J Comput Neurosci
DOI 10.1007/s10827-010-0275-y

A network of spiking neurons that can represent interval timing: mean field analysis

Jeffrey P. Gavornik · Harel Z. Shouval

Received: 18 January 2010 / Revised: 18 August 2010 / Accepted: 26 August 2010

Abstract Despite the vital importance of our ability to accurately process and encode temporal information, the underlying neural mechanisms are largely unknown. We have previously described a theoretical framework that explains how temporal representations, similar to those reported in the visual cortex, can form in locally recurrent cortical networks as a function of reward modulated synaptic plasticity. This framework allows networks of both linear and spiking neurons to learn the temporal interval between a stimulus and paired reward signal presented during training. Here we use a mean field approach to analyze the dynamics of non-linear stochastic spiking neurons in a network trained to encode specific time intervals. This analysis explains how recurrent excitatory feedback allows a network structure to encode temporal representations.

Action Editor: Nicolas Brunel

J. P. Gavornik · H. Z. Shouval (✉)
Department of Neurobiology and Anatomy,
The University of Texas Medical School at Houston,
6431 Fannin St., Houston, TX 77030, USA
e-mail: harel.shouval@uth.tmc.edu

J. P. Gavornik
Department of Electrical and Computer Engineering,
The University of Texas at Austin, 2501 Speedway,
Austin, TX 78712-0240, USA

Present Address:
J. P. Gavornik
The Picower Institute for Learning and Memory,
Massachusetts Institute of Technology,
77 Massachusetts Avenue, 46-3301,
Cambridge, MA 02139-4307, USA

Keywords Recurrent network · Mean field theory · Synaptic plasticity · Spontaneous activity · Reinforcement learning

1 Introduction

Disparate visual stimuli can be used as markers for internal time estimates, for example when determining how long a traffic light will remain yellow. The idea that neurons in the primary visual cortex might contribute explicitly to this ability contradicts our understanding of V1 as an immutable visual feature detector and the prevailing notion that temporal processing is a higher-order cognitive function (Mauk and Buonomano 2004). These expectations of V1 function are challenged by the finding that neurons in rat V1 can learn robust representations of the temporal offset between a visual stimulus and water reward presented during a behavioral task (Shuler and Bear 2006). Experimental results suggesting that temporal processing might begin in other primary sensory regions have been reported as well (Super et al. 2001; Moshitch et al. 2006). These observations led us to investigate how local networks or single neurons can learn, as a function of reward, temporal representations in low-level sensory areas.

In a previous work (Gavornik et al. 2009), outlined below, we demonstrated that recurrent networks can use reward modulated Hebbian type plasticity as a mechanism to encode time. Here, we presents a mean field theory (MFT) analysis of temporal representations generated by a network of conductance based integrate and fire neurons (described in Section 3). This analysis specifically addresses the mechanistic question of how lateral excitation between non-linear spiking

Published online: 10 September 2010

neurons can be used as the substrate to encode specific durations of time. We first perform MFT analysis on a noise free system (Section 4) then describe and compare the results of this analysis to those simulated in the full network (Section 5) and show that the temporal report is invariant to the magnitude of the stimulus and that these representations can be used to accurately encode short time intervals. Next, we show how the strength of recurrent connections effects spontaneous activity levels (Section 6). Finally, we describe how the network operates in the super-threshold bistable regions (Section 7).

2 A model of learned network timing

Theoretical studies have shown that a careful tuning of lateral weights can generate neural networks with attractor states that can possibly account for the neuronal dynamics observed in association with working memory (Amit 1989; Seung 1996; Eliasmith 2005). Until recently the potential for local cortical networks to encode temporal instantiations by learning specific slow dynamics in response to sensory stimuli had not been considered. Based on the reports of timing activity in rat V1, we proposed a theoretical framework showing how a network with local lateral excitatory connectivity can learn to represent temporal intervals as a function of paired stimulus and reward signals (Gavornik et al. 2009).

Our model consists of a fully recurrent population of neurons receiving stimulating feed-forward projections. This structure (Fig. 1(a)) is roughly analogous to the visual cortex where V1 neurons receive projections from the LGN and interact locally with other V1 neurons. In order to explain how a network can learn temporal representations we described a paradigm, called Reward Dependent Expression (RDE), wherein Hebbian plasticity is modulated by a reward signal paired with feed-forward stimulation during training. Briefly, RDE posits (1) that the action of a reward signal results in long term potentiation through the permanent expression of activity driven molecular processes, described as "proto-weights" and (2) that ongoing activity in the network inhibits the expressive action of the reward signal. These assumptions allow RDE to solve a temporal credit assignment problem associated with the offset between stimulus and reward during early training sessions. Additionally, the learning rule naturally allows the network to fine-tune synaptic weights by preventing additional potentiation as the network nears its target activity level. After training with RDE, the network responds to specific feed-forward stimulation patterns for a period of time equal to the temporal offset between the stimulus and a paired reward signal presented during training. Trial-to-trial fluctuations in evoked response duration combined with the highly non-linear relation between synaptic weights and the network report time (see Fig. 5(a)) impose practical limits on the ability of RDE to encode long report periods.

RDE was formulated specifically to explain how plasticity between excitatory neurons in the visual cortex could encode temporal intervals cued by visual stimulation, but its principles may apply in other brain regions as well. The temporal representations created by RDE consist of periods of post-stimulus activity whose durations are interpreted to encode neural

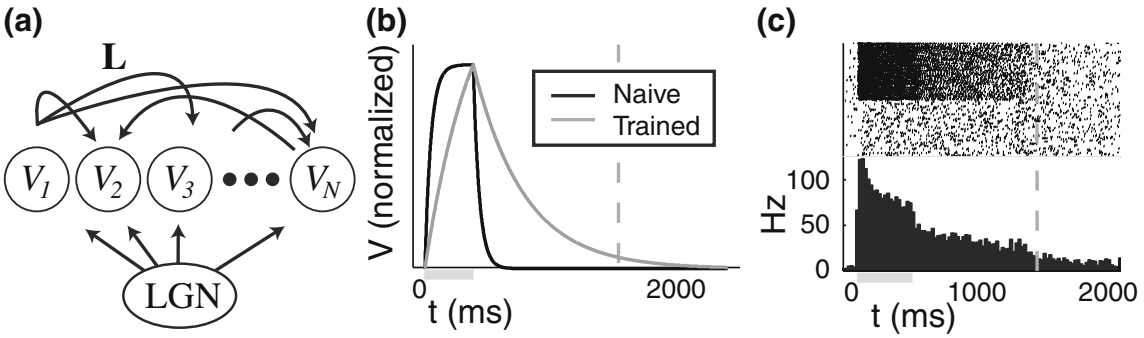

Fig. 1 Temporal representations created by RDE. (**a**) Neurons in the recurrent layer of our network model are stimulated by retinal activation via the LGN. **L** is the matrix defining lateral excitation. (**b**) With a linear neuron model, time is encoded by the exponential decay rate of an activity variable V. (**c**) In the spiking neuron model, evoked activity (shown by spike rasters, where each row represents a single neuron in the network, and the resultant histogram) in a responsive sub-population of the network persists until the time of reward. In both models, the stimulus is active during the period marked by the *gray bar* and the reward time is indicated by the *dashed line*. See Gavornik et al. (2009) for details of learning with RDE

instances of "time". The form of these representations are qualitatively consistent with the "sustained response" reported in rat V1 (Shuler and Bear 2006). Our original work demonstrated that a learned network structure can result in this form of representation using both a rate-based linear neuron model and non-linear integrate and fire neurons. Neuronal activity dynamics in the linear case are purely exponential and easy to analyze (Fig. 1(b)). Activity dynamics following stimulation, in a network of spiking neurons are characterized by a rapid drop to a plateau level, with activity slowly decreasing during the period of temporal report, and a second drop back to the baseline level at the time of reward (Fig. 1(c)). This behavior is quantitatively similar to the experimental data and differs from roughly linear ramping activity seen in other brain areas. A key component of the previous work was to demonstrate that RDE allows the network to precisely tune recurrent synapses to encode specific times; this is import since recurrent network models can be exquisitely sensitive to synaptic tuning. Notably we have shown that RDE can be used to learn these times even in a network of stochastic spiking neurons. Although the mechanism responsible for encoding time in our model is recurrent excitation, we also demonstrated that RDE works in the presence of both feed-forward and recurrent inhibition. After training, the average dynamics in networks including inhibition were similar to the dynamics in purely excitatory networks.

The aim of this paper is to determine quantitively how the spiking network represents time, and how neuronal non-linearity shapes the observed form of its dynamics. Understanding how the excitability of individual spiking neurons can lead to this network-level dynamical activity profile, which the linear analysis can not explain, is critical to understanding mechanistically how temporal representations might form in biological neural networks.

3 Spiking network model

The model network consists of a single layer of $N = 100$ neurons with full excitatory lateral connectivity (Fig. 1(a)). The recurrent layer is assumed to be roughly analogous to V1, which has a large number of synapses with local origin and where extrastriate feedback accounts for a small percentage of total excitatory current (Johnson and Burkhalter 1996; Budd 1998). Recurrent layer neurons are driven by monocular inputs and receive feed-forward projections that are active only during periods of stimulation.

Spiking neurons were simulated with a conductance based integrate and fire model. The equation governing the sub-spiking threshold dynamics of the membrane potential, V, of a single neuron i is:

$$C\frac{dV_i}{dt} = g_L(E_L - V_i) + g_{E,i}(E_E - V_i) \tag{1}$$

where C is the membrane capacitance, and E_L and g_L are the reversal potential and conductance associated with the leak current. This equation applies when $V_i < V_\theta$, where V_θ is the spike threshold. The variable $g_{E,i}$ represents the total excitatory conductance with current driven by the reversal potential E_E. Inhibitory synaptic connections do not contribute to the formation of temporal representations and are omitted here for the sake of clarity. Each synapse contributes the product of its activation level and weight to the total conductance:

$$g_{E,i} = \sum_{j=1}^{J} W_{ij} s_j(t) \tag{2}$$

where J is the number of excitatory synapses driving the neuron and $s_j(t)$ is the activity level of synapse j at time t. The synaptic weight variable W_{ij} is used here to indicate that conductance is determined by all synaptic connections, both feed-forward and recurrent. The subset of W consisting of only the lateral excitatory connections is an $N \times N$ matrix $\mathbf{L}$ and, for the sake of this analysis, we will assume homogeneous connectivity.

Synaptic resources are assumed to be finite and saturate following multiple pre-synaptic spiking events; maximal trans-membrane current occurs when 100% of synaptic resources are active. Synaptic activation dynamics are modeled independently for each synapse according to:

$$\frac{ds_i}{dt} = -\frac{s_i}{\tau_s} + \rho(1 - s_i)\sum_{\text{pre}} \delta\left(t - t_{\text{pre}}\right) \tag{3}$$

The synaptic activation level jumps by a fixed percentage, ρ, with each pre-synaptic spike and decays with time constant τ_s. A biological interpretation is that the neurotransmitter released by each spike binds a fixed percentage of available post-synaptic receptors, and that bound neurotransmitter dissociates at a constant rate.

Parameters were chosen to be biologically plausible, based on values used in a previous computational work (Machens et al. 2005). The resting membrane voltage was set to −60 mV, and reversal potentials for

excitatory and leak ionic species were −5, and −60 mV respectively. Spiking occurred when membrane voltage reached a threshold value $V_\theta = -55$ mV. After spikes, the membrane voltage was reset to $V_{\text{reset}} = -61$ mV and held for a 2 ms absolute refractory period. The leak conductance was 10e-3 μS and membrane capacitance was set to give a membrane time constant of 20 ms. Each spike is assumed to utilize approximately 15% of the available synaptic resources ($\rho = 1/7$) and, as in other models (Lisman et al. 1998; Compte et al. 2000), synaptic activation decays with a slow time constant appropriate for NMDA receptor activation dynamics ($\tau_s = 80$ ms).

After training, a feed-forward pulse that drives the network to a high activity state is sufficient to evoke a report of encoded time (Gavornik et al. 2009). During the stimulation period, recurrent layer neurons receive random spikes with arrival times drawn from a time-varying Poission distribution chosen to mimic LGN activity (Mastronarde 1987). In the original work, each neuron in the recurrent layer also received random spiking input from independent Poisson processes with intensity levels set to produce a low level of spontaneous activity in the network (see Section 6).

4 Mean field theory analysis

4.1 Extracting the I/O function for a conductance based neuron

The spiking network model is a high dimensional system comprised of order-N coupled differential equations describing the membrane voltage and synaptic activation dynamics of all of the neurons and synapses in the network. The MFT approach ignores the detailed interactions between individual neurons within this large population and instead considers a single external "field" that approximates the average ensemble behavior. Stochasticity in the conductance based integrate and fire model described above results from random synaptic inputs. Accordingly, the approach here will be to replace random synaptic activations and resultant currents by their mean values and to analyze dynamics in terms of the input-output relationship of a single neuron. This is similar to the approaches that have been used previously to analyze and solve many-body system problems in various neural networks (Renart et al. 2003; Amit and Brunel 1997; Amit et al. 1985). Since the temporal representation forms in the recurrent layer of our network we will start by analyzing the case where all of the excitatory input originates from recurrent feed-back.

The first requirement for the mean-field analysis is an accurate description of the firing rate, ν, of the integrate and fire neuron model as a function of synaptic input over the operating range of a "temporal report". This relationship can be investigated numerically by driving the spiking neuron model at a constant rate and simply counting the spikes resulting in a fixed amount of time. If any of the neuron parameters change, including the strength of synaptic weights responsible for afferent current, the curve resulting from the numerical approach must be regenerated, limiting its usefulness as a tool to understand network dynamics. An alternative approach is to quantify the I/O curve analytically.

In the spiking neuron model voltage changes at a rate proportional to the total ionic current (Eq. (1)). The resting membrane potential is set by the reversal potential of the leak conductance, which is assumed to be constant in time. Excitatory conductance, however, is a function of random synaptic input. For a single excitatory synapse, the equation governing synaptic conductances (Eq. (3)) can be re-written as a stochastic differential equation:

$$\frac{dS_E(t)}{dt} = X(t)\rho(1 - S_E(t)) - \frac{S_E(t)}{\tau_s} \tag{4}$$

where $X(t)$ is a random binary spiking process, $S_E(t)$ is a random process describing excitatory synaptic activation, and the other parameters are as defined previously. Assuming that $X(t)$ is a temporally uncorrelated stationary Poisson process, the expectation of $S_E(t)$ evolves in time as a function of the expectation of $X(t)$ according to the first-order differential equation:

$$\frac{d}{dt}\mathsf{E}[S_E(t)] = \mathsf{E}[X(t)]\rho(1 - \mathsf{E}[S_E(t)]) - \frac{\mathsf{E}[S_E(t)]}{\tau_s} \tag{5}$$

For the Poisson process, $\mathsf{E}[X(t)] \equiv \mu$, which is the pre-synaptic firing frequency driving the synapse. Defining $s_E(t) \equiv \mathsf{E}[S_E]$ and assuming the initial condition $s_E(0) = 0$, then:

$$s_E(t) = \frac{\mu\rho\left(1 - e^{-t\left(\mu\rho + \frac{1}{\tau_s}\right)}\right)}{\mu\rho + \tau_s^{-1}} \tag{6}$$

for $t \geq 0$. The resulting steady state value of $s_E(t)$, for a constant value of μ, as $t \to \infty$, is:

$$s_E^{\infty}(\mu) = \frac{\rho\mu}{\rho\mu + \tau_s^{-1}} \tag{7}$$

Excitatory conductance through a single synapse is the product of the maximal conductance, defined as the synaptic weight W, and the synaptic resources activation level. That is:

$$g_E(t) = Ws_E(t) \tag{8}$$

The expression for average synaptic activation can now be used to write an equation for the mean conductance of a single synapse independent of time. Assuming that network activity has been approximately constant long enough to keep the synapse near its steady state value, which is the case following feedforward stimulation in the network model described in Section 2, Eq. (8) can be simplified further by replacing $s_E(t)$ in the limit with $s_E^\infty(\mu)$, which results in a constant steady-state excitatory conductance value:

$$g_E^\infty(\mu, W) = Ws_E^\infty(\mu) \tag{9}$$

The firing rate of the conductance based neuron model can be estimated analytically as a function of the mean input current. Equation (1) describes the subthreshold dynamics of the integrate and fire neuron model, where net current across the membrane is a function of driving force and conductances associated with the various ionic species. This can be re-written as:

$$C\frac{dV}{dt} = -Vg_{\text{tot}} + I_{\text{rev}} \tag{10}$$

where the total conductance is $g_{\text{tot}} = g_E(\mu) + g_L$ and the reversal potential currents are $I_{\text{rev}} = g_E(\mu)E_E + g_L E_L$. The output spike frequency is the inverse of the time, t_{spike}, required for the voltage to increase from the reset level, V_{reset}, to the spike threshold, V_{thresh}, and can be calculated directly from Eq. (10) by separating the variables and integrating. The result, based only on the mean current without fluctuations for a single input spike frequency, is:

$$t_{\text{spike}} = \frac{C}{g_{\text{tot}}}\log\left(\frac{I_{\text{rev}} - g_{\text{tot}}V_{\text{reset}}}{I_{\text{rev}} - g_{\text{tot}}V_\theta}\right) \tag{11}$$

If t_{spike} is real, the corresponding spike frequency is equal to ${t_{\text{spike}}}^{-1}$; otherwise the spike frequency is 0. It is now possible to write an analytical function, $\phi(\mu, W)$, relating the output spike frequency to the mean input spike rate (though the steady state conductance level) and synaptic weight by combining equations above.

$$\phi(\mu, W) = \left[t_{\text{ref}} + \frac{C}{g_E^\infty(\mu, W) + g_L} \times \log\left(\frac{g_E^\infty(\mu,W)E_E + g_L E_L - (g_E^\infty(\mu,W)+g_L)V_{\text{reset}}}{g_E^\infty(\mu,W)E_E + g_L E_L - (g_E^\infty(\mu,W)+g_L)V_\theta}\right)\right]^{-1} \tag{12}$$

An upper limit to the output spike frequency is set by the absolute refractory period, t_{ref}. That is, $\phi_{\max} = 1/t_{\text{ref}}$.

Figure 2 demonstrates that the spike rate predicted by the MFT analytical ϕ curve agrees well with numerical estimates of ν over a range of W above some input frequency threshold. This spiking threshold, which changes as a function of synaptic weight, occurs above the input where fluctuation driven output is seen in numerical I-O solution (see also Section 6 and the discussion). Its value is determined by the minimum input current required to drive the membrane voltage all the way to threshold, which occurs when the numerator in the log operand of Eq. (12) is equal to 0. In terms of the excitatory conductance, the threshold value is:

$$g_E^\theta = \frac{V_\theta(g_I + g_L) + g_L E_L}{E_E - V_\theta} \tag{13}$$

and the corresponding input frequency threshold is:

$$\mu^\theta = \frac{g_E^\theta}{\rho\tau_s(W - g_E^\theta)} \tag{14}$$

For a given synaptic weight, spiking will occur whenever $\mu \geq \mu^\theta$. The region above μ^θ is a mean driven firing region, in which firing rates are well approximated by the deterministic theory (Fig. 2(a)) whereas for input frequency lower than μ^θ any firing that does occur is driven by fluctuations from the mean, and cannot be accounted for by the deterministic approximation.

4.2 Pseudo-steady state approximation

Each synapse in the recurrent layer takes a little over 100 ms to reach its steady state activation level when driven by 50 Hz input. Since the stimulation protocol specifies a stimulation period of 400 ms, this means that recurrent synapses in the network have reached approximate steady state activity levels by the beginning of the decay phase. Furthermore, from Eq. (6),

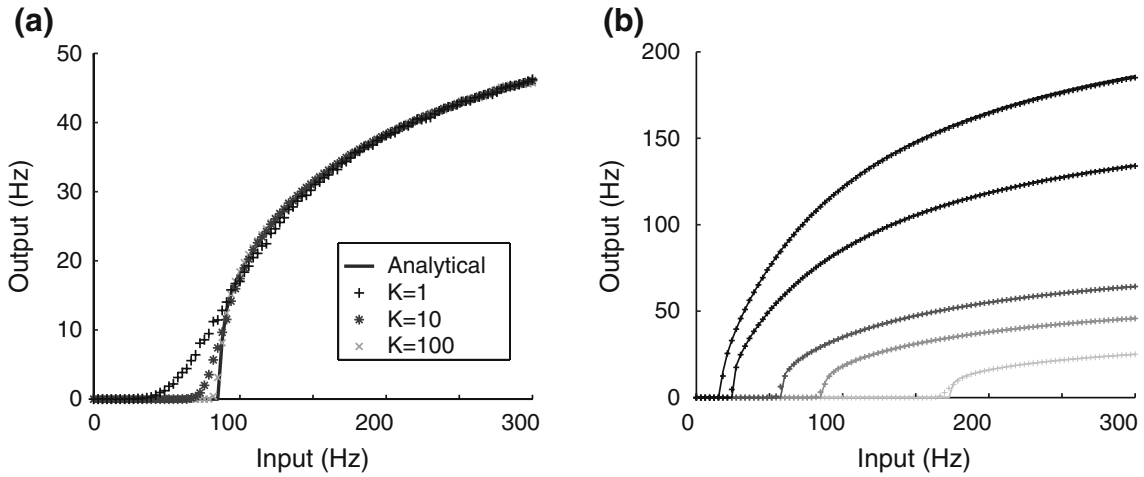

Fig. 2 Input-output relationship of an IF neuron. (**a**) The analytical ϕ curve (*black line*) calculated using MFT analysis (Eq. (12)) with $W = 3.4\text{e-}3$ μS compared to numerically generated estimates of the output rate ν (symbols). K indicates the number of independent synapses driving activity in the numerically simulated neuron; individual synaptic weights are scaled by K so that the cumulative synaptic weight is constant for each of the three cases shown (K = 1,10,100). As K increases, the numerical approximations approach the analytical curve. Note that in the model described in Section 3, $K = N = 100$ for the recurrent synapses. Deviations exist primarily in the low frequency input region where output is driven by fluctuations (see Eqs. (13) and (14)). (**b**) The analytical solution (*solid lines*) compares well with numerical results (plus signs, K = 100) for values of W ranging from 1.5e-3 μS (*light gray*) to 6.0e-3 (*black*). All parameters are as listed in Section 3

synaptic activation tracks its steady state value with a time constant equal to $\frac{1}{\mu\rho+\tau_S^{-1}}$, which is much faster than the rate that the spike frequency changes during the temporal report (a decay rate on the order of approximately 1 s). This implies that a model based on synaptic conductance values equivalent to their steady state levels as defined in Eq. (9) should capture decay dynamics during the temporal report period well .

The formulation of Eq. (12) assumes only feed-forward input and is based on the relationship between pre-synaptic spiking activity and excitatory conductance (Eqs. (2) and (4)). In the fully recurrent network, however, the recurrent layer's output is also part of its own input. Since excitatory current is a function of both the synaptic weight and activation level, the loop between synaptic activation and spike frequency can be closed by replacing the generic W component of $g_E^\infty(\mu, W)$ from Eq. (12) with the value of the laterally recurrent weights, L.

The full dynamics of the spiking model are described by Eq. (1) and (3). Since the synaptic time constant is assumed to be significantly longer than the effective membrane time constant, the time course of synaptic activation will dominate membrane voltage dynamics. This suggests that the differential form of V (Eq. (1)) can be replaced with an instantaneous function of the average input rate. Accordingly, the dynamics of the recurrent network model during the falling phase can be described in terms of the synaptic activation variable, s, and recurrent weight value, L, by replacing the stochastic variable $X(t)$ from Eq. (4) with the output frequency calculated using the MFT I/O curve (Eq. (12)):

$$\frac{ds}{dt} = \phi(s, L)\rho(1 - s) - \frac{s}{\tau_s} \tag{15}$$

where $\phi(s, L)$ is equivalent to $\phi(\mu, W)$ from Eq. (12) with $g_E^\infty(\mu, W)$ replaced by sL. This pseudo-steady state approximation replaces the full system of equations with a single ODE and can be used as written or with a numerical estimate of ν replacing the analytical ϕ function .

5 Dynamics of encoded temporal reports

5.1 Comparison of mean field dynamics to full network dynamics

Equation (15) describes the derivative of synaptic activation as the difference between source term ($\phi(s, L)\rho(1 - s)$), and a negative sink term (s/τ_s). The relationship between these two terms as a function of s for different values of the recurrent coupling weight is shown graphically in Fig. 3(a). In the absence of

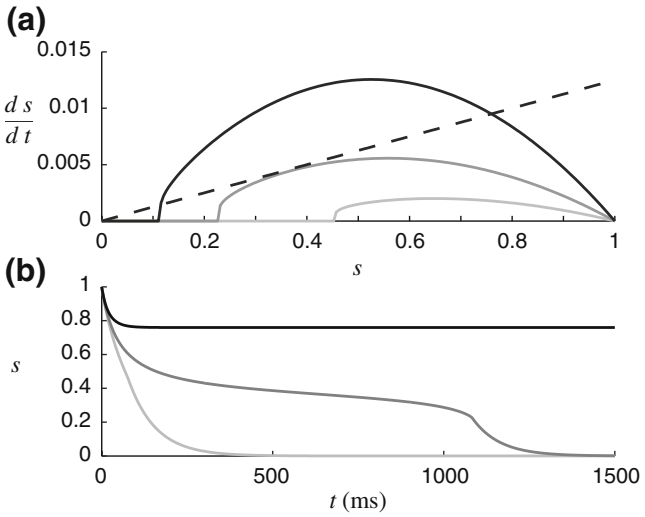

Fig. 3 Relaxation dynamics of reduced mean field model. (**a**) Source (*solid curves*) and sink (*black dashed line*) components of the pseudo-steady state equation (Eq. (15)) for three values of the excitatory recurrent weight. $L = 2.2\text{e-}3$ μS (*light gray*), 4.4e-3 μS (*gray*), and 8.8e-3 μS (*black*). (**b**) Resulting dynamics. Critical slowing occurs when recurrent weights move the positive component sufficiently close to the negative component. A stable "up" state appears if the weights grow large enough that the lines intersect

external input, the system relaxes to a stable steady state at zero as long as the negative component is larger than the positive component. The instantaneous decay rate is equal to the the difference between the two components. It is immediately clear from this plot that larger recurrent weights increase the response duration by moving the positive source curve closer to the sink line.

Network dynamics predicted by MFT are found by numerically integrating Eq. (15) from an initial condition of $s(0) = 1$. Dynamics calculated using the deterministic ϕ function for several values of L are shown in Fig. 3. These results demonstrate the mechanism responsible for forming temporal representations; as L increases, the positive and negative components of Eq. (15) get closer together, ds/dt gets smaller, and decay dynamics slow down. A temporal representation results when the recurrent network structure, in effect, creates a temporal bottleneck in the relaxation dynamics. If the recurrent weights are set too high, a stable fixed point appears at the upper intersection of the source and sink terms, corresponding to a high level of persistent firing (see Section 7).

The pseudo-steady state model explains several features of temporal representations seen in the spiking network that are absent in the linear model (Fig. 1). First, instead of decaying at a constant rate, spiking activity falls quickly to a plateau level following stimulation. From the MFT analysis it is evident that this results from the large gap, due to synaptic saturation, between the positive and negative curves at high s values. After the fast initial drop, activity in the spiking model decays slowly until it reaches some threshold level and then falls precipitously back to baseline levels. The rapidity of this fall depends on the shape of the I/O curve (Fig. 2). A sharp boundary between quiescence and spiking activity, as predicted by the MFT analysis, will result in a steep drop while the gradual transition seen using the numerical approach (which includes the noise dominated I/O region) will elicit a more gradual decay.

The dynamics in Fig. 3 look qualitatively similar to those seen in the full spiking model. Figure 4 demonstrates that the dynamics predicted by the pseudo steady-state model for two values of L accurately describe the dynamics of the full spiking model for the same values of L (scaled by the number of neurons responsive to the stimulus in the full model) both in terms of the synaptic activation variable and firing rates. The following relationship, found by setting Eq. (15) to

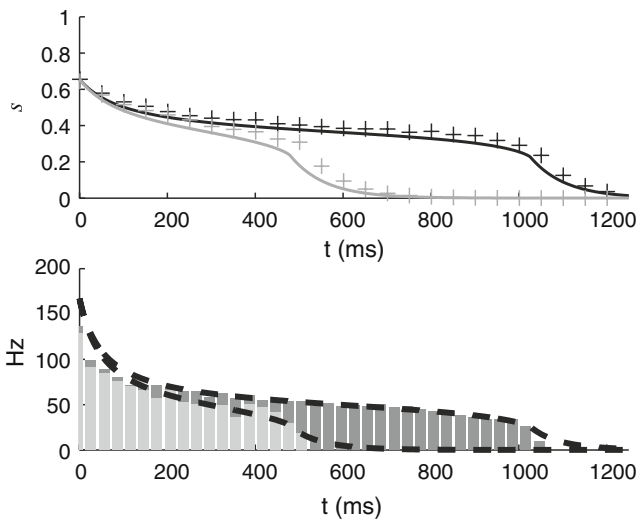

Fig. 4 Pseudo steady-state model prediction compared to full I&F model. The plots above show trajectories generated by solving the pseudo steady-state equation compared to values extracted from the full spiking network model for two values of lateral recurrent weights. In the *top plot*, the *solid lines* are the trajectories predicted by integrating Eq. (15) and the stars indicate the average synaptic activation variable of 100 neurons participating in a temporal representation over a single run. In the *bottom plot*, the *bars* show the PSTH of the spiking neurons overlaid with the spike frequencies predicted from the mean-field theory (*dashed black lines*). The initial condition $s(0) = 0.625$ was taken from the simulations of the complete network at the end of the stimulus period

zero, is used to convert from synaptic activation levels to spike frequencies:

$$\nu^{\infty} = \frac{s}{\tau_s \rho (1 - s)} \tag{16}$$

5.2 Limit of encodable time

We can heuristically define the "encoded time" of our network in relation to Eq. (15). Here, T_E is the time required for the network to relax back to some value close to zero following stimulation. With this definition relaxation time depends on the value of s at the end of the stimulus presentation period, but it makes sense to assume a starting point corresponding to full activation in order to define a measure of the encoded temporal representation. The existence of an exact solution to Eq. (15) will depend on the form of $\phi(s, L)$, but we can solve for T_E by numerically integrating from 1 to some value close to 0. The result of this calculation is shown in Fig. 5(a).

With the parameters used in this paper, our model is limited to maximum temporal representations on the order of 1–2 s. This is approximately the same response duration reported by Shuler and Bear, although it is not known wether this duration represents an upper limit in V1 or is an artifact of the specific stimulus-reward offset pairing presented during training.

5.3 Invariance of temporal report to stimulus intensity

An interesting observation from Fig. 3 is that the "temporal report" (that is, the duration of the plateau during which s decays very slowly) occurs over a very narrow range of s values. Any initial activation greater than the maximal plateau value will report approximately the same interval. Conversely, any activation below the fall-off threshold will report no temporal representation. This means that the network will reliably report encoded temporal values so long as stimulation is sufficiently robust to raise the activity level high enough. This is shown graphically in Fig. 5(b).

There are two implications of this threshold-invariance that may be of importance in biological networks: (1) a reliable temporal report requires only a vigorous query of the trained network and not a carefully graded stimulus (2) sub-threshold response dynamics of individual neurons in the trained network will be no different from those in the naive network. Since temporal reports require coincident activations of an ensemble of neurons, temporal representations could conceivably exist on top of other network structures without changing neural dynamics in the nominal

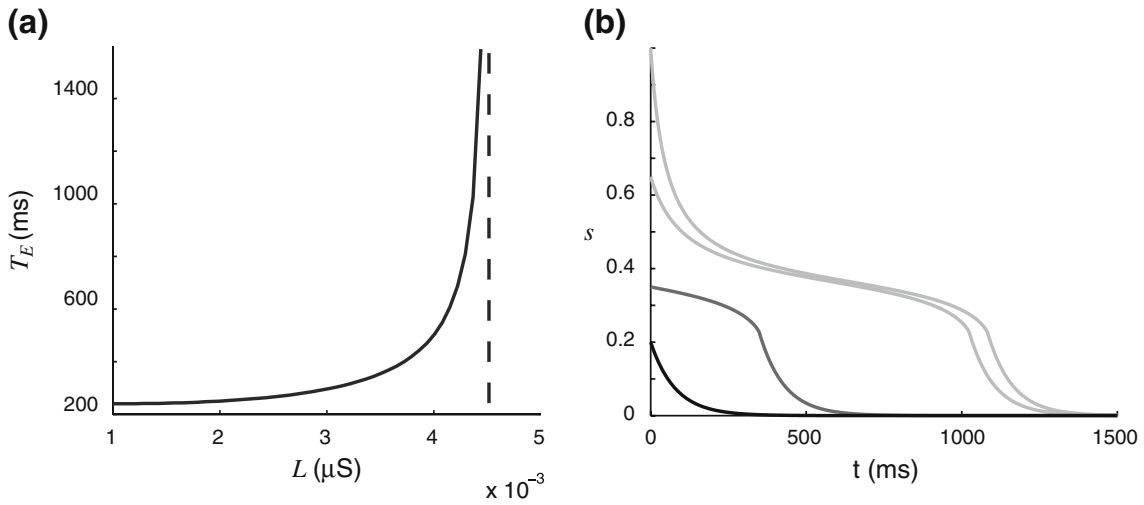

Fig. 5 (**a**) The duration of encoded time estimated by integrating s from 1 to 0.05. With the parameters used in this paper, encoding times above approximately 1.5 s becomes difficult due to the sensitivity of the encoded representation to very small changes in L. (**b**) Invariance of temporal representation to stimulus magnitude. This plot shows the pseudo-steady state system response of a single "trained" network with initial conditions representing different stimulus levels. If vigorous stimulation drives the network to a sufficiently high level, the temporal report is approximately the same (*light gray lines*). An intermediate stimulus show a degraded temporal report (*gray*), and the response to low-level stimulation (*black*) is identical to the report of an isolated neuron. Temporal reports above a threshold value of $s(0) \approx 0.45$ are very similar

activity range and would not be evident without the correct querying stimulation pattern.

6 Dynamics and steady state with spontaneous activity

The analysis presented above assumes that activity dynamics during the decay phase are set only by synaptic connections within the recurrent layer and clearly describes the mechanism responsible for creating temporal representations in our model and the relationship between recurrent weights and decay period dynamics. In principle, the same analysis will also describe changes in steady state firing rates so long as the I/O curve includes an accurate description of the fluctuation dominated region where spontaneous activity occurs (see Fig. 2). In our previous work (Gavornik et al. 2009), spontaneous activity was simulated by including independent excitatory feed-forward synapses into the recurrent population. As shown in Fig. 6, training a recurrent network on a timing task increases the rate of spontaneous activity as the magnitude of the recurrent weights grows even though the strength of the feed-forward synapses driving the activity do not change.

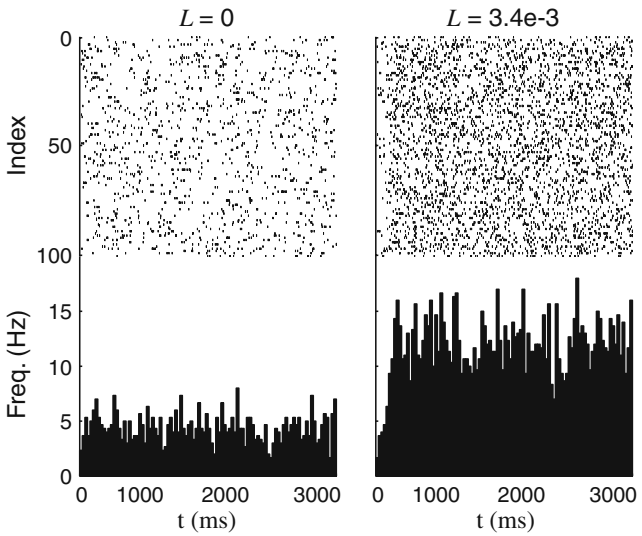

Fig. 6 Spontaneous activity in the full network model. Here, spontaneous spiking in the recurrent layer is driven by stochastic feed-forward synapses each with maximal conductance of 2.1e-2 μS and a synaptic time constant of 10 ms. The synapses are driven by independent poisson spikes arriving at and average rate of 12.5 Hz. These plots (raster plots for each neuron in the recurrent layer over a binned histogram showing firing rate) show that the spontaneous activity rate driven by these inputs increases from approximately 4 Hz in a network with no recurrent connections to ≈12 Hz when the total value of $L = 3.4\text{e-}3$ μS. A similar increase in the spontaneous firing rate was also seen in the experimental data

The analytical solution (Eq. (12)) considers activity through a class of synapses with a single time constant, all of which spike at the same rate. While this approach is sufficient to account for the recurrent activity in our network, where all synapses have the same time constant, it does not capture changes in the low spontaneous firing rate resulting from synapses with different time constants that spike at a different rate from the recurrent layer neurons. Although a general analytical solution to this problem is difficult to obtain, dynamics can be approximated using the pseudo-steady state approach by generating numerical estimates of the I/O curves that include both recurrent and feed-forward synapses. As in the full network model shown in Fig. 6, spontaneous activity is generated by feed-forward synapses with a time constant $\tau_f = 10$ ms and maximum conductance of 2.1e-2 μS driven independently by 12 Hz Poisson spikes. Dynamics are calculated by replacing the ϕ function in Eq. (15) with numerical estimates of ν as a function of μ.

As seen in Fig. 7(a), the source term is simply proportional to $(1 - s)$ when $L = 0$. For $L > 0$, the source curve is much more complicated. This complexity results from the inclusion of two different time constants associated with the feed-forward and feedback components of excitatory current and will not be further discussed here. Since the feed-forward activity requires a high variability to produce a high CV in the spontaneous activity (cortical neurons have a CV that is close to 1), the numerical estimates of ν are noisier than those shown in Fig. 2. Steady state activity levels are found at the intersection of the source and sink curves, and in both cases match those seen in the full network (predicted spontaneous rates of ≈4.2 Hz for $L = 0$ and ≈12.5 Hz for $L = 3.4\text{e-}3$ μS). Figure 7(b) demonstrates that the dynamics calculated using the numerical ν estimate match the dynamics from the full network model very well.

7 Bistability in the recurrent network

It is clear from Fig. 3 that the pseudo steady state model has a single fixed point at $S = 0$ when L is low. If L is increased beyond some threshold value, however, the positive and negative components of Eq. (15) will intersect and the system will have three fixed points. For example, the black source curve in Fig. 3(a) intersects the dashed sink line at three points. In this case the system is bistable and may, depending on the activation

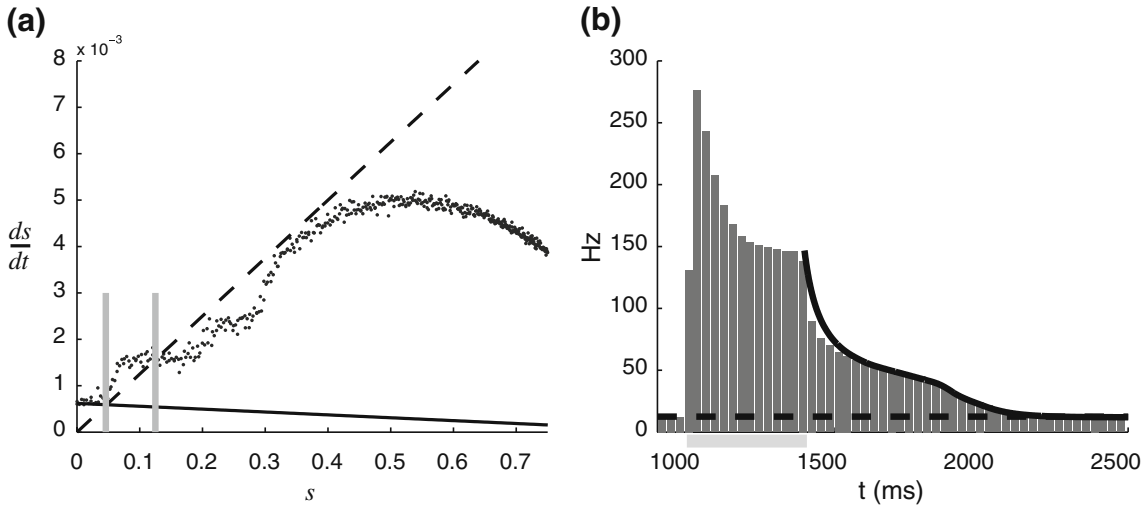

Fig. 7 Accounting for changes in spontaneous activity. (**a**) The sink (*dashed black line*) and source curves for the cases that $L = 0$ (*solid black line*) and $L = 3.4\text{e-}3$ μS (*black points*, numerical estimates made over multiple simulation runs). Spontaneous activity occurs at, and is set by, the intersection of the source and sink curves (marked for both L values by two *vertical gray lines*). This plot shows that the spontaneous activity level increase from 4.2 Hz ($s = 0.046$) to 12.5 Hz ($s = 0.125$) as a function of recurrent excitation. (**b**) Dynamics during the decay period (*solid black line*) calculated using this numerical curve and Eq. (15) match those seen in the full network model (*gray bars* show spike frequency in the network averaged over 50 runs over the period of stimulation, marked by the light gray bar, and temporal report). The *dashed line* marks the spontaneous activity level estimated from the full network and matches the value predicted from (**a**)

level, move towards a high "up" state (set by the upper intersection) rather than decaying towards zero (the intersection at the origin). The RDE timing model analyzed in this work requires the recurrent weights be set below the threshold value so that the system will always decay to a single fixed point.

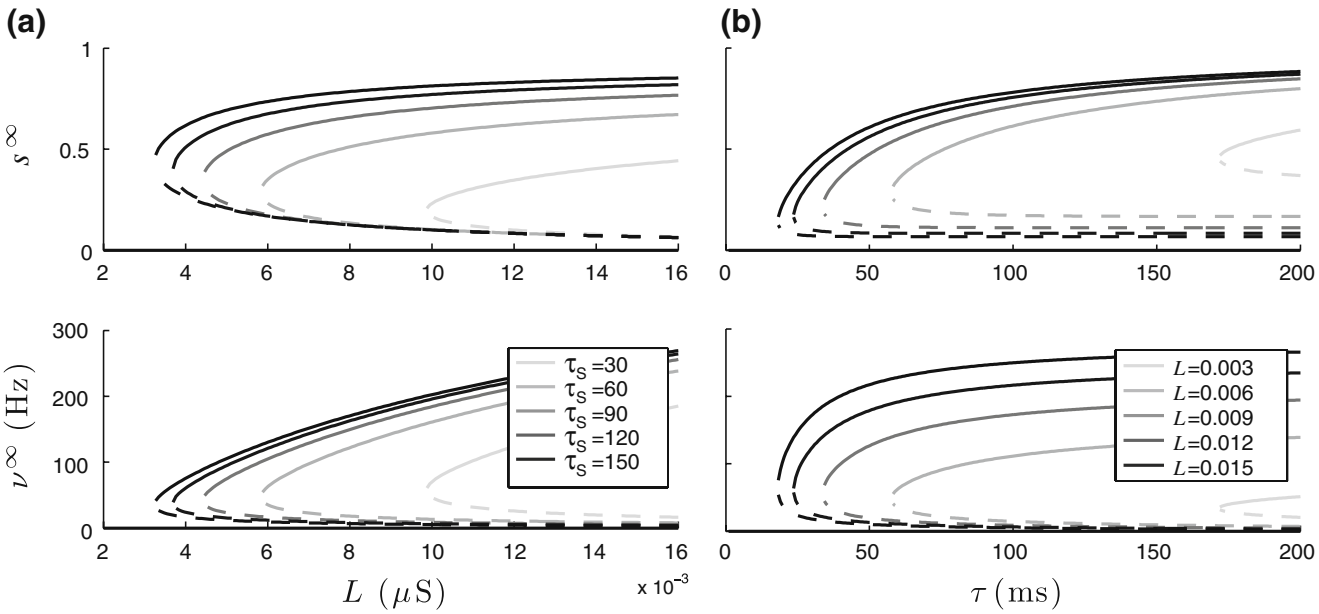

Fig. 8 Bifurcation diagrams of the fixed-point values from the pseudo steady-state model (Eq. (15)) demonstrate bistability. *Solid lines* indicate stable fixed points, and *dotted lines* indicates unstable fixed points. The *top* plots shows the synaptic activation variable at steady states (s^∞) and the *bottom* plots show the resultant firing rate (ν^∞). (**a**) Bistability as a function of lateral recurrent weights (L) over a range of τ_s values. (**b**) Bistability as a function of τ_s for several fixed values of L. With zero spontaneous activity, all solutions have fixed point at $s = 0$

The MFT model can be used to demonstrate and analyze the range of bistability as a function of L. The zeros of Eq. (15) (that is, $\frac{ds}{dt} = 0$) were found numerically while varying L and holding all other parameters constant. This analysis was repeated over a range of τ_s values. The resulting bifurcation diagram is shown in Fig. 8(a). The spike frequencies for each fixed point value of s were also calculated using Eq. (16). As expected, the system is monostable with a solution at $s = 0$ for low values of L; above some threshold, a second stable steady state emerges. This "up" state corresponds to persistent firing at a fixed rate, thought to underlie the persistent activity associated with working memory (Wang 2001; Miller et al. 2003). Note that although the lowest possible stable "up" state s value increase with τ_s, the minimum possible spike frequency decreases. This accords with previous findings that slow NMDA currents are critically important in working memory models that spike at biologically plausible frequencies (Wang 2001; Seung et al. 2000).

Bistability can also emerge as a function of other key parameters including τ_s. As before, a bifurcation diagram was generated by finding fixed points numerically over a range of τ_s with several fixed values of L. The results are shown in Fig. 8(b). Again, the system is monostable below a threshold and a bistable, with steady solutions at zero and a high "up" state.

8 Discussion

Our previous work explains qualitatively how temporal representations of the type reported by Shuler and Bear (2006) can form in local recurrent networks as a function of reward modulated synaptic potentiation (Gavornik et al. 2009). It could not explain the quantitative form of representations that develops when RDE is applied to a network of spiking neurons. Specifically, analysis based on a linear neuron model fails to explain why the spike rate falls so precipitously immediately after feed-forward stimulation ends, the rapid drop back to baseline firing rates at the end of the temporal report, and the increase in spontaneous firing rates that occurs with training. By reducing the stochastic system of 100 coupled non-linear differential equations to a single deterministic ODE, the pseudo-steady state model based on the MFT approach (Eq. (15)) explains these features. Temporal representations form in networks of non-linear spiking neurons when the level of recurrent excitatory feedback is slightly less than the intrinsic neural activity decay rate, resulting in a critical dynamical slowdown. The particular shape of the temporal representation, as seen in Figs. 1, 4 and 7 depends on the non-linear input-output relationship of the individual neurons. Unlike in the linear case, where activity decays exponentially, spiking temporal representations are relatively invariant to stimulus magnitude above some threshold (Fig. 5(b)).

It should be noted that the analytical MFT approach described here considers only first order spike statistics. While the mean-based analytical solution of the input-output curve matches the numerically extracted curve well above the spike-frequency threshold (Eq. (13)) it can not account for the sub-threshold region where relatively low-rate spiking is driven by input fluctuations (Fig. 2). The agreement between the encoded time predicted by the analytical mean-field approximation and the full spiking model seen in Fig. 4 indicates that the impact of this omission on dynamics is minimal with this parameter set. This can be understood since the noise-dominated region of the ϕ function exists primarily at activation levels below the narrow bottleneck responsible for the the critical slowing, as seen in Fig. 3(a). Since the analytical solution predicts zero output for low input spike frequencies it can not explain changes in spontaneous activity levels when each neuron in the recurrent layer is driven by low levels of random feed-forward activity. The pseudo-steady state model can accurately predict these changes if the analytically calculated ϕ function in Eq. (15) is replaced by a numerical estimate of the I/O curve that is generated including the feed-forward input (Fig. 7). The same approach would also work with an analytical I/O curve including an accurate description of the noise dominated spike region, although the calculation of such for the conductance based neuron model used here is beyond the scope of this work.

The CVs of neurons in our network are close to 1 when spontaneously active and drop to a value closer to 0.4 following stimulation. A similar phenomenon, whereby external stimulus onset decreases neural spiking variability, has been reported in various brain regions (Churchland et al. 2009). The CVs of V1 neurons during the temporal report period have not been experimentally characterized, although they are likely to be higher than exhibited in our spiking network. Previous models have shown bistability with a high CV (Barbieri and Brunel 2008; Roudi and Latham 2007). For example, a balanced network models including high reset values and short term synaptic depression or neuronal adaptation can exhibit high CVs concurrent with stored memory retrieval and bistable "up" states (Barbieri and Brunel 2008). It is an open question how the inclusion of the additional mechanisms used in these models to increase variance would impact our timing model, which depends on the

development of regular slow dynamics near bifurcation points.

Figure 8 demonstrates that the mechanism underlying RDE has the potential to create regions of bistability. In the bistable regime, our model becomes very similar to models of persistent activity thought to underlie working memory processes (Lisman et al. 1998; Wang 2001; Seung 1996; Brody et al. 2003; Miller et al. 2003). A learning rule similar to RDE could be used to tune a recurrent network to produce desired "up" levels.

It has been suggested that temporal processing might involve the same mechanisms that underlie working memory (Lewis and Miall 2006; Staddon 2005), and correlates of working memory have been observed in the monkey visual cortex (Super et al. 2001). Despite this, there is no experimental evidence that the high stable state is reached, or used, by the neurons in V1. Functionally, it is unlikely that persistent firing in the low level visual processing areas resulting from a brief stimulus would be desirable for the visual system. Presumably, homeostatic mechanisms not included in our model could ensure that the up state is never reached in V1. The bifurcation analysis helps set upper limits on the allowable range of parameters in the recurrent network over which the model of temporal representation is valid. Based on the similarity between our RDE model of temporal representation and previous models of working memory, it is possible that similar neural machinery is recruited for both tasks by the brain.

Our model, as presented and analyzed in this work, contains no role for inhibition. Indeed, RDE posits that recurrent excitation provides the neural basis for temporal representations. We have previously demonstrated that recurrent excitation can overcome both feed-forward and recurrent inhibition and that RDE is able to entrain temporal representations in networks that include biologically realistic ratios of excitation and inhibition (Gavornik 2009). Recurrent network models including significant amounts of inhibition have been successfully analyzed using the mean field approach (Brunel and Wang 2001; Renart et al. 2003, 2007).

Another possible way to explain the data in Shuler and Bear (2006) is that temporal representations form within individual neurons independent of network structure. In an accompanying work (Shouval and Gavornik 2010), we present a model demonstrating how single neurons can learn temporal representations through reward based modulation of their intrinsic membrane conductances. The mechanistic basis of the prolonged spiking in this alternate model is a positive feedback loop between voltage gated calcium channels and calcium dependent cationic channels. Although analysis of the single cell model reveals mathematical similarities with the mean field approach presented here, there are functional differences between the two models that can be explored experimentally. It is also possible that temporal representations in the brain could form through a hybridization of the two models.

The success of RDE in explaining how temporal representations can arise in local cortical networks, such as V1, suggests that neural mechanisms responsible for temporal processing may be more distributed throughout the brain than previously thought. Additional reports of persistent activity in other sensory cortices may further support this hypothesis. The mechanistic similarity between the temporal representations postulated by RDE and those thought to contribute to working memory could imply that a simple mechanism, widely available throughout the cortex, can be recruited for widely different tasks. This work demonstrates that a MFT approach can explain how temporal representations can form in networks of non-linear spiking neurons.

References

Amit, D. J. (1989). *Modeling brain function: The world of attractor neural networks*. Cambridge [England], New York: Cambridge University Press. 89015741 Daniel J. Amit. ill.; 24 cm. Includes bibliographies and index.

Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex, 7*(3), 237–252.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985). Spin-glass models of neural networks. *Physical Review A, 32*(2), 1007–1018.

Barbieri, F., & Brunel, N. (2008). Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? *Front Neuroscience, 2*(1), 114–122.

Brody, C. D., Romo, R., & Kepecs, A. (2003). Basic mechanisms for graded persistent activity: Discrete attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology, 13*(2), 204–211.

Brunel, N., & Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience, 11*(1), 63–85.

Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: A quantitative analysis of connectivity. *Proceedings of the Royal Society B: Biological Sciences, 265*(1400), 1037–1044.

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2009). Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nature Neuroscience, 13*(3), 369–378.

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex, 10*(9), 910–923.

Eliasmith, C. (2005). A unified approach to building and controlling spiking attractor networks. *Neural Computation, 17*(6), 1276–1314.

Gavornik, J. P. (2009). *Learning temporal representations in cortical networks through reward dependent expression of synaptic plasticity*. Ph.D. dissertation, The University of Texas at Austin.

Gavornik, J. P., Shuler, M. G., Loewenstein, Y., Bear, M. F., & Shouval, H. Z. (2009). Learning reward timing in cortex through reward dependent expression of synaptic plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 106*(16), 6826–6831.

Johnson, R. R., & Burkhalter, A. (1996). Microcircuitry of forward and feedback connections within rat visual cortex. *Journal of Comparative Neurology, 368*(3), 383–398.

Lewis, P. A., & Miall, R. C. (2006). A right hemispheric prefrontal system for cognitive time measurement. *Behavioural Processes, 71*(2–3), 226–234.

Lisman, J. E., Fellous, J. M., & Wang, X. J. (1998). A role for NMDA-receptor channels in working memory. *Nature Neuroscience, 1*(4), 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*(5712), 1121–1124.

Mastronarde, D. N. (1987). Two classes of single-input X-cells in cat lateral geniculate nucleus. II. Retinal inputs and the generation of receptive-field properties. *Journal of Neurophysiology, 57*(2), 381–413.

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience, 27*, 307–340.

Miller, P., Brody, C. D., Romo, R., & Wang, X. J. (2003). A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cerebral Cortex, 13*(11), 1208–1218.

Moshitch, D., Las, L., Ulanovsky, N., Bar-Yosef, O., & Nelken, I. (2006). Responses of neurons in primary auditory cortex (A1) to pure tones in the halothane-anesthetized cat. *Journal of Neurophysiology, 95*(6), 3756–3769.

Renart, A., Brunel, N., & Wang, X. (2003). Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. In J. Feng (Ed.), *Computational neuroscience: A comprehensive approach* (pp. 431–490). Boca Raton: CRC Press.

Renart, A., Moreno-Bote, R., Wang, X. J., & Parga, N. (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Computation, 19*(1), 1–46.

Roudi, Y., & Latham, P. E. (2007). A balanced memory network. *PLoS Computational Biology, 3*(9), 1679–1700.

Seung, H. S. (1996). How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences of the United States of America, 93*(23), 13339–13344.

Seung, H. S., Lee, D. D., Reis, B. Y., & Tank, D. W. (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron, 26*(1), 259–271.

Shouval, H. Z., & Gavornik, J. P. (2010). A single cell with active conductances can learn timing and multi-stability. *Journal of Computational Neuroscience*. doi:10.1007/s10827-010-0273-0.

Shuler, M. G., & Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science, 311*(5767), 1606–1609.

Staddon, J. E. (2005). Interval timing: Memory, not a clock. *Trends in Cognitive Sciences, 9*(7), 312–314.

Super, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science, 293*(5527), 120–124.

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistentactivity. *Trends in Neurosciences, 24*(8), 455–463.