.

We next examined, quantitatively, the dependence of the firing rate as a function of eye position in our model. We used the following method. Suppose that one has to calculate the firing rate after an ON saccade $f_{ON}(E)$. We evaluate the state of the neuron with the total input $E + \Delta I$ in the ON state (dark gray dependence in Figure 3A) and subsequently follow the OFF firing rate dependence (light gray in Figure 3A) from input $E + \Delta I$ to E. This latter operation reproduces the drop in the total recurrent current after the end of a saccade. A similar procedure is followed for the OFF saccades. This method is illustrated in Figures 8B and 8D. The results of systematically applying this method are presented in Figure 9.

Neurons can be separated into four groups with qualitatively different behaviors of the firing rates as functions of eye position. Figures 9A and 9B show the regions in the parameter space occupied by these groups with the corresponding dependencies of the firing rates on the eye position displayed in Figure 9C. For neurons with a large value of endogenous

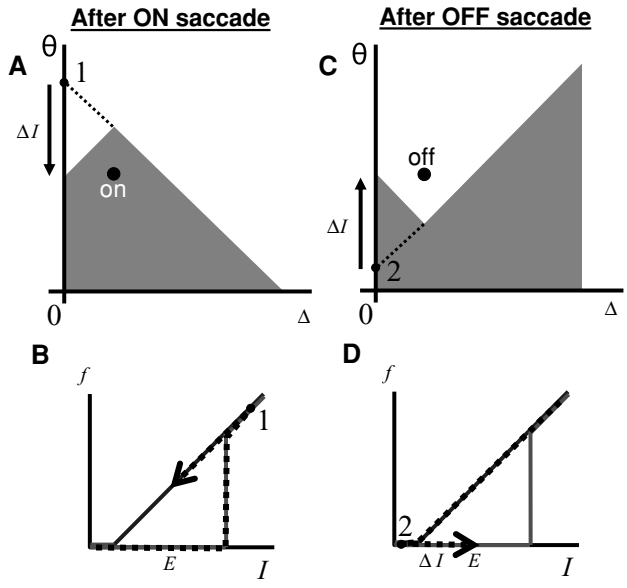

Figure 8: The reversed direction of hysteresis in the model with differential parameters. The response of the neuron indicated by the black dot after the ON saccade (A) is higher than after the OFF saccade (B). This is because this neuron is ON/OFF in these states as indicated. (B, D) Calculation of the firing rate as a function of eye position during fixations used to generate Figure 9. For the same eye position E after ON saccade (A and B), the firing rate is larger than after the OFF saccade (C and D). This is in agreement with experimental findings (see Figure 1).

hysteresis (see group G4), a regular direction of hysteresis is observed—the same as the endogenous one. This is not surprising, because the endogenous hysteresis is so strong for these neurons that no drop in the recurrent current can reverse it. The neurons with smaller values of endogenous hysteresis (parameter Δ), which belong to groups G1 and G2 in Figure 9A, display a reversed hysteresis as follows from the qualitative argument illustrated in Figure 8. Finally, a small group of neurons, G3, shows no hysteresis at all. We call these neurons marginal. The marginal neurons are important in establishing a relationship with our previous results (Koulakov et al., 2002), as explained below.

It is of interest to calculate the relative numbers of neurons belonging to different groups. This can be accomplished by using the conditions for areas occupied by different groups of neurons summarized in Figure 9A and equation 3.1. Thus, for the exponential distribution of neurons in the parameter space, G1 contains 1/2 of all neurons, while G2 and G4 contain

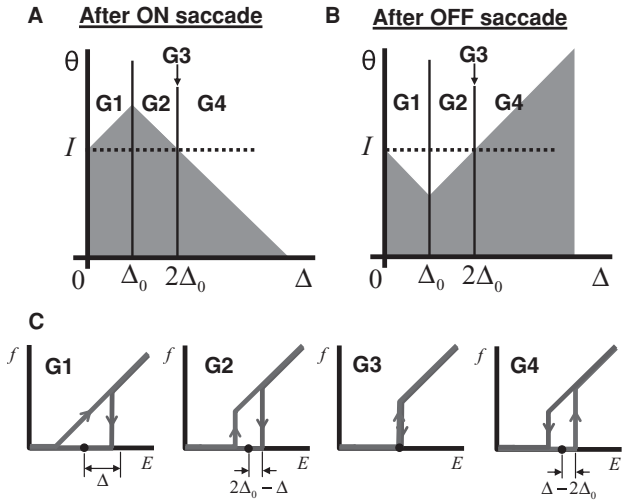

Figure 9: The firing rate as a function of eye position during fixations for different groups of neurons (G1–G4). (A, B) The activation functions after the ON and OFF saccades, respectively. (C) Firing rate as a function of eye position for neurons from four groups (G1–G4) indicated in *A*. The dark and light gray curves correspond to the preceding ON and OFF saccades, respectively.

a 1/4 fraction each. Therefore, 3/4 of the neurons (G1 and G2) should demonstrate a reversed hysteresis, while for 1/4 of the neurons (G4), the sign of the hysteresis should remain unchanged. These fractions are specific to the exponential distribution of the endogenous hysteresis $\rho(\Delta)$ given by equation 3.1. For a different-from-exponential $\rho(\Delta)$, a different set of relative abundances of neurons in the groups is expected. Although the relative fractions of different groups of neurons (G1 versus G2) depend on the distribution of hysteresis widths, equation 3.1, the firing rate dependences shown in Figure 9C do not. For a different distribution, the neural responses as a function of eye position are exactly the same in this model.

Group G3 contains neurons with no hysteresis, despite the fact that without global feedback, they should display a bistable response as in Figure 3A. This group is defined by the condition that their hysteresis half-width Δ is equal to $2\Delta_0$. Since this group resides on the interface between G2 and G4, the number of neurons in this group is vanishingly small for the exponential distribution of Δ. This, however is not true for the case when all neurons have the same value of Δ, such as in the case of delta function distribution 3.14. In this case, all the neurons belong to the marginal group. This is because the position of these neurons in the parameter space $\Delta = 2\Delta_0$ determined by equation 3.15 coincides with the definition of the marginal neurons G3 in Figure 9A. Therefore in this case,

one should expect no hysteresis in the firing rate as a function of eye position displayed by all neurons. This conclusion applies to neurons in our previous study (Koulakov et al., 2002).

3.1.3 A More Realistic Implementation. We then investigated if our simplified theory can apply to a more realistic model for a neural integrator involving a biophysically plausible implementation of neurons and synaptic conductances. We implemented our model with 40 two-compartmental neurons. Each neuron is represented by two compartments: somatic and dendritic (see Figure 10A). The somatic compartment contains sodium and potassium conductances for the generation of action potentials. The firing frequency depends on the overall input into the dendritic compartment. The dendritic compartment is capable of generating membrane voltage-based bistability due to the nonlinearity of NMDA conductances. Since the amount of NMDA conductance is different for different cells (see section 2) the cells display a hysteretic input-output relationship, with the width of hysteresis varying from cell to cell, similar to the simple model (see Figure 11A).

We then connected the cells by the recurrent feedback. The strength of all recurrent synapses is the same, which corresponds to the all-to-all connectivity with roughly the same recurrent current on the input to all neurons (see Figure 10B). When these cells are connected by recurrent feedback, the network can integrate a transient input current (see Figures 10C–10E) as a function of time.

As predicted by the simple model, the neurons fall into four categories: with little or no hysteresis (see Figure 11B), inverted hysteresis (see Figure 11C), marginal neurons (see Figure 11D), and regular direction of hysteresis (see Figure 11E). These classes correspond to the dependences derived in the simplified model and illustrated in Figure 9C. We have therefore shown that the conclusions of the simplified model sustain the test by a more realistic implementation.

3.2 Multistable Neurons. We next examined the behavior of the recurrent system of neurons that themselves exhibit multistability. Cellular multistability is assumed to emerge from the bistable properties of several dendritic compartments of the same neuron (see Figure 12A). We demonstrate here that under certain conditions, the recurrent network of multistable neurons may be mapped mathematically onto the system of bistable neurons considered above. The properties of the multistable neurons in the network are therefore similar to the behaviors of bistable neurons, including the history-dependent profiles of the activation function (see Figure 9) and the reversal of the sign of hysteresis.

We first consider the properties of multistable neurons in isolation, without recurrent connections present. Each neuron includes several dendritic compartments (see Figure 12A). The dendritic compartments are assumed

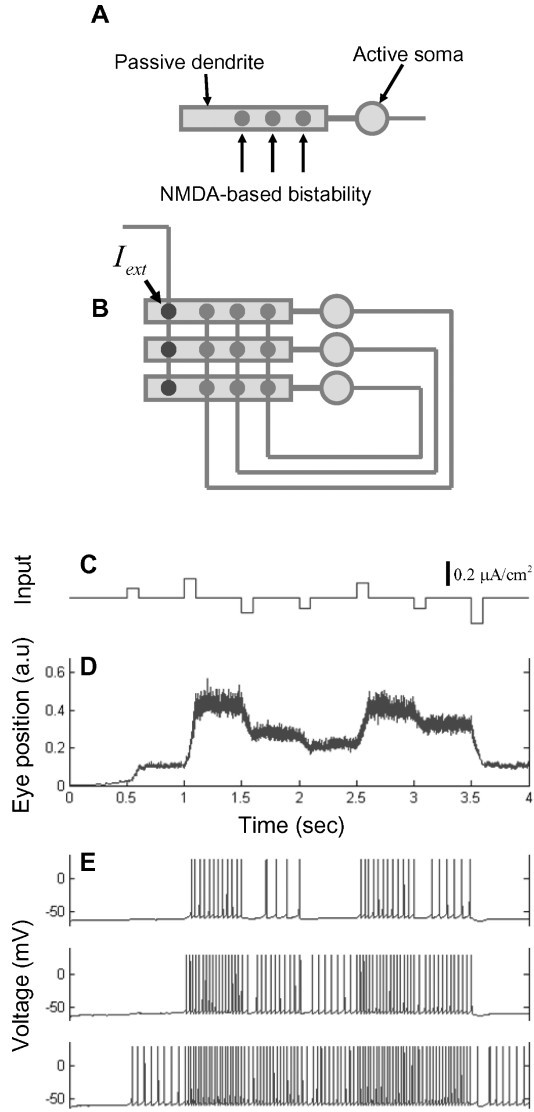

Figure 10: Model with Hodgkin-Huxley neurons. (A) Each neuron has a dendritic compartment that is bistable and a soma that generates action potentials. To generate bistability, the dendrite receives tonic NMDA current. (B) Neurons are connected into the network by all-to-all global feedback that generates additional NMDA current into each dendrite. The network contains 40 neurons, only 3 of which are shown. (C) An example of external input supplied to every dendritic compartment. (D) Resulting changes in the average feedback current reflect the integral of input. This variable could therefore be associated with the eye position. (E) The membrane voltage for three example neurons.

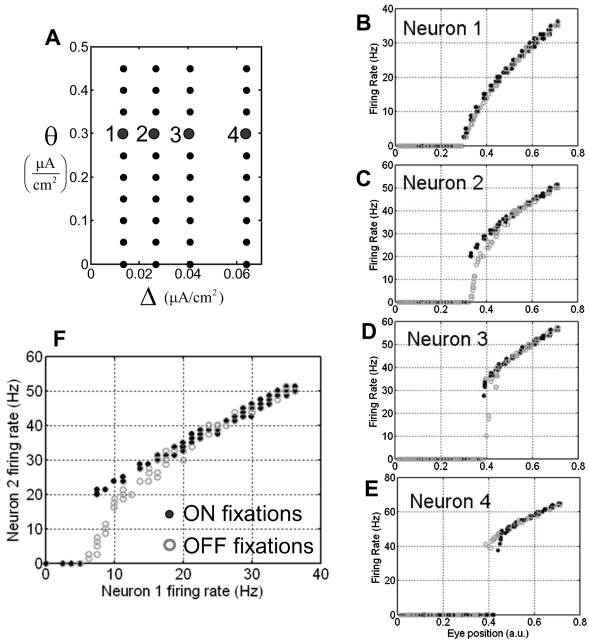

Figure 11: Reversal of hysteresis in the network of Hodgkin-Huxley neurons. (A) The arrangement of 40 neurons in the parameter space of hysteresis positions and half-widths. (B–E) The firing rate traces for four example neurons indicated in *A*. The black/gray markers display values obtained after ON/OFF saccades. (B) For some neurons, the observed hysteresis is small, similar to some experimental observations (see Figure 1A). (C, D) For neurons 2 and 3, the black dependence is above the gray one. These neurons belong to groups G1 or G2 defined in Figure 9 and display reversed hysteresis (compare to Figure 1B). (E) Neuron 4 belongs to G4. (F) Firing rate of one neuron versus the other (*C* versus *B*) displays hysteresis, as observed experimentally (Aksay et al., 2003).

to be electrotonically isolated from each other. This implies that the response of each dendritic compartment is independent of the state of other compartments in the same dendritic tree and on the firing activity of the cell's soma. Each of the compartments is assumed to be bistable. The bistable ranges of the compartments are assumed to be distributed over a large range of values with no substantial difference in the hysteresis width. These assumptions lead to the dependence of the firing rate on the input current in the form of a staircase (Koulakov et al., 2002; Goldman et al., 2003), as discussed below. This dependence is illustrated in Figure 12B.

When the cell receives input current that increases, the dendritic compartments are sequentially switched on, and the firing rate of the cell

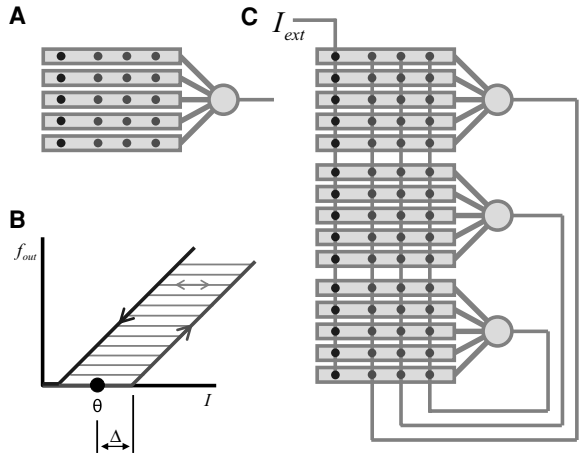

Figure 12: Model with multistable neurons. (A) Neuron with multiple bistable dendritic compartments. Synapses are shown by full circles. (B) Firing rate as a function of input current for such a neuron is multistable: each stable firing state corresponds to the fixed number of active dendritic compartments. Introducing the parameters θ and Δ allows us to map the problem mathematically to the previous model with bistable neurons. (C) The recurrent network of multistable neurons with all-to-all connections that can be solved using this method.

increases following the light gray line in Figure 12B. When the current decreases, the dendritic compartments are switched off, leading to a decrease in the firing rate (dark gray line in Figure 12B). The space between the light gray, and dark gray lines corresponds to the bistable range of the dendritic compartments. This implies that none of them will change their state, and therefore the firing rate is preserved in the region between the dark gray and the light gray lines. This is signified by the horizontal segments in Figure 12B. The bistability in the behavior of dendritic compartments results in the hysteretic loop in the response of the cell (Lisman et al., 1998; Koulakov et al., 2002). Note that the activation line (light gray) is below the deactivation line (dark gray), which means that for an isolated cell, the hysteresis has a regular sign (cf. Figure 3). An additional twist is provided by the states that are accessible inside the hysteretic loop. We will argue that the multistable neurons with these response properties connected into a recurrent network (see Figure 12C) display reversed hysteresis through the mechanism similar to that previously discussed.

We now consider the multistable neurons connected into an all-to-all network (see Figure 12C). As in our previous model, we assume that all neurons receive the same value of input current composed of the external and recurrent currents. The latter component is determined by the number

of neurons that are currently active. We assume, similar to the model with bistable neurons, that the neuron is active if the firing rate of this neuron is above zero. This assumption leads to the similarity between the model with multistable and bistable neurons (see section 3.1). Indeed, due to the saturation of slow NMDA conductance or synaptic depression, the recurrent NMDA current varies weakly as a function of the firing rates of the presynaptic neurons. Therefore, the exact form of the firing rate dependence is not important for the network current. Thus, the neurons shown in Figures 3A and 12B are not distinguishable from the viewpoint of other neurons in the network. In particular, we can describe each multistable neuron by two parameters: the median hysteresis position θ and half-width Δ (see Figure 12B). The neurons then can be arranged on the 2D parameter plane similar to the bistable case. Our conclusions about the dynamics of activation function can be transferred from the bistable case to the multistable case without any modification. Therefore, as far as the properties of the entire network are concerned, two networks based on bistable and multistable neurons are indistinguishable. The differences between the networks emerge when the firing rate of individual neurons is determined from the history dependence of the recurrent current. We illustrate this point next.

Consider a multistable neuron in the network. During the saccade in the ON direction, the firing rate of this neuron was increasing according to the light gray dependence (see Figure 13B). When the eye movement comes to conclusion, the external input driving it terminates. The moment of termination of the external input is shown in Figures 13A and 13B by point 1. Establishment of the eye position during fixations leads to a recoil in the recurrent current in this model by the amount equal to $2\Delta_0$, similar to the case of bistable neurons. From the point of view of the single neuron in the network, the input current is decreased by $2\Delta_0$ leading to the transition from point 1 illustrated in Figure 13B by the dotted arrow. The newly established firing rate during eye fixation is shown by the tip of the arrow. After the OFF saccade, the removal of external driving input leads to a symmetric increase in the recurrent current given by $2\Delta_0$ (see Figure 13C). The corresponding firing rate for this neuron after the OFF saccade is represented by the tip of the dotted arrow in Figure 13D. Notice that for the same eye position in Figures 13B and 13D, the firing rate is higher after an ON saccade than after the OFF saccade. This is in contrast to the response of the neuron that is not connected by the feedback (see Figure 12B). Thus, the hysteresis is expected to be reversed for this neuron due to the presence of network feedback.

Through systematic application of the geometric procedure illustrated in Figures 13B and 13D to all possible groups of neurons in the parameter plane of the system, we obtained the dependence of the firing rate on the eye position illustrated in Figure 14. This figure shows that the neurons with substantially small hysteresis, that is, from groups G1 and G2, display inverted hysteresis as suggested by the qualitative analysis in Figure 13. For

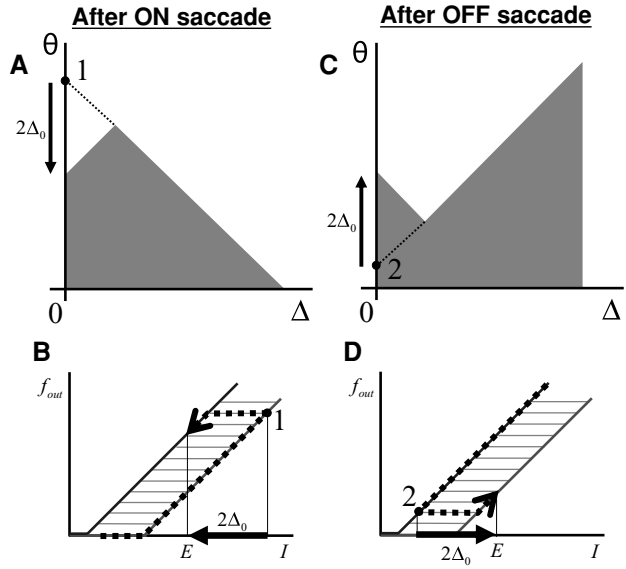

Figure 13: Illustration of the calculation of the firing rate as a function of eye position during fixations. For the same eye position E after ON saccade (A, B) the firing rate is larger than after the OFF saccade (C, D). This is similar to experimental findings (see Figure 1).

an inverted hysteresis, the native hysteresis width Δ has to be smaller than twice the recoil in the recurrent current $2\Delta_0$. For neurons with strong native hysteresis from group G4, a regular sign of hysteresis is expected. These conclusions are similar to the results of the model with bistable neurons (see Figure 9).

The important feature displayed in Figure 14C is that the values of response inside the hysteretic loop are not accessible during fixations. This observation is an artifact of considering the saccades of large amplitude, which are independent of each other. If the actual eye movement included many smaller saccades, the values inside the hysteretic loop in Figure 14 would be possible. Thus, one could make an experimental prediction that in the experimental data in Figure 1, smaller-amplitude saccades result in the responses near the center of the hysteretic loop.

3.3 Mistuned Integrator. In the previous sections, we discussed the properties of the perfectly tuned integrator. In this case, the current needed to recruit an additional neuron was exactly equal to the increase in the recurrent current resulting from recruiting this neuron. This condition is sufficient for many states of the system to be equally stable—that is, for the system to be multistable. The condition was quantitatively described

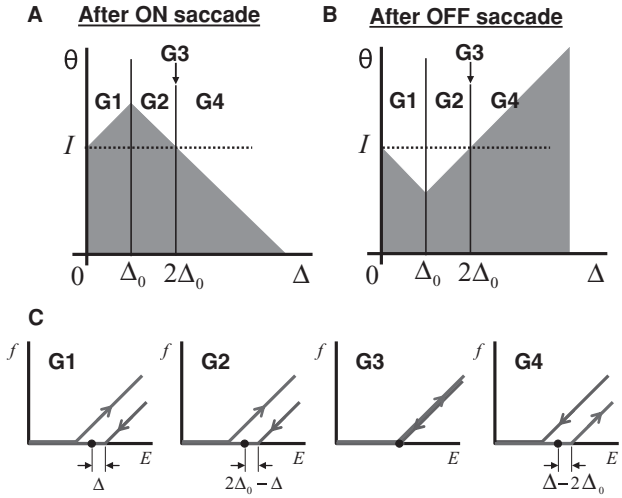

Figure 14: The firing rate as a function of eye position during fixations. The separation of the parameter space into areas (A, B) is similar to the bistable case (see Figure 9). (C) The response is given by the inverted hysteretic loop. These dependencies are obtained for the case of large saccades. For smaller saccades, the values of the firing rate inside hysteretic loops are accessible, which is one of the experimental predictions of this model.

by setting the parameter $\alpha = I_0 C\overline{\Delta}$ to 1. Here I_0 describes the contribution to the recurrent current from a single neuron and $C\overline{\Delta}$ is the inverse spacing between neuronal thresholds for activation, representing the increase in the recurrent current needed to activate an extra neuron. What if the recurrent feedback strength is weaker ($\alpha < 1$) or stronger ($\alpha > 1$)? We find that for substantial deviations from the perfectly tuned condition ($\alpha = 1$), stable eye fixations are also possible. Our model displays the same degree of robustness as simpler hysteretic systems (Koulakov et al., 2002; Goldman et al., 2003). The activation function during fixations is displayed for the case of weak feedback in Figure 15. Although the fixation is stable, the recoil in the recurrent current after the ON saccade is larger than the increase after the OFF saccade (see Figure 15A versus Figure 15B). The activation function therefore loses its symmetry between the ON and OFF saccade cases pertinent to the perfectly tuned integrators (see Figures 9 and 13).

3.4 Subthreshold Inputs. Our model includes some neurons that have no hysteresis ($\Delta = 0$). At the same time, we have concluded that sustained integration requires the external input exceeding a certain threshold equal to the average value of hysteresis for the ensemble (see Figure 5). The natural question is whether the neurons with no hysteresis can somehow integrate even a subthreshold input ($I_{ext} < \overline{\Delta}$).

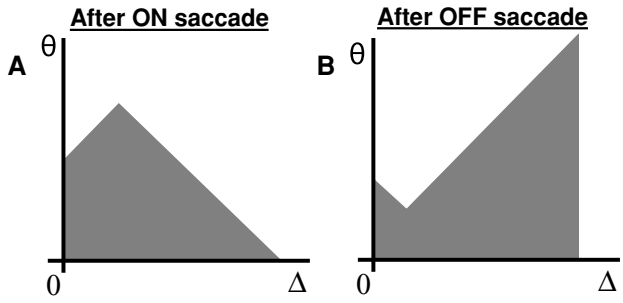

Figure 15: Robustness of a mistuned integrator. Mistuning $\alpha = I_0 C\overline{\Delta} \neq 1$ (defined in equation 3.9) leads to the asymmetric activation function after the ON and OFF saccades. The activation functions shown correspond to $\alpha < 1$, the weak feedback.

In the model described above, the subthreshold inputs can indeed induce persistent changes in the activation function and the firing rates of integrator neurons. It is true, however, that these changes will not depend on the duration of the stimulus. They depend only on the stimulus amplitude. Thus, persistent integration is indeed not possible, which means that the longer stimulus will not induce larger changes in the number of active neurons. This statement is valid for the model with no noise, that is, with no spontaneous switching of the bistable units. We demonstrate below that the presence of a finite switching time will lead to the leaky integrator and will make sustained integration possible even for subthreshold inputs.

Consider the state of the activation function during fixation after an ON saccade (see Figure 16A). A small positive input ($0 < I_{ext} < \overline{\Delta}$; see Figure 16B) leads to no sustained changes in the activation function. This implies that when the input is extinguished (see Figure 16C), the activation function is the same as before the stimulus (see Figure 16A). This is because the stability equation, 3.12, has only one solution in this case, as shown in Figures 16A and 16C. However, the negative inputs ($I_{ext} > 0$) do lead to a sustained decrease in the eye position (see Figures 16D to 16F). This partially justifies the intuition that the units with small hysteresis can react to stimuli of small amplitude. Of course, this intuition fails for positive inputs (see Figures 16A to 16C) since they do not lead to sustained changes in the integrator state.

3.5 Integration Time Constants. In this section we study the effect of spontaneous transitions in the bistable neurons on the stability of the integrator as a whole. So far we have assumed that the bistable or multistable neurons possess the property of perfect memory; there are no transitions between different states for these neurons. Thus, in the case of bistable neurons, we assumed that the neuron will remain in the ON or OFF state for a

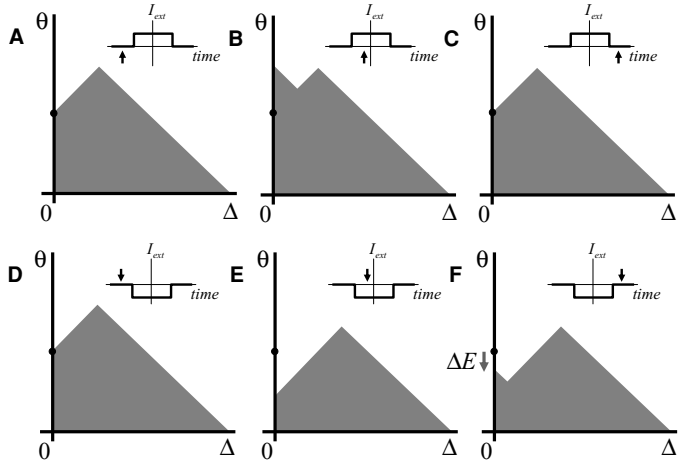

Figure 16: Response of the integrator to the subthreshold inputs $I_{ext} < \overline{\Delta}$ is asymmetric. After the ON saccade, the inputs in the same direction lead to no sustained changes in the activation function (A–C). The inputs in the opposite direction result in a sustained decrease in the eye position (D–F) shown by the gray arrow. The initial eye position and the moment of time are indicated by the full circle and the arrow in the inset.

very long time, even though a transition to the other state is possible. In the case of multistable neurons, we assumed that the state of bistable dendritic compartments is preserved, which leads to the infinite in time retention of the value of the firing rate in the absence of external inputs. In this section, we examine the effects of noise-driven transitions between different memory states. We address the behavior of bistable neurons for definiteness. We will derive the connection between the decay time constant of the bistable units and the integrator leak time.

In the case of perfect memory, the activation function $h(\Delta, \theta)$ can take two values, 1 and 0, corresponding to the ON and OFF states, respectively. When the spontaneous noise-driven transitions are possible, the activation function may take values between 0 and 1. This is because the activation function defines the average state of neurons in the small neighborhood of a point in the parameter space. If spontaneous transitions are allowed, the average value can deviate from pure values of 0 and 1. The activation function $h(\Delta, \theta)$ then describes the fraction of neurons in the ON state near point (Δ, θ). It is also equal to the conditional probability of finding a unit in the active state. The evolution of this function can be described by the relaxation equation:

$$\frac{\partial h(\Delta, \theta)}{\partial t} = \frac{h_0(\Delta, \theta) - h(\Delta, \theta)}{\tau(\Delta, \theta)}. \tag{3.18}$$

Here $h_0(\Delta, \theta)$ is the activation function in the equilibrium, while $\tau(\Delta, \theta)$ is the relaxation time constant describing how fast this equilibrium is reached. The important feature of this equation is that both the equilibrium relaxation function and the time constant depend on the position in the parameter space. Indeed, in an area of unconditional stability (I and III in Figure 3C), the equilibrium function equals 0 and 1, respectively. No other values are permitted because there is no bistability in these regions. In the area of bistability (II in Figure 3C), the equilibrium function takes intermediate values that are determined by the fraction of units in the ON state in the equilibrium. Also, in areas I and III, the equilibrium values are reached with a very fast time constant. We assume that the bistable units can be flipped essentially instantaneously. In the area of bistability (area II), the equilibrium is reached over much longer timescales determined by the time of spontaneous transitions between two stable states of the neuron. Thus, if one assumes that the transition from the ON state to OFF state (decay) is characterized by the time constant $\tau_{1\to0}$ while the opposite transition (spontaneous activation) occurs with time constant $\tau_{0\to1}$, the equilibrium activation function and the time constant of relaxation to the equilibrium value in area II are

$$h_0 = \tau_{1\to0}/(\tau_{1\to0} + \tau_{0\to1}) \tag{3.19}$$

$$\tau_h = \tau_{0\to1}\tau_{1\to0}/(\tau_{1\to0} + \tau_{0\to1}). \tag{3.20}$$

To simplify subsequent equations, we will make an assumption that $\tau_{0\to1} = \tau_{1\to0}$, so that in area II, $h_0 = 1/2$ and $\tau_h = \tau_{1\to0}/2$. The important feature of the biological system is that the decay timescale τ_h is substantially larger than synaptic time constant τ_s. The ratio between τ_h and τ_s is usually exponential (Bialek, 2000; Koulakov, 2001).

To demonstrate that the seemingly simple equation 3.18 can lead to interesting results, we solved the equation for the case when $\tau(\Delta, \theta) = \tau_h$ is constant everywhere in area II. (A more complex case is considered in supplementary materials 2.) Our goal is to understand how the decay of bistable neuronal units translates into the decay of integrator memory or, in other words, how the integrator becomes leaky. That the integrator is leaky implies that the recurrent current changes with time. For example, if the recurrent connections are weak ($\alpha < 1$), the current decays that is, $dI/dt = \dot{I}$ is negative. The equilibrium activation function in this case is sliding down in the parameter space at a rate equal to $\dot{I}$, as illustrated in Figure 17A. Note that the equilibrium activation function is the limit toward which the real activation function is moving at each moment in time, according to equation 3.18. The equilibrium activation function therefore can be nonstationary, in which case the final target for the activation function is dynamically changing. Thus, for $\dot{I} < 0$, the real activation function lags behind the equilibrium values by the time constant τ_h (see Figure 17). To

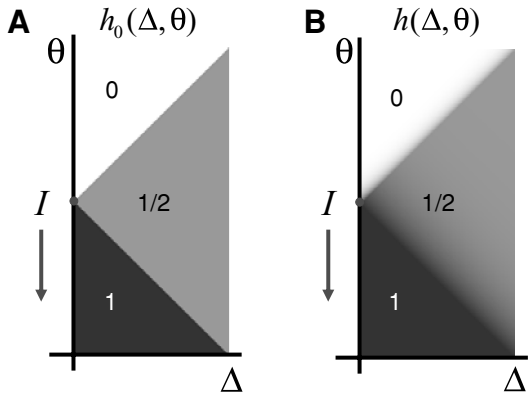

Figure 17: The activation function in the case of nonstationary recurrent current that decreases as a function of time. (A) The equilibrium activation function for the case $\tau_{0\to1} = \tau_{1\to0}$. (B) The actual activation function is lagging behind the equilibrium values.

solve equation 3.18, we notice that once a neuron enters area II, its activation function will relax to the equilibrium value independent of other neurons with the time constant τ_h. As a result, the solution for the activation function in area II becomes

$$h = 1/2 + \mathrm{e}^{-\Delta t/\tau_h}/2. \tag{3.21}$$

Here Δt is the time spent by the neurons in area II. This time is given by the distance traveled in area II divided by the speed of the movement of the area II border $|\dot{I}|$:

$$\Delta t = (\theta - I + \Delta)/|\dot{I}|. \tag{3.22}$$

This equation assumes that the speed is uniform and does not depend on time. This approximation is valid if the acceleration of the eye movement is negligible. A more precise condition for the validity of this approximation is derived in appendix C.

The number of active neurons can be easily calculated from equation 3.21 using equation 3.5. We obtain

$$\begin{aligned} n &= C\overline{\Delta}I + \frac{C}{2}\int_0^\infty d\Delta \int_{I-\Delta}^{I+\Delta} d\theta e^{-(\Delta+\theta-I)/\tau_h|\dot{I}|-\Delta/\overline{\Delta}} = \\ &= C\overline{\Delta}I - C\overline{\Delta}\tau_h\dot{I}\frac{\overline{\Delta}}{2\overline{\Delta}+\tau_h|\dot{I}|}. \end{aligned} \tag{3.23}$$

In formulating this equation we again assumed, similar to the remainder of the study, that $I >> \overline{\Delta}$, which allowed extending integration over Δ to infinity. Equation 3.23 is valid for an arbitrary sign of the time derivative of the current $\dot{I}$: both positive and negative. Note that for $\dot{I} < 0$, the correction to the number of active neurons (second term in equation 3.23) is positive. This is because the activation function extends into area II due to a finite delay time introduced by hysteresis (the gray plume in area II in Figure 17B).

To simplify the analysis further, we assume that the rate of change of the current is actually small. This is possible if the integrator is close to being perfectly tuned. The more exact condition for this assumption to be valid becomes clear in the following equations. Assume that $\tau_h|\dot{I}|$ is much smaller than $\overline{\Delta}$ in the denominator in equation 3.23. Under this condition, we can write that

$$n(t) = C\overline{\Delta} \cdot I(t) - C\overline{\Delta} \cdot \tau_h \dot{I}/2. \tag{3.24}$$

Since the recurrent current is proportional to the number of active neurons in this model,

$$I\,(t + \tau_s) = I_0 n(t), \tag{3.25}$$

we then arrive at the equation describing the dynamics of input current to all units:

$$(\alpha - 1)I = \frac{\alpha \tau_h}{2} \frac{dI}{dt}. \tag{3.26}$$

Here the tuning parameter $\alpha = I_0 C\overline{\Delta}$ is the same as defined above. It is clear from this equation that the mistuning of the integrator should be small, $(1 - \alpha)I << \overline{\Delta}$, for us to neglect the second term in the denominator of equation 3.23. Thus, the leaky integrator equation is valid if the mistuning is not too large.

Equation 3.26 describes the decay of the integrator current in the absence of the external inputs. Assuming $\alpha \approx 1$, the integrator time constant that follows from this equation is

$$\tau_{\text{leak}} = \frac{\tau_h}{2|1 - \alpha|}. \tag{3.27}$$

We conclude that the integrator leak time constant is determined by the rate of decay of bistability in this model. This is in contrast to the models without substantial hysteresis in which the timescale for the integrator leak is provided by the synaptic time constant τ_s (Robinson, 1989; Seung et al., 2000a). This point suggests another interpretation for the robustness of this model. Indeed, since the observed τ_{leak} is about 30 sec, with the timescale of

the decay of hysteresis τ_h of a few seconds (Bialek, 2000; Koulakov, 2001), the parameter α has to be tuned to unity with the precision of about 10%. Thus, the presence of hysteretic neurons allows putting a much weaker constraint on the integrator tuning to reach the same value of leak.

With the help of equation 3.23, one can derive a more general equation for the dynamics of the integrator:

$$(1-\alpha)\, I + \left(\frac{\alpha \overline{\Delta} \tau_h}{2\overline{\Delta} + \tau_h |dI/dt|} + \tau_s \right) \frac{dI}{dt} = I_{ext}. \tag{3.28}$$

This equation is valid for a constant external current. Equation 3.28 can be solved for dI/dt. The result for $\alpha = 1$ is

$$\tau_s \frac{dI}{dt} = \frac{I_{ext} - s\overline{\Delta}}{2} + \sqrt{\left(\frac{I_{ext} - s\overline{\Delta}}{2} \right)^2 + \frac{2\overline{\Delta}\, |I_{ext}|\, \tau_s}{\tau_h}}, \tag{3.29}$$

where $s = \text{sign}\,(I_{ext})$. This equation has the following limits for small and large values of the external current:

$$\tau_s \frac{dI}{dt} \approx 2 I_{ext} \frac{\tau_s}{\tau_h}, \quad |I_{ext}| << \overline{\Delta} \tag{3.30}$$

and

$$\tau_s \frac{dI}{dt} = I_{ext} - \overline{\Delta} \cdot \text{sign}\left(\frac{dI}{dt} \right), \quad |I_{ext}| >> \overline{\Delta}. \tag{3.31}$$

In the limit when transitions between two states of bistable neurons can be neglected, $\tau_h \to \infty$, equation 3.29 yields solution 3.10, previously obtained.

Another point evident from the solutions of kinetic equation is that the rate of integration is determined by the synaptic timescale τ_s (see equation 3.31), while the decay of delayed activity is proportional to the average time of spontaneous decay of the bistability τ_h (see equation 3.27). The latter is an exponential function of the former (Bialek, 2000; Koulakov, 2001) and may significantly exceed the synaptic timescale. Thus, the neural integrator described here is capable of integrating large external stimuli effectively due to a small time constant τ_s while remaining robust due to the larger memory timescale τ_h.

Finally, we note that according to equation 3.30, in the perfectly tuned case ($\alpha = 1$), the current in the integrator is changing ($dI/dt \neq 0$) even for infinitely small external current. This implies that the integrator with spontaneous transitions does not have the threshold for integration. The rate of integration of weak inputs is determined, however, by the hysteresis time constant τ_h, which is a remnant of the threshold existing for perfectly

bistable units without spontaneous transitions between the ON and OFF states (see Figure 5).

3.6 Experimental Predictions. The main experimental prediction that follows from this model concerns the comparison between the firing rates during fixations and slow eye movements. The latter include both smooth pursuit and vestibulo-ocular reflex (VOR) responses. We have shown that the model yields an inverted sign of hysteresis for some neurons in the ensemble, similar to experiential observations. Thus, ON responses are above OFF responses in Figures 9 and 14 for G1 and G2. We argue here that the responses of the same neurons are different during smooth eye movements, sometimes in a qualitative manner. For the purposes of this section, we will use the multistable model with no decays (see section 3.2 and Figure 14).

We first notice that the activation function is the same for the smooth eye movements as for the case when the neurons are not connected by global recurrent connections (compare Figures 4A and 5A). Thus, the firing rate as a function of eye position can be derived from the dependence with no recurrent connections by choosing the appropriate value of input current. The total value of input current is composed of the recurrent and external current. We have associated the former with the eye position (see equation 3.17). The external current has a minimum value that allows sustained integration $|I_{ext}| = \overline{\Delta}$. We thus obtain for the smooth eye movements

$$I = E + \overline{\Delta} \cdot \mathrm{sign}(I_{ext}). \tag{3.32}$$

Here we assume that external input is equal to the threshold value for sustained integration: $|I_{ext}| = \overline{\Delta}$. The firing rate as a function of eye position is obtained from the dependencies with no feedback by shifting the ON dependence to the left by $\overline{\Delta}$ and the OFF dependence to the right by $\overline{\Delta}$. The dependence is shown in Figure 18B for the multistable neurons.

As clear from Figure 18, the hysteresis during smooth eye movements is related to that during fixations. By plotting the width of hysteresis for individual cells, one could map out the dependence experimentally. In particular, the neurons with small hysteresis during fixations will acquire history dependence in the response during smooth eye movements. This feature could be detected experimentally.

4 Discussion

We studied the recurrent networks built out of hysteretic units. We considered the case when neurons belonging to the same integrator network have different values of hysteresis. The properties of the integrators with differential hysteresis were first analyzed by Goldman et al. (2003).

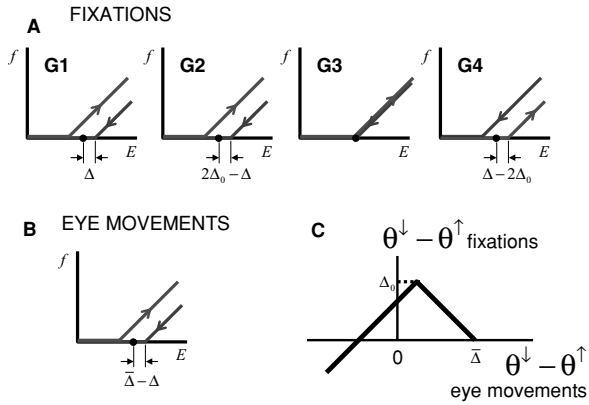

Figure 18: The differences in responses between the case of eye fixations (A) and smooth movements (B) predicted by this model. Note that response depicted in B is predicted for all groups of cells G1 through G4. (C) The distance between OFF and ON thresholds for fixations versus small velocity eye movements that follows from A and B.

We analyzed the properties of such networks assuming a large number of neurons, which allowed treating the behavior of the system statistically, using distribution functions. To simplify our analysis, we assumed that neurons are connected in the all-to-all fashion; each neuron makes synapses with the dendrites of all other neurons in the network. This assumption allowed us to study the properties of this network analytically, without the use of a computer. We expect, however, that many properties derived here will be valid in networks with sparse random connectivity. Because at the basis of our model lies the neuronal activation function that resembles sand dunes in shape (see Figure 6C), we call our model the "sandpile" integrator.

One of the important properties derived here is that the hysteresis of the integrator neurons is strongly affected by the recurrent connections. Assume that before neurons are connected, their firing rate exhibited history dependence of a particular sign. For example, assume that the firing rate was always higher after the decrease in the external input into the neuron than after an increase. Such history dependence is often displayed by neurons with intracellular mechanisms of positive feedback or in local strongly connected clusters of neurons (Koulakov et al., 2002). We called, somewhat arbitrarily, this type of history dependence a regular native hysteresis or a hysteresis of positive sign. As follows from our consideration, the history dependence after these neurons are connected in the network is different. Many neurons actually change the sign or direction of hysteresis from positive (regular) to negative (anomalous). This implies that the firing rates after an increase in the external current (ON saccades) are higher than after

a decrease in the current. This feature is acquired by the neurons when they are connected into the network, which is needed to maintain the memory about the eye position. Similarly, it is quite easy to show that the neurons with negative native hysteresis may reverse their sign to positive when connected into networks.

This phenomenon of the reversal of the direction of history dependence in the recurrent networks of hysteretic neurons could reconcile the observed direction with the possible mechanisms of generating hysteresis in the VPNI. Indeed, the observed sign of hysteresis is almost always negative (Pastor, Torres, Delgado-Garcia, & Baker, 1991; Aksay et al., 2003). Simplistically, this may imply that the mechanisms other than those involving positive feedback are responsible for this sign of hysteresis, such as the mechanisms based on negative feedback. Our findings indicate that the neurons could have a positive feedback active inside the cells generating the regular direction of history dependence, which, after the neurons are connected into the network, becomes reversed. In a sense, two positive feedbacks, intracellular and extracellular, may manifest themselves in reversed hysteresis mimicking the negative feedback systems.

A similar phenomenon of the reversal of the hysteresis sign was obtained in the model of Goldman et al. (2003). This study examined more complex network connectivity in which neurons targeted specific dendritic compartments. The number of bistable dendritic compartments was equal to the number of neurons in the network. In the simplest case, when the dendritic compartments targeted by a single neuron have the same properties (network connectivity has an outer product form), their model was solved exactly without the use of computer. Their findings in this case indicate that the cells do not display hysteresis in the firing rate to eye position dependence. When the outer product form assumption was relaxed, the neurons displayed the hysteresis in the firing rate similar to the one observed experimentally, including the sign. In the latter case, the model of Goldman et al. (2003) could not be solved exactly. Goldman et al. demonstrated their findings computationally and using a qualitative argument. Here we presented a different view on a simpler system than Goldman et al., discussed. Due to the simplicity of our networks (all-to-all connectivity) it was possible to demonstrate the hysteresis in the firing rate and provide an exact solution to our system without the use of computer. Perhaps our method with the extended dimensionality of the parameter space could yield the exact solution to the model of Goldman et al. in the non-outer-product case.

Previous studies of the goldfish VPNI have failed to find substantial bistability at the level of the individual neuron when these neurons were stimulated intracellularly (Aksay et al., 2001). These findings could be reconciled with our model if one assumes that the stimulations were applied to nonbistable neuronal compartments. This could occur if only a fraction of dendritic tree were actually bistable or if the bistability on the single-neuron level were conditional, that is, existed only for a small range of eye positions

near the activation threshold. This possibility is discussed in our previous study (Koulakov et al., 2002). In particular it is possible that the bistability is created in part due to NMDA current in the recurrent synapses, which is dependent substantially on the eye position (Koulakov et al., 2002). The model with conditional bistability is beyond the scope of the study presented here, in which we considered the simplest possible models. The models considered here are therefore not suitable for addressing the issue of single-neuron bistability in the intracellular stimulation experiments (Aksay et al., 2001). At the same time, experimental studies have clearly shown hysteresis in the response of integrator neurons when they are stimulated endogenously as an ensemble (Aksay et al., 2003). These experiments are addressed in our study. These points indicate that further investigation, both theoretical and experimental, of the response of individual neurons to external intracellular stimulation is warranted.

We studied the effect of spontaneous transitions in the bistable units on the dynamics of the integrator. We found that in the presence of such transitions, the integrator becomes leaky. The leak time constant is determined by the rate of spontaneous transitions, at least when the mistuning of the integrator is small. For the case of constant rate of spontaneous transitions, we have derived the simple formula relating two time constants: those of integrator leak and of spontaneous transitions (see equations 2.1 and 3.27). The relationship is similar to the formula earlier derived for the connection between the fast synaptic time constant and integrator leak for integrators with no hysteresis. A more complex case of nonconstant rate of spontaneous transitions is considered in supplementary material 2. Our findings also indicate that the integration rate for the hysteretic integrator is determined by the faster synaptic time constant. This statement is valid for large inputs. Thus, hysteretic integrators are both robust and capable of responding quickly to the external stimuli.

The effects of noise on integrator networks were also addressed by Goldman et al. (2003). They show that adding noise allowed small inputs to be integrated and caused slow decay. These conclusions are similar to our result. In our study, the transitions caused by noise are treated quantitatively and are described by an analytical equation for the rates of decay (see equations 3.28 and 1.1). Similarly to Goldman et al., our equations show that the decay is fast for larger perturbations, while integration of very small smooth inputs will be governed by the slower time constant of spontaneous transitions in the bistable units.

Our consideration of the dynamics of transitions in the hysteretic neurons was limited for two reasons. First, we assumed that the transition time constant is the same for all units, independent of the width of hysteresis. Second, we assumed that the spontaneous transition rate is the same at all positions in the region of bistability. Our approach allows relaxing these assumptions at the expense of simplicity (as done in supplementary material 2). We also expect that making small hysteresis units less robust to

noise-driven transitions, as predicted by prior theoretical studies (Bialek, 2000; Koulakov et al., 2002), will allow integrating small amplitude inputs.

Our consideration of network dynamics was limited to the evolution of the distribution function in the two-dimensional parameter space (hysteresis width and position). Additional parameters could be included by making the coordinate space for the distribution function three-, or more than three-, dimensional. Such variables would allow a study of network topologies more complex than all-to-all.

5 Conclusion

We studied an exactly solvable model for recurrent networks of hysteretic neurons. This model displays the reversal of the direction of hysteresis when the recurrent connections are included. The leak time constant of the integrator with hysteretic neurons is determined by the rate of spontaneous noise-driven transitions in the individual neurons. We argue that experimental data are consistent with the positive feedback existing on both intracellular and network levels.

Appendix A: Derivation of Equation 3.6

In this appendix we show that slow dynamics of the integrator can be described by equations with synaptic delay. Similar to the remainder of our study, we assume here that recurrent synaptic current is a linear function of the number of active neurons $n(t)$:

$$I(t) = \int_{-\infty}^{t} K(t - t')n(t')dt' + I_{ext}. \tag{A.1}$$

Here $K(t)$ is the function that describes time-dependent synaptic transmission. This function can be associated with the postsynaptic current in response to a pulse of input, such as action potential. If the variations of the presynaptic variable $n(t)$ are slow, one can approximate them by the first-order Taylor series expansion,

$$n(t') \approx n(t) + (t' - t)\frac{dn(t)}{dt}, \tag{A.2}$$

in equation A.1 to obtain the following approximation:

$$I(t) \approx I_0 n(t) - I_1 \frac{dn(t)}{dt} + I_{ext}. \tag{A.3}$$

This expression describes approximately the variation of the current if the number of active neurons $n(t)$ is changing slowly. In this expression, we introduced the moments of the postsynaptic current defined as follows:

$$I_0 = \int_0^\infty K(t)dt,\ I_1 = \int_0^\infty K(t)|t|dt. \tag{A.4}$$

Equation A.3 can be transformed using the first-order Taylor expansion:

$$\begin{aligned} I(t) &\approx I_0 \left[n(t) - \tau_s \frac{dn(t)}{dt} \right] + I_{ext} \approx \\ &\approx I_0 \left[n(t - \tau_s) \right] + I_{ext}. \end{aligned} \tag{A.5}$$

where we have introduced the following time constant:

$$\tau_s = \frac{I_1}{I_0} = \frac{\int_{-\infty}^0 K(t)|t|dt}{\int_{-\infty}^0 K(t)dt}. \tag{A.6}$$

Thus, the slow dynamics of current is described by the delayed equation A.5 that is similar to equation 3.6. According to equation A.6, synaptic delay τ_s is determined by the average period over which synaptic current arrives in response to a pulsed input. The synaptic delay time is therefore weakly dependent on the onset of synaptic current.

Appendix B: Parameters of the Stable State of the Integrator

Here we apply the stability condition 3.12 to some distributions of hysteresis widths $\rho(\Delta)$ and obtain the parameter of stable configuration Δ_0 (see Figure 6B). The boundary function in Figure 6B is

$$b(\Delta) = \begin{cases} \Delta, & \Delta < \Delta_0 \\ 2\Delta_0 - \Delta, & \Delta \geq \Delta_0 \end{cases}. \tag{B.1}$$

The goal is to find Δ_0 using equation 3.12. Equation 3.12 can be rewritten, using equation B.1, as follows:

$$\int_0^{\Delta_0} \rho(\Delta)\Delta d\Delta + \int_{\Delta_0}^\infty \rho(\Delta)(2\Delta_0 - \Delta)d\Delta = 0. \tag{B.2}$$

We assumed here that the integral over Δ can be extended to infinity, which is valid for large values of I. For the exponential distribution $\rho(\Delta)$ given by equation 3.1 from equation B.2, we obtain

$$\Delta_0 = \overline{\Delta} \ln 2. \tag{B.3}$$

For the delta function distribution of hystereses 3.14, $\Delta_0 = \overline{\Delta}/2$, which can be verified by direct substitution to equation B.2. This leads directly to equation 3.15.

Appendix C: Conditions for Neglecting Acceleration of the Recurrent Current

Here we derive the condition at which the acceleration in the recurrent current can be neglected in the kinetic equation. The effects of acceleration are negligible if the variation of the current's velocity is small within the time constant of the decay of bistability $\Delta\dot{I} \sim \tau_h \ddot{I} \ll \dot{I}$, so that the integral in equation 2.23 is not strongly affected. This results in the condition $\tau_h \ll \tau_{\text{leak}}$, that is, $|\alpha - 1| \ll 1$, in view of equation 3.27.

References

Aksay, E., Baker, R., Seung, H. S., & Tank, D. W. (2000). Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *J. Neurophysiol., 84,* 1035–1049.

Aksay, E., Gamkrelidze, G., Seung, H. S., Baker, R., & Tank, D. W. (2001). In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nat. Neurosci., 4,* 184–193.

Aksay, E., Major, G., Goldman, M. S., Baker, R., Seung, H. S., & Tank, D. W. (2003). History dependence of rate covariation between neurons during persistent activity in an oculomotor integrator. *Cereb. Cortex., 13,* 1173–1184.

Bialek, W. (2000). *Stability and noise in biochemical switches*. Available online at http://lanlarxivorg/abs/cond-mat/0005235.

Camperi, M., & Wang, X. J. (1998). A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability. *J. Comput. Neurosci., 5,* 383–405.

Fukushima, K., & Kaneko, C. R. (1995). Vestibular integrators in the oculomotor system. *Neurosci. Res., 22,* 249–258.

Fuster, J. M. (1995). *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.

Goldman, M. S., Levine, J. H., Major, G., Tank, D. W., & Seung, H. S. (2003). Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cereb. Cortex, 13,* 1185–1195.

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

Koulakov, A. A. (2001). Properties of synaptic transmission and the global stability of delayed activity states. *Network, 12,* 47–74.

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

Lin, Y., Skeberdis, V. A., Francesconi, A., Bennett, M. V., & Zukin, R. S. (2004). Postsynaptic density protein-95 regulates NMDA channel gating and surface expression. *J. Neurosci. 24,* 10138–10148.

Lisman, J. E., Fellous, J. M., & Wang, X. J. (1998). A role for NMDA-receptor channels in working memory. *Nat. Neurosci., 1,* 273–275.

Machens, C. K., Romo, R., & Brody, C. D. (2005). Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science, 307,* 1121–1124.

Major, G., & Tank, D. (2004). Persistent neural activity: Prevalence and mechanisms. *Curr. Opin. Neurobiol., 14,* 675–684.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cereb. Cortex, 13,* 1257–1269.

Miller, P., Brody, C. D., Romo, R., & Wang, X. J. (2003). A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cereb. Cortex, 13,* 1208–1218.

Pastor, A. M., Torres, B., Delgado-Garcia, J. M., & Baker, R. (1991). Discharge characteristics of medial rectus and abducens motoneurons in the goldfish. *J. Neurophysiol., 66,* 2125–2140.

Robinson, D. A. (1974). The effect of cerebellectomy on the cat's vestibulo-ocular integrator. *Brain Res., 71,* 195–207.

Robinson, D. A. (1989). Integrating with neurons. *Annu. Rev. Neurosci., 12,* 33–45.

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

Rosen, M. J. A. (1972). A theoretical neural integrator. *IEEE Trans. Biomed. Eng., 19,* 362–367.

Seung, H. S., Lee, D. D., Reis, B. Y., & Tank, D. W. (2000a). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron, 26,* 259–271.

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

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol., 86,* 1916–1936.

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *J. Neurosci., 19,* 9587–9603.

Received December 13, 2006; accepted December 27, 2007.