Power-Law Neuronal Fluctuations in a Recurrent Network Model of Parametric Working Memory

Paul Miller and Xiao-Jing Wang

J Neurophysiol 95:1099-1114, 2006. First published Oct 19, 2005; doi:10.1152/jn.00491.2005

J Neurophysiol 95: 1099–1114, 2006.
First published October 19, 2005; doi:10.1152/jn.00491.2005.

Power-Law Neuronal Fluctuations in a Recurrent Network Model of Parametric Working Memory

Paul Miller and Xiao-Jing Wang
Volen Center for Complex Systems, Brandeis University, Waltham, Massachusetts

Submitted 11 May 2005; accepted in final form 14 October 2005

Miller, Paul and Xiao-Jing Wang. Power-law neuronal fluctuations in a recurrent network model of parametric working memory. *J Neurophysiol* 95: 1099–1114, 2006. First published October 19, 2005; doi:10.1152/jn.00491.2005. In a working memory system, persistent activity maintains information in the absence of external stimulation, therefore the time scale and structure of correlated neural fluctuations reflect the intrinsic microcircuit dynamics rather than direct responses to sensory inputs. Here we show that a parametric working memory model capable of graded persistent activity is characterized by arbitrarily long correlation times, with Fano factors and power spectra of neural activity described by the power laws of a random walk. Collective drifts of the mnemonic firing pattern induce long-term noise correlations between pairs of cells, with the sign (positive or negative) and amplitude proportional to the product of the gradients of their tuning curves. None of the power-law behavior was observed in a variant of the model endowed with discrete bistable neural groups, where noise fluctuations were unable to cause long-term changes in rate. Therefore such behavior can serve as a probe for a quasi-continuous attractor. We propose that the unusual correlated fluctuations have important implications for neural coding in parametric working memory circuits.

INTRODUCTION

Neuronal activity in the cerebral cortex is highly irregular (Buzsaki 2004; Compte et al. 2003; Shadlen and Newsome 1994 Softky and Koch 1993). Fluctuations in spike discharges are often correlated between simultaneously recorded cells on a trial-by-trial basis (Bair et al. 2001; Gawne and Richmond 1993; Lee et al. 1998; Zohary et al. 1994); this covariance greatly impacts the extent to which noise can be averaged out over a neuronal pool and so determines the accuracy with which a stimulus feature can be extracted from a large population of neurons (Abbott and Dayan 1999; Averbeck and Lee 2004; Shadlen and Newsome 1996; Shamir and Sompolinsky 2004; Sompolinsky et al. 2002). Correlations arise from common inputs caused by both overlapping afferents and local synaptic interconnections. For example, cells in the primary visual cortex become correlated because of shared thalamic inputs as well as lateral connections within the cortex. Exactly how correlated noise enhances or decreases the coding efficiency for an oriented visual stimulus critically depends on the (feedforward vs. recurrent) network architecture (Seriès et al. 2004).

Correlated fluctuations provide a valuable probe into the dynamic nature of a neural network. For responses of sensory cells to external stimuli, the effects of direct sensory inputs and local connectivity are confounded. In contrast, neurons in working memory systems exhibit persistent activity in the absence of external stimulation, when the subject must hold information transiently in the mind. Hence fluctuations of persistent activity (Compte et al. 2003; Constantinidis and Goldman-Rakic 2002) are believed to result from the intrinsic dynamics of a working memory microcircuit. The cellular mechanisms of mnemonic persistent activity represent a topic of intense current interest (Brody et al. 2003; Major and Tank 2004; Wang 2001). Of special interest are working memory circuits that encode an analog quantity, such as those integrating a velocity signal into a persistent change in position or direction (Aksay et al. 2000; Taube and Bassett 2003) or storing in active short-term memory the spatial location or temporal frequency of a sensory stimulus (Funahashi et al. 1989; Romo et al. 1999). After a transient input, neurons in such a network exhibit sustained changes of activity that are tuned with a stimulus feature in a graded manner. This suggests the system could possess a continuous family of stable self-sustained activity states. Neuronal tuning curves that arise from such activity states are often bell-shaped functions or monotonic functions of the encoded quantity.

For a parametric working memory system, if the delay firing rates of two monotonically tuned neurons are plotted against each other, a curved line results, as shown in Fig. 1*A*. Such a curved line is referred to as a line attractor (Seung 1996). The realization of a continuum of monotonically tuned persistent firing states requires fine-tuning of network parameters (Miller et al. 2003; Seung et al. 2000b). In an alternative scenario, the network could display multiple persistent states, not truly a continuum, by virtue of a set of discrete stable activity patterns that can be robustly realized without fine-tuning of parameters (Goldman et al. 2003; Koulakov et al. 2002). Different stimuli cause the network to change discontinuously from one pattern of activity to another, whereas background noise does not cause a switch in the state of persistent activity. Hence the discrete network model has a limited sensitivity to small differences in stimuli, but is robust to noise in the circuit. Whether the continuous or discrete model better describes neural integrators remains an open question (Major and Tank 2004).

Thus far, most experimental and computational studies have been concerned with the trial-averaged firing rates. Fluctuation analysis offers a different approach to characterize and potentially test distinct working memory models. In a continuous attractor network, noise is integrated in time, so the mnemonic activity could drift in the same manner as a random walk. Because correlation functions for a random walk do not decay exponentially with time (Mandelbrot and Ness 1968), we

Address for reprint requests and other correspondence: X.-J. Wang, Volen Center for Complex Systems, Brandeis Univ., Waltham, MA 02454 (E-mail: xjwang@brandeis.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

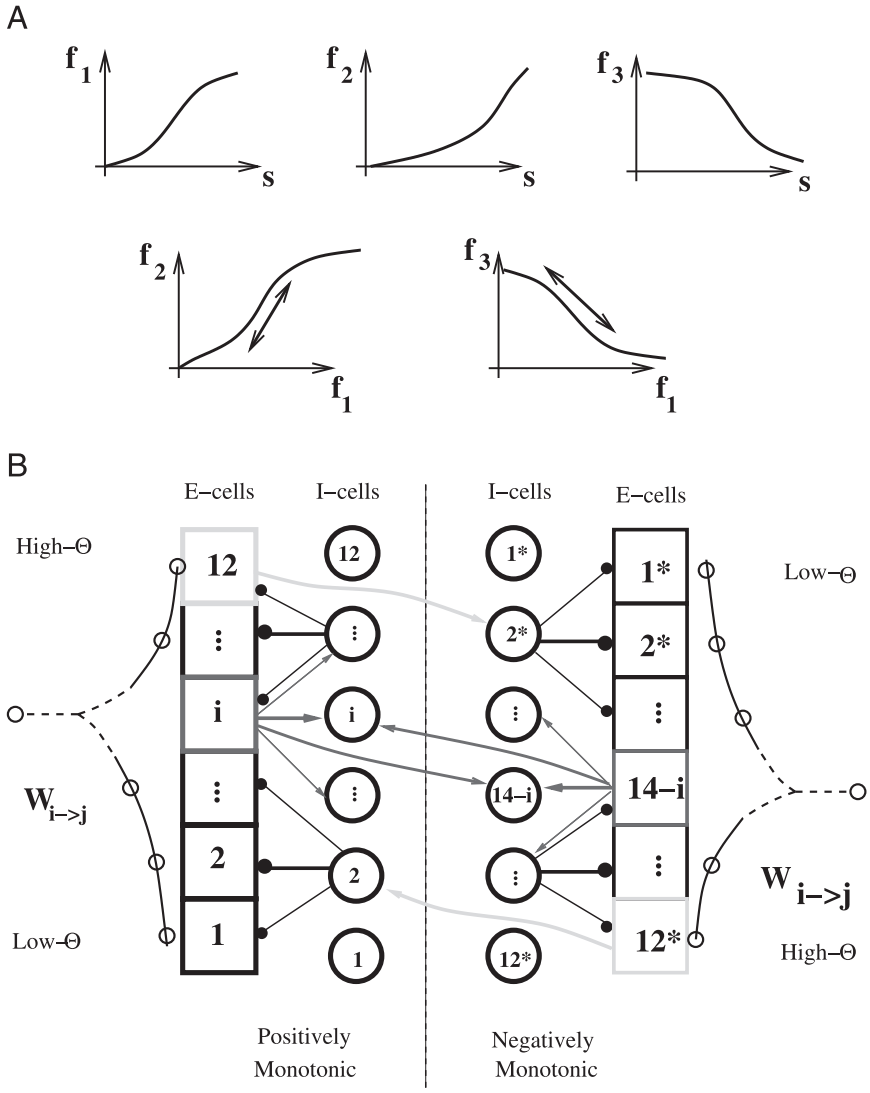

FIG. 1. Network properties. *A*: schematic representation of a line attractor. *Top*: tuning curves of 3 different neurons. *Bottom*: representation of the continuous attractor in the planes of pairs of neurons. Arrow indicates direction in which fluctuations are not rapidly damped in time, because they shift the system along the continuous attractor of the network. *B*: schematic model architecture. There are 2 networks of positively and negatively monotonic neurons, respectively. Each network has 12 excitatory pyramidal cell populations (squares) and 12 inhibitory interneuron populations (circles). Synaptic connections are stronger within the same population than between populations. Connectivity is asymmetrical, so that the activation threshold, Θ, by stimulus is the lowest for neural population 1 and highest for neural population 12. The 2 networks interact through pyramid-to-interneuron connections, resulting in cross-inhibition. Stimulus inputs are given equally to all excitatory cells within a network, but they differ across the 2 oppositely tuned networks.

reasoned that unusual fluctuation properties should be a conspicuous feature of continuous "parametric" memory networks, in which mnemonic neural firing is monotonically tuned to an analog quantity (Aksay et al. 2000; Miller et al. 2003; Romo et al. 1999; Seung 1996; Seung et al. 2000b). A more robust discrete integrator network (Goldman et al. 2003; Koulakov et al. 2002) would not show the same random drifts in response to noise, unless the noise were strong enough to cause transitions between the multiple discrete states (Miller 2006). If the scale of noise fluctuations sets the coarseness of the system, a system is considered as quasi-continuous whenever the difference in discrete rates is less than the noise-induced variation in rate during a trial.

The purpose of this study is to analyze temporal fluctuations in firing patterns of working memory models with monotonically tuned persistent states. We show that power-law fluctuations are a salient feature of continuous attractor networks.

METHODS

Computational network model

We developed a cortical microcircuit model for the task of parametric working memory, as published previously (Miller et al. 2003). The task requires macaque monkeys to encode and maintain in memory a vibrational frequency, before comparison with a second vibrational frequency. In the secondary somatosensory cortex of a monkey, the spiking response of a neuron to the vibrotactile stimuli varies either positively monotonically or negatively monotonically with the stimulus frequency. Neurons in the prefrontal and premotor cortices, which receive inputs from the secondary somatosensory cortex, show tuned mnemonic activity that can be sustained for up to 6 s during a delay after the end of the initial stimulus (before a 2nd comparison stimulus). Our model network is intended to reflect the activity of prefrontal cortical neurons. We included two sets of populations, corresponding to sets of neurons that receive inputs separately from the secondary somatosensory cortex. One set of inputs is positively monotonically tuned while the other set is negatively monotonically tuned to the vibrational stimulus frequency.

We used interconnected integrate-and-fire neurons, grouped into populations that had strong intrapopulation connections and weaker connections between populations. The connection strengths decrease exponentially with the difference between population numbers (Fig. 1*B*). The decay in connection strength was more gradual from high-to-low population numbers, so that the neurons in populations with low index received more excitatory current from other cells than those with higher index. This led to a range of thresholds for responses to external input, with the low-index populations being the most excitable and having the highest spontaneous rates.

We couple the two sets of oppositely tuned populations by cross-inhibition. The positively monotonic populations are labeled 1 for the most excitable to 12 for the least excitable, and the negatively monotonic neurons are labeled 1* for the most excitable to 12* for the least excitable. Cross-inhibition is not uniform, but strongest between populations labeled by i and $(14 - i)$*, where i runs from 2 to 12. Hence the increase in activity of a highly excitable positively monotonic population, after a weak stimulus, helps to reduce the activity of a less excitable negatively monotonic population. A larger stimulus results in a strong increase in activity for the less excitable positively monotonic populations and a concurrent decrease in activity of the more excitable negatively monotonic populations.

We realized a continuous attractor, with moderate recurrent excitation, where the total synaptic weight from all other excitatory populations is of the same order as the self-excitation within a population. For each population, the background external input amplitude and intrapopulation recurrent excitation strength were adjusted to a particular combination, corresponding to a point called a "cusp" in the two-dimensional parameter space, indicated by C in Fig. 2*A*. For a strength of recurrent excitation greater than at the cusp, bistability can exist between distinct up (persistently active) and down (resting) states, with a gap between them (A in Fig. 2*A*). For weaker recurrent excitation than at the cusp, the population firing rate varies smoothly with the excitatory drive (B in Fig. 2*A*). At the cusp, there is a possibility of an approximately vertical line segment, indicating a range of stable firing rates for the population with a fixed background excitatory input (position C, Fig. 2*A*, *right*).

We generated a set of discrete attractors for robust working memory (Koulakov et al. 2002) by creating many bistable populations with a range of thresholds. We achieved this by increasing the recurrent connection strength within each population while reducing the connection strengths between populations (see point A in Fig. 2*A*) and increasing the average leak conductances. The bistable populations each converge onto a readout population of 400 neurons, whose firing rates encode in a step-like manner the number of active bistable populations. We use the statistical properties of the readout cells for a fair comparison between the discrete model and the continuous model. The set of bistable populations should just be considered as one particular mechanism for generating the more realistic properties of the readout cells, which are suitable for experimental comparison.

We chose stimulus strengths for the discrete attractor, so that each stimulus would cause the discrete system to be in the same state for all trials with that stimulus. In this way, we can use the discrete system as a control to show that typical networks with a single stable state after a stimulus do not exhibit the same type of fluctuations seen in the continuous attractor network.

Single neuron parameters

Individual neurons are simulated using the single-compartment leaky-integrate-and-fire model (Tuckwell 1988), such that the membrane potential, V_i, of cell, i, follows the current-balance equation

$$C_m \frac{dV_i}{dt} = -g_L(V_i - V_L) - g_E S_{E,i}(V_i - V_E) - g_I S_{I,i}(V_i - V_I) - g_{ext} s_{ext,i}(V_i - V_E) - g_{cue} s_{cue}(V_i - V_E) \quad (1)$$

where C_m is the total membrane capacitance, g_L is the leak conductance, V_L is the leak potential, g_E and V_E are the conductance and reversal potential for excitatory channels and g_I and V_I are the conductance and reversal potential for inhibitory channels, respectively. g_{ext} and g_{cue} are the fixed conductances for background noisy input and applied, stimulus-dependent input, respectively, whereas s_{ext} and s_{cue} are the corresponding time-dependent gating variables (see *Eqs. 5* and *6*). When the membrane potential reaches a threshold, V_{thr}, the neuron spikes, and the membrane potential is reset at V_{reset} for an absolute refractory period, τ_{ref}, before continuing to follow *Eq. 1*.

The total synaptic drive for excitation or inhibition (S_E or S_I) is the sum of synaptic inputs from all presynaptic neurons j

$$S_i = \sum_j W_{j\to i} s_j(t) \quad (2)$$

where $W_{j\to i}$ is the relative synaptic weight from cell j to cell i, and s_j is the synaptic current gating variable activated by the presynaptic neuron j firing spikes at times $t_{spike,j}$. Specifically, for excitatory synapses, we have

$$\frac{ds_j}{dt} = \alpha_s \times \overline{P_R(t)}(1 - s_j)\delta(t - t_{spike,j}) - \frac{s_j}{\tau_s} \quad (3)$$

and for inhibitory synapses

$$\frac{ds_j}{dt} = \delta(t - t_{spike,j}) - \frac{s_j}{\tau_s} \quad (4)$$

with synaptic time constants τ_s. The probability of vesicular release, $\langle P_R(t)\rangle$, is described in the next subsection.

Background noisy input to all neurons is simulated using uncorrelated Poisson spike trains at a rate, r_{ext}, through nonsaturating synapses, of conductance g_{ext}, which are gated according to

$$\frac{ds_{ext}}{dt} = \delta(t - t_{spike,ext}) - \frac{s_{ext}}{\tau_{ext}} \quad (5)$$

with synaptic time constant τ_{ext} following spikes at times, $t_{spike,ext}$.

Similarly, during the stimulus, Poisson spike trains of rate, λ, generate additional excitation through α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor (AMPAR)-mediated synapses of conductance, g_{cue}, multiplied by a gating variable, s_{cue}, which follows

$$\frac{ds_{cue}}{dt} = \delta(t - t_{spike,cue}) - \frac{s_{cue}}{\tau_{ext}} \quad (6)$$

The rate, λ, represents the combined spike rate of multiple afferents from secondary somatosensory cortex, and increases or decreases linearly with stimulus frequency for the positively monotonic or negatively monotonic populations, respectively.

In the network models presented here, background and stimulus inputs are mediated by α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptors, with τ_{ext} = 2ms, recurrent excitation through *N*-methyl-D-aspartate (NMDA) receptors (Wang 1999, 2001) with τ_s = 100 ms and V_E = 0 mV, and inhibition through $GABA_A$ receptors with τ_s = 10 ms and V_I = −70 mV. In the continuous attractor network, the cellular parameters are as follows for excitatory cells: C_m = 0.5 nF, g_L = 38.4 nS, V_L = −70 mV, V_{reset} = −60 mV, V_{thr} = −45 mV, τ_{ref} = 2 ms, g_{ext} = 6 nS, r_{ext} = 1.2 kHz, g_E = 36 nS, g_I = 12 nS. For inhibitory cells, they are as follows: C_m = 0.2 nF, g_L = 17.6 nS, V_L = −70 mV, V_{reset} = −60 mV, V_{thr} = −50 mV, τ_{ref} = 1 ms, g_{ext} = 1.6 nS, r_{ext} = 1.8 kHz, $g_E = g_{cue}$ = 36 nS, g_I = 12 nS. The discrete integrator network has a range of leak conductances for the excitatory cells, equally spaced from g_L = 30.4 nS for cells in population-1 to g_L = 40 nS for cells in population-12. The inhibitory neurons have g_L = 20 nS, g_{ext} = 3 nS, and r_{ext} = 1.0 kHz. Otherwise single-cell properties are the same as in the continuous attractor network.

Short-term plasticity of excitatory synapses

All excitatory synapses exhibit short-term presynaptic facilitation and depression (Hempel et al. 2000; Varela et al. 1997). We implement the scheme described by Matveev and Wang (2000), which assumes a docked pool of vesicles containing neurotransmitter, where each released vesicle is replaced with a time constant, τ_d. The finite

pool of vesicles leads to synaptic saturation, because when the presynaptic neuron fires more rapidly than vesicles are replaced, no extra excitatory transmission is possible. Such synaptic depression contributes to stabilizing persistent activity at relatively low rates.

We assume that there is at most one vesicle release per spike; hence the release probability at any individual synapse, $P_R(t)$, is

$$P_R(t) = 1 - [1 - p_v(t)]^{N(t)} \quad (7)$$

where $p_v(t)$ is the release probability for an individual vesicle and $N(t)$ is the number of docked vesicles (smaller than a maximum N_0). We make the simplification that there are many synapses between each pair of connected neurons, such that the average release probability per synapse, $\langle P_R(t)\rangle$, simply scales the amplitude of synaptic transmission, as shown in *Eq.* 3. Similarly, we do not keep track of a discrete $N(t)$ for every individual synapse, but assume that the several synapses between two neurons have a binomial distribution with an average number of docked vesicles, $\langle N(t)\rangle$ (where the brackets $\langle\rangle$ represent the average over this binomial distribution, with mean $\langle N(t)\rangle$) and maximum N_0. Hence $\langle N(t)\rangle$ is a continuous variable obeying

$$\frac{d\langle N\rangle}{dt} = \frac{N_0 - \langle N\rangle}{\tau_d} - \langle P_R(t)\rangle\delta(t - t_{spike}) \quad (8)$$

decreasing by $\langle P_R(t)\rangle$ after a spike at time t_{spike}. By averaging over the binomial distribution, we have

$$\langle P_R(t)\rangle = \langle 1 - [1 - p_v(t)]^{N(t)}\rangle \quad (9)$$

$$= 1 - [1 - p_v(t) \times \langle N(t)\rangle/N_0]^{N_0} \quad (10)$$

and this value of $\langle P_R(t)\rangle$ is used in *Eq.* 3.

The vesicular release probability is given by the product of three gating variables, $p_v(t) = O_1(t)O_2(t)O_3(t)$. A gating variable $O_i(t)$ (i = 1,2,3) increases because of calcium influx triggered by an action potential, followed by a decay with time constant τ_f^i between spikes. Specifically, the following simple update rule is used: a gating variable $O_i(t)$ (i = 1,2,3) follows

$$O_i(n+1) = 1 - \{1 - O_i(n)\exp[-(t_{n+1} - t_n)/\tau_f^i]\}C_f^i \quad (11)$$

Our simulations use the following values for the parameters in the continuous network: N_0 = 16, τ_d = 0.5 s, C_f^1 = 0.45, τ_f^1 = 50 ms, C_f^2 = 0.75, τ_f^2 = 200 ms, C_f^3 = 0.9, τ_f^3 = 2 s. Excitatory neurons in the discrete attractor network are identical except with τ_d = 0.1 s.

Network connectivity

Connection strengths between neurons depend only on their group numbers and are all-to-all. All weights are normalized by (i.e., divided by) the number of neurons in the presynaptic group, so that average network properties should be independent of the system size.

The set of weights W_{EI}, W_{IE}, W_{II} all follow the same form

$$W_{i\to j}^{EI} = W_{max}^{EI}\exp\left(\frac{-|i-j|}{N_{grps}\sigma_{EI}}\right) \quad (12)$$

where N_{grps} = 12 is the total number of groups used, W_{max}^{EI} is the maximum connection strength between groups of the same label ($i = j$), and σ_{EI} determines the breadth of connections to other groups. W_{max}^{IE}, σ_{IE}, W_{max}^{II}, and σ_{II} have similar definitions.

The recurrent excitation has a slightly different form. First, the connections within the same group are significantly stronger than those between groups, so we define a separate set of parameters for the $W_{i\to i}^{EE} = W_i$. Second, the connection strengths between different groups i and j are asymmetric

$$W_{i\to j}^{EE} = W_0^{EE}\exp\left(\frac{-|i-j|}{N_{grps}\sigma_i A_{EE}}\right) \quad (13)$$

for $i > j$ and

$$W_{i\to j}^{EE} = W_0^{EE}\exp\left(\frac{-A_{EE}|i-j|}{N_{grps}\sigma_i}\right) \quad (14)$$

if $i < j$, where A_{EE} is an asymmetry factor. The continuous attractor network has A_{EE} = 1.5 so that connections are stronger from higher to lower threshold neurons.

The cross-inhibition, W_{cross}^{EI}, is the strength of connection from each inhibitory population, labeled by i to the excitatory population of opposite tuning labeled by 14 − i for $2 \le i \le 12$.

The full set of parameters are as follows for the continuous network:

W_0^{EE} = 0.16, W_1 = 0.244, W_2 = 0.239, W_3 = 0.237, W_4 = 0.238, W_5 = 0.239, W_6 = 0.24, W_7 = 0.241, W_8 = 0.242, W_9 = 0.243, W_{10} = 0.244, W_{11} = 0.245, W_{12} = 0.246; σ_1 = 0.5, σ_2 = 0.4, σ_3 = 0.39, σ_4 = 0.385, σ_5 = 0.385, σ_6 = 0.388, σ_7 = 0.392, σ_8 = 0.397, σ_9 = 0.402, σ_{10} = 0.408, σ_{11} = 0.414, σ_{12} = 0.42; A_{EE} = 1.5; W_{max}^{EI} = 1.65, σ_{EI} = 0.25, W_{max}^{IE} = 0.5, σ_{IE} = 0.2, W_{max}^{II} = 2.0, σ_{II} = 0.5. W_{cross}^{EI} = 0.25.

For the discrete network: W_0^{EE} = 0.14, W_1 = 0.35, W_2 = 0.365, W_3 = 0.378, W_4 = 0.39, W_5 = 0.401, W_6 = 0.412, W_7 = 0.423, W_8 = 0.434, W_9 = 0.445, W_{10} = 0.455, W_{11} = 0.465, W_{12} = 0.475; σ_1 = 10, σ_2 = 10, σ_3 = 10, σ_4 = 10, σ_5 = 10, σ_6 = 10, σ_7 = 10, σ_8 = 10, σ_9 = 10, σ_{10} = 10, σ_{11} = 10, σ_{12} = 10; A_{EE} = 1.0; W_{max}^{EI} = 0.3, σ_{EI} = 0.4, W_{max}^{IE} = 0.3, σ_{IE} = 0.4, W_{max}^{II} = 0.5, σ_{II} = 0.5; W_{cross}^{EI} = 0.25.

The readout cells of the discrete network are excited by cells from the populations, i, with weights W_i^{ER} as follows: W_1^{ER} = 0.45, W_2^{ER} = 0.4, W_3^{ER} = 0.35, W_4^{ER} = 0.4, W_5^{ER} = 0.25, W_6^{ER} = 0.2, W_7^{ER} = 0.2, W_8^{ER} = 0.2, W_9^{ER} = 0.2, W_{10}^{ER} = 0.2, W_{11}^{ER} = 0.2, W_{12}^{ER} = 0.2.

Covariance functions

In the following sections, a key quantity that enters the calculations and affects all the statistics is the covariance in spike rates, $cov_{ij}^{(r)}$ (t_1, t_2) between neurons i and j (which can refer to the same neuron with $i = j$) at times t_1 and t_2. The special case $cov_{ii}^{(r)}$ (t_1, t_2) is the variance in spike rates across trials for a neuron i at time t_1. When analyzing data, we count the number of spikes by neuron, i, in trial, n, in specific bins of width δt (where δt = 25 ms unless otherwise stated) between $t = (k - 1/2)\delta t$ and $t = (k + 1/2)\delta t$ as $N_i^n(k)$. Hence the relevant covariance in spike count becomes

$$\mathrm{cov}_{ij}^{(N)}(k\delta t, k\delta t + \tau) = \frac{1}{N_{trials}}\sum_n [N_i^n(k) - \bar{N}_i(k)][N_i^n(k + \tau/\delta t) - \bar{N}(k + \tau/\delta t)]$$

$$= \frac{1}{N_{trials}}\sum_n [N_i^n(k)N_j^n(k + \tau/\delta t)] - \left[\frac{1}{N_{trials}}\sum_n N_i^n(k)\right]\left[\frac{1}{N_{trials}}\sum_{n'} N_j^{n'}(k + \tau/\delta t)\right] \quad (15)$$

where we have substituted for the trial-averaged count, $\bar{N}_i(k) = (1/N_{trials})\sum_n N_i^n(k)$, explicitly in the final line. We achieve the average across different trials, n, in our computer simulations by using different initializations of the random number generator.

Correlation functions and correlograms

We calculate the unnormalized cross-correlogram by averaging the covariance function over the measurement interval according to the formula (Bair et al. 2001; Brody 1998, 1999)

$$C_{ij}(\tau) = \frac{1}{T - \tau}\sum_{k=0}^{(T-\tau)/\delta t}\mathrm{cov}_{ij}^{(N)}(k\delta t, k\delta t + \tau) \quad (16)$$

The time lag is τ and the total temporal window of data measurement

is T. Both T and t are integral multiples of $\delta\tau$. The integration window for detection of spikes offset by τ is limited to $T - \tau$. The above definition is for positive t. For negative τ, the summation limits are from $k = -\tau/\delta\tau$ to $T/\delta\tau$ and prefactor is $1/(T + \tau)$ such that $C_{ij}(-\tau) = C_{ji}(\tau)$.

We normalize the covariance functions by dividing by the geometric mean of the variance at the two time-points (Brody 1999). This leads to the correlation coefficient, $r_{ij}(t,t + \tau)$, between spike times t and $t + \tau$ of neurons i and j

$$\rho_{ij}(t,t + \tau) = \frac{\mathrm{cov}_{ij}(t,t + \tau)}{\sqrt{\mathrm{cov}_{ii}(t,t)\mathrm{cov}_{jj}(t + \tau,t + \tau)}} \quad (17)$$

We average over time, t

$$\overline{\rho_{ij}(\tau)} = \frac{1}{T - \tau}\int_0^{T-\tau} \rho_{ij}(t,t + \tau)\mathrm{d}t \quad (18)$$

to produce the correlograms in Fig. 3, *A* and *C*.

Figure 4*B* is calculated using a variant of *Eq.* 16 with $T - \tau$ replaced by T' to remove any artificial dependence on τ because of its inclusion in the integration limit. Different values of T' produce the four different curves in Fig. 4*B*.

Power spectrum

The appropriate power spectrum for a nonstationary process is the Wigner-Ville spectrum (Flandrin 1989), which is well established in signal processing and physics (Mallat 1999). It is defined as a function of frequency, ω, and time, t, as

$$W(t,\omega) = \int_{-\infty}^{\infty} \mathrm{cov}(t + \tau/2,t - \tau/2)\exp(-i\omega\tau)\mathrm{d}\tau \quad (19)$$

where $\mathrm{cov}(t_1,t_2)$ is the covariance function (between 2 times, t_1 and t_2) described in an earlier subsection. The integration over τ, the time-difference ($\tau = t_2 - t_1$) leads to a result independent of t, the mean time [$t = (t_1 = t_2)/2$] for a stationary process.

For a nonstationary process, we can calculate a power spectrum in frequency alone, $P(\omega)$, independent of time, t, by averaging the Wigner-Ville spectrum across the measurement period (Flandrin 1989) as

$$P(\omega) = \frac{1}{T}\int_0^T W(t,\omega)\mathrm{d}t \quad (20)$$

In practice, to obtain good statistics from noisy spike trains, such a temporal average is essential.

Calculation of power spectrum from spike trains

In Fig. 4*C* we plot the temporally averaged Wigner-Ville spectrum, $P(\omega)$, for neurons in our simulations, using a measurement interval of $T = 10$ s. Specifically, we calculated the average power spectrum by combining spikes from different neurons in a population and using *Eqs. 15*, *19,* and *20* to give

$$P(\omega_n) = \frac{1}{N_{pop}^2}\frac{1}{N_{trial}}\sum_{\lambda=1}^{N_{trial}}\sum_{i,j=1}^{N_{pop}}\sum_{k,k'=1}^{T/(\delta t)} N_i^{\lambda}(k)N_j^{\lambda}(k')\cos[\omega_n\delta t(k - k')] - \sum_{k,k'=1}^{T/(\delta t)}\frac{1}{N_{pop}^2}\frac{1}{N_{trial}^2}\Big[\sum_{\lambda=1}^{N_{trial}}\sum_{i=1}^{N_{pop}} N_i^{\lambda}(k)\Big]\Big[\sum_{\lambda'=1}^{N_{trial}}\sum_{j=1}^{N_{pop}} N_j^{\lambda'}(k')\Big]\cos[\omega_n\delta t(k - k')] \quad (21)$$

where $\omega_n = (n\tau)/T$, and the sums are over all binned spikes [where k is the time index for the bin that contains spikes at times t such that $k - (\delta t)/2 < t \leq k + (\delta t)/2$)] from $N_{pop} = 25$ different neurons from the same population (with identical network inputs, but different background noise) in trial-λ out of N_{trials}. In the log-log plot of Fig. 4*C*, we show values for low-n up until statistical noise causes the power to be negative for some data points (in which case we cannot take the logarithm) and estimate the power coefficient, α, from a straight line fit of log-power versus log-frequency through the points shown.

Fano factor

The Fano factor, $F(T)$, is a measurement of trial-to-trial variability of the spike count of an individual neuron. It is obtained from the variance in spike count during a measurement window of temporal length, T, and is expressed as a function of time

$$F(T) = \frac{\sum_n [N_{i,n}(T) - \langle N_i(T)\rangle]^2}{\langle N_i(T)\rangle} \quad (22)$$

Noise correlation

We evaluate the noise correlation by counting the total number of spikes during a delay of 6 s after the cue offset, for a given neuron, i, in the nth trial, as $N_{i,n}$. The noise correlation, X_{ij}, between two neurons, i and j, is defined as (Lee et al. 1998)

$$X_{ij} = \frac{\sum_n (N_{i,n}N_{j,n} - \langle N_i\rangle\langle N_j\rangle)}{\sqrt{\sum_n (N_{i,n}^2 - \langle N_i\rangle^2)}\sqrt{\sum_n (N_{j,n}^2 - \langle N_j\rangle^2)}} \quad (23)$$

where $\langle N_i\rangle = (\Sigma_n N_{i,n})/N_{trial}$ is the trial-averaged number of spikes for the ith neuron. Clearly $X_{ii} = 1$, the maximum possible correlation when $i = j$, and in general, $-1 \leq X_{ij} \leq 1$. The noise correlation is calculated for each cue and averaged across cues in Fig. 5, *A* and *B*.

The total spike count covariance, plotted in Fig. 5*C*, is simply the numerator in the preceding function, that is

$$\sum_n (N_{i,n}N_{j,n} - \langle N_i\rangle\langle N_j\rangle).$$

RESULTS

We examined spike-time correlations in a recurrent network model of noisy spiking neurons for somatosensory parametric working memory (Miller et al. 2003). The model was designed to simulate prefrontal cortical neurons in monkeys during a somatosensory delayed discrimination task (Romo et al. 1999). In this task, the monkeys were trained to discriminate the frequencies of two vibrotactile stimuli, presented before and after a delay period of 3–6 s, so that the behavioral response (a lever press to indicate whether the 1st or 2nd stimulus was at a higher frequency) depended on remembering the first stimulus frequency across the delay period. It was discovered that neurons in prefrontal and premotor cortices exhibited persistent delay activity that was tuned monotonically to the initial frequency. Such neurons maintain the necessary information for the animal to perform the delayed comparison task.

The model network contains 12,000 neurons that fire spontaneously and stochastically (with a CV averaging >1 at firing rates <10 Hz, dropping to 0.5 as firing rates increase to 40 Hz) because of noisy excitatory inputs. The neurons are arranged in two sets of 12 neuronal populations. In line with the experimental data (Romo et al. 1999), the activity of one set of populations increases with the stimulus frequency, whereas the activity of the other set decreases with increasing stimulus frequency. Hence the tuning of populations is positively monotonic or negatively monotonic, respectively. Recurrent excitation dominates within each set of populations, whereas the two

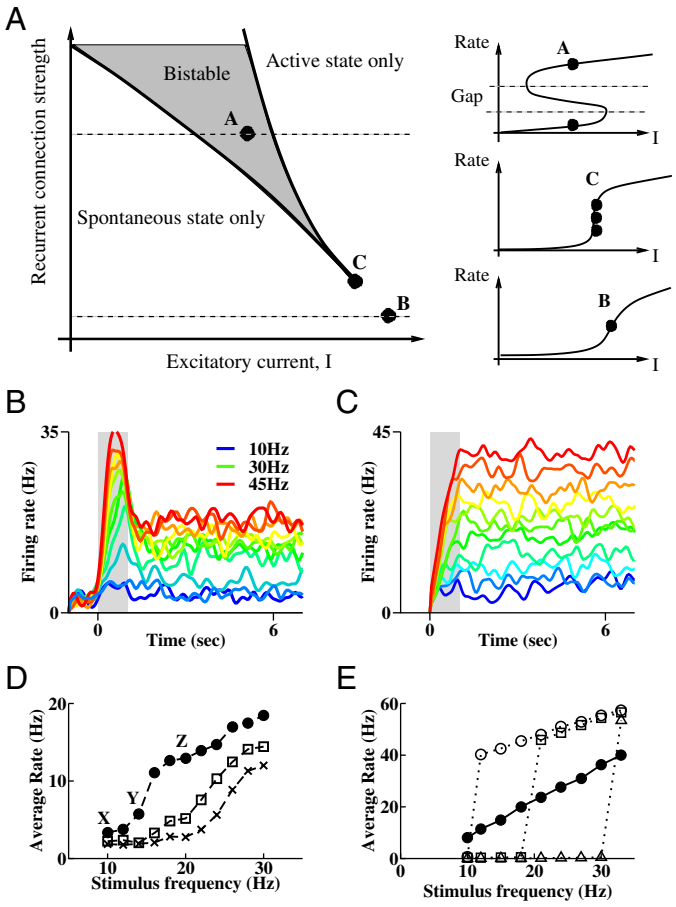

FIG. 2. Continuous or discrete attractor networks. *A*: schematic phase diagram showing possible stable states for a single population as a function of excitatory drive (*x*-axis) and strength of intrinsic excitatory feedback (*y*-axis). Boundary of a bistable region, where 2 stable states exist with a gap between them (*A*) is given by a cusp (*C*). For recurrent excitation weaker than at the cusp, the system varies continuously from low to high activity (*B*). For recurrent excitation greater than at the cusp, bistability exists in a range of external drive. *B*: trial-averaged neural firing rate for a positively monotonic excitatory cell in the network with quasi-continuous states. Solid bar represents cue duration of 1 s. Colors from blue through green to red represent increasing stimulus frequency. *C*: trial-averaged neural firing rate for a positively monotonic excitatory readout cell in the network with discrete states. Same representation as *B*. *D*: tuning curves showing average firing rate during the delay for 3 neurons from different populations in the quasi-continuous model. The points X, Y, and Z mark the stimuli used in Fig. 3*B*. *E*: tuning curves with average firing rate in the delay for 3 neurons from different bistable populations in the network with discrete states (open symbols) and readout population (filled symbols). Only the readout cells are designed to have realistic properties that can be compared with experimental data.

oppositely tuned sets are connected by cross-inhibition (see Fig. 1*B*). The results comparing the mnemonic activity of our model network with the delay activity of monkey neurons are published elsewhere (Miller et al. 2003).

In our model, after a transient stimulus, neurons are able to fire persistently, with slow drift, over a range of firing rates. The feedback to each population is tuned (schematically to position C in Fig. 2*A*) so that each population is on the cusp of bistability, enabling the system to be close to possessing a continuous attractor (Aksay et al. 2000; Durstewitz 2003; Loewenstein and Sompolinsky 2003; Miller et al. 2003; Seung 1996; Seung et al. 2000b). The delay activity after different strengths of stimulus is shown in Fig. 2, *B* and *D*. Neurons are able to fire persistently over a wide range of rates. Initial stimuli of different intensities cause a transient increase in firing rates, before the network settles into a state of approximately constant firing rate, but with fluctuations caused by noise. We label the system in this case as quasi-continuous.

Alternatively, we can adjust parameters so that each population provides strong excitatory feedback to itself and receives less excitation from other sources (position A in Fig. 2*A*). In such a case, the population can be strongly bistable, often with a large gap in firing rates between a down state and an up state. If groups of neurons are bistable, but have differing excitabilities or thresholds, the network can possess a number of robust, discrete stable states (Goldman et al. 2003; Koulakov et al. 2002), with the number of states approximately equal to the number of bistable groups. Figure 2, *C* and *E*, shows the persistent activity for such a multistable network. The possible firing rates of an individual neuron have a large gap, between states where the rate is almost zero, and states where the firing rate is high (Fig. 2*E*, open symbols). Such a large gap in firing rates results from the strong recurrent feedback within a population, so that once the neurons in a population are able to fire a little, they excite each other with strong synaptic input, causing increasing activity in an escalating manner. Synaptic depression and receptor saturation limit the rise in firing rate to a persistent level of 40–50 Hz in our model. Any transition between the multiple discrete states in the network involves a discrete jump in the rate of at least one population of neurons, from their spontaneous value to their lowest persistent rate (across the gap in Fig. 2*A*). Such a large jump in firing rates of neurons in a population between stable states makes the discrete system more stable to noise fluctuations (Koulakov et al. 2002).

To produce more realistic cells in the discrete attractor network, to be compared with those in the continuous attractor network, we added a readout population to the discrete network. The readout population receives excitatory input that is a weighted sum from all populations of the same sign of tuning in the network. The tuning curve of the readout population (Fig. 2*E*, solid symbols) more closely resembles those of neurons in the continuous network and experimental data (Romo et al. 1999). Cells from the readout population also have a similar, rate-dependent CV (although about 10% lower) to neurons in the continuous attractor network. In all statistical analyses, we use such cells from the readout population of the discrete network.

Hence using the same framework of interconnected populations of neurons, we compared the noise and correlations of these two types of memory system. Our main results are independent of details of the model. However, they rely on noise moving the system along a continuous range of states but being unable to cause large jumps in activity between discrete states. A set of discrete attractors with small gaps in rate between states, or with large noise fluctuations, so that noise could cause spontaneous transitions between the discrete states, would show the same behavior we describe here for a continuous attractor (Miller 2006).

Random walk of rate

If a continuous attractor underlies our parametric working memory model, fluctuations in the state of the network are

described by a random walk. To show this conceptual point, consider a simplified, linear model of a typical cell in a neural group whose firing rate $r(t)$ obeys the following dynamic equation

$$\frac{dr}{dt} = -\frac{r - GI}{\tau_s} + \sqrt{A}\eta(t) \quad (24)$$

where GI is the mean firing rate produced by the input current I, $\eta(t)$ is an uncorrelated noise, $\sqrt{A}$ is the noise amplitude, and τ_s is a synaptic or membrane time constant (a few to tens of milliseconds). Noise leads to an autocorrelation which decays exponentially

$$C_{ii}(\tau) \propto \exp(-\tau/\tau_s).$$

Parametric working memory requires neurons to display sustained changes of firing activity by feedback mechanisms either within a cell (Camperi and Wang 1998; Egorov et al. 2002; Goldman et al. 2003; Loewenstein and Sompolinsky 2003) or through a reverberatory network (Koulakov et al. 2002; Miller et al. 2003; Seung 1996; Seung et al. 2000b). In either case, a positive feedback implies a component of the input current that increases with its own firing rate, so making the simplification of linear feedback, $I = I_0 + Wr$, the above equation becomes

$$\frac{dr}{dt} = -\frac{r - GWr - GI_0}{\tau_s} + \sqrt{A}\eta(t) \simeq -\frac{r}{\tau_{eff}(W)} + \frac{GI_0}{\tau_s} + \sqrt{A}\eta(t) \quad (25)$$

where $\tau_{eff}(W) = \tau_s/(1 - GW) > \tau_s$. If there is no feedback, $W = 0$, and in time τ_s, the firing rate reaches its steady-state value, GI_0, as in *Eq. 24*. With strong feedback ($GW \rightarrow 1$), one has $\tau_{eff} \gg \tau_s$. In the limit $\tau_{eff} \rightarrow \infty$, the first term in the right hand side of the equation vanishes. The firing rate integrates the input I as well as noise over time, in the sense of calculus: $r(t) = (GI_0/\tau_s)t + A\int^t \eta(t')dt'$. Moreover, after the input is withdrawn, the firing rate is maintained stationary at any level (within an operating range) except for noise-induced drifts; hence the system behaves like a line attractor. The average rate is given by $< r > = GI_0\, t_I/\tau_s$, where t_I is the time when the input is withdrawn, whereas the trial-to-trial variance in rate increases linearly with time: $\langle (r - \langle r \rangle)^2 \rangle = At$.

The analogy between the random walk model and our parametric working memory model lies in the fact that the continuum of persistent activity states form a one-dimensional curved line in the space of population firing rates, as schematically shown in Fig. 1*A*. If the firing rates are temporarily perturbed by noise to move off this line, the network's dynamics cause the firing rates to return back to the line, but do not prevent any drifts along the continuous attractor. Such noise-induced drifts along the attractor are similar to a random walk. To assess whether the simple random walk model can quantitatively capture the behavior of our large-scale and highly nonlinear network model, we analyze the fluctuation properties of population firing rates and spike counts.

Random walk analysis of a line attractor

A closer analysis of the effects of noise on a system with a continuous attractor reveals that the correlation functions are determined by the products of gradients of the tuning curves of two neurons (Ben-Yishai et al. 1995; Pouget et al. 1998). The argument is shown in Fig. 1*A*, where the top three figures depict the tuning curves (average firing rate as a function of stimulus) for three neurons. The first two are positively monotonic; the third is negatively monotonic. When the system has a continuous attractor, the only fluctuations that are not rapidly damped out are those along the attractor. The attractor can be seen as a curve in a plot of the firing rate of one neuron versus another, with each point in the curve corresponding to the two firing rates of persistent activity after a particular stimulus. Noise has the same effect as the stimulus in shifting the firing rates along the attractor. Two positively monotonic neurons [with firing rates r_1,r_2 and tuning curves $f_1(s)$,$f_2(s)$] either both increase or decrease their firing rates together, after a larger or smaller stimulus, and after noise fluctuations. Hence their noise correlations are expected to be positive.

If we consider the firing rate of mnemonic persistent activity as a function of stimulus, s, and noise, $\eta(t)$, expanded to first order in the noise, we have

$$r_1 = f_1[s + \eta(t)] \simeq f_1(s) + \eta(t)df_1/ds \quad (26)$$

$$r_2 = f_2[s + \eta(t)] \simeq f_2(s) + \eta(t)df_2/ds \quad (27)$$

The noise correlation and unnormalized cross-correlation functions are proportional to the covariance of firing rates, $\langle r_1(t)r_2(t')\rangle - \langle r_1(t)\rangle\langle r_2(t')\rangle$, which is given by the product of the noise terms in *Eqs. 26* and *27*. Hence the cross-correlation is proportional to the product of gradients of the two tuning curves [$(df_1/ds).(df_2/ds)$] (Ben-Yishai et al. 1995; Pouget et al. 1998). If the gradients of tuning curves of two neurons have opposite sign [as in Fig. 1*A*; $(df_1/ds)(df_3/ds) < 0$], the noise correlation is negative between the corresponding neurons. That is, a fluctuation along the attractor simultaneously causes the positively monotonic neurons to increase their firing rates and negatively monotonic neurons to decrease their rates or vice versa. Similarly, for an individual neuron, the covariance function is proportional to $(df_i/ds)^2$, so all effects of noise are greater after cues that fall on a steep part of a neuron's tuning curve.

Noise during the stimulus presentation leads to variation in the initial encoded values of s and has a correlated effect on the firing rate of neurons in the same way as noise during the delay. The main effect of both stimulus noise and drift noise in a line attractor is encoded in the one-dimensional variable, s. The firing rate of each neuron responds to the network noise by an amount proportional to the square of the gradient of its tuning curve. In the preceding analysis, we assumed that fluctuations are small compared with changes in the gradient of the tuning curve, so df_i/ds remains constant for each neuron for a complete trial, allowing *Eqs. 26* and *27* to be expanded only to first order. If the first-order approximation is valid, the proportionality constant, A, is the same for both stimulus noise and drift noise. That allows us to write the total noise at time t as

$$\sigma_i^2(t) = A_i(t + t_0) \quad (28)$$

for any neuron, i, where t_0 quantifies noise during the stimulus and is a constant for all neurons that form the line attractor and $A_i \propto (df_i/ds)^2$.

Other results for a random walk are described in detail in the APPENDIX. We summarize the key features that can be compared with our network results here. The covariance function,

$\mathrm{cov}(t_1,t_2)$, for a random walk starting at $t = 0$ is independent of the time difference, $|t_2 - t_1|$, and instead equals the variance of the process given above in *Eq. 28* where t is the earlier of t_1 or t_2 (Gillespie 1992) (cf. *Eqs. A4* and *A5*).

The other statistical measures presented in this paper are derived from the covariance or variance functions. The time-average of the covariance function is proportional to the integration limits used in the time-averaging (see *Eqs. A6–A9*). The average (*Eq. 16*), leads to a linear decay in $|\tau|$

$$C_{xx}(\tau) = A(T + 2t_0 - |\tau|)/2 \quad (29)$$

for the unnormalized correlation function of a random walk, which will be compared with Fig. 3*A–C*. An alternative method results in curves independent of τ, but linearly increasing with T'

$$C'_{xx}(\tau) = A(T' + 2t_0)/2 \quad (30)$$

for the unnormalized correlation function of a random walk, which will be compared with Fig. 4*B*.

A hallmark of random walk behavior is an inverse-square power law for the Wigner-Ville power spectrum (*Eq.* 19) (Flandrin 1989). When averaged over a finite time interval, T, specific frequencies, $\omega_n = n\pi/T$ should be used, so that the inverse-square power law is apparent in the averaged spectrum. We find that initial variance in the random walk leads to oscillations in the power spectrum between even and odd frequencies (*Eqs. A11* and *A12*), such that

$$P(\omega_n) = 2A/\omega_n^2 \quad (31)$$

for even n and

$$P(\omega_n) = 2A(1 + 2t_0/T)/\omega_n^2 \quad (32)$$

for odd n.

The Fano factor, $F(T)$, for a random walk, also shows power-law behavior, increasing quadratically with time. Initial variance in the starting points contributes a linear term, such that for a Poisson-like renewal process

$$F(T) = 1 + \frac{A(T^2 + 3t_0T)}{3\bar{r}_0} \quad (33)$$

In the following subsections, we present the results of our spiking network and compare them with those expected for a random walk in a line attractor.

Correlation function

Fluctuations are commonly quantified by autocorrelation functions of single neurons and cross-correlation functions between them (Bair et al. 2001; Brody 1999; Perkel et al. 1967a,b; Singer and Gray 1995). The autocorrelation function is a function of two times (1 for each spike) and should only be written as a function of the time-difference, or time-lag, τ, if the underlying process is stationary (Gardiner 1985). Many processes measured in neuroscience are not stationary, so the shuffle-correction is used to remove the principle effects of systematic variations in the average rate. We use the term correlogram to denote the shuffle-corrected covariance function, normalized by the product of SD in spike count (producing a correlation coefficient), averaged across the measurement interval (Brody 1999). We use the term unnormalized correlogram when referring to a temporal average of the covariance function.

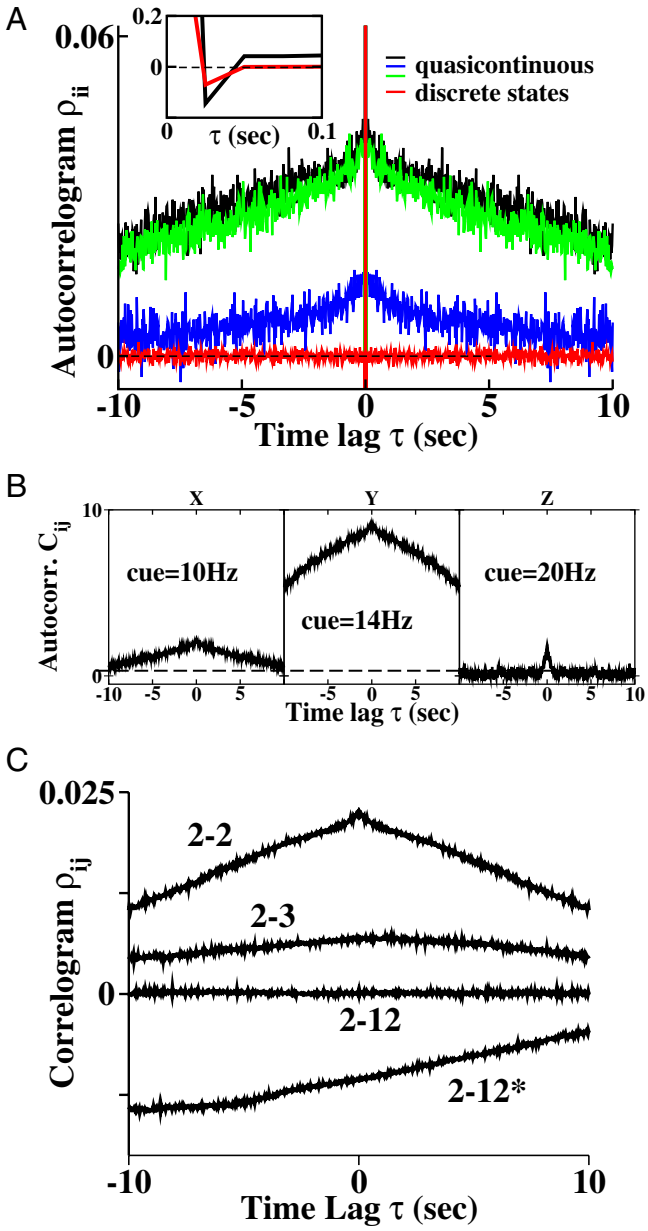

FIG. 3. Long-time correlograms. *A*: autocorrelogram measured during delay period for 3 populations of the quasi-continuous network (black, blue, and green) and 1 from the discrete network (red). Central dip is caused by the refractory period. Note that the neurons in the network with quasi-continuous states have a non-zero correlation for large time-lags, decaying slowly, approximately linearly with time. *Inset*: autocorrelogram reaches 0 after a few tens of milliseconds for a neuron in the discrete network. *B*: size of offset in the unnormalized autocorrelogram is cue-dependent and is greatest at the steepest parts of the tuning curve for any particular neuron. Stimuli correspond to those marked X, Y, and Z in Fig. 2*D*. *C*: cross-correlograms between neurons of the same population (2–2), two closely tuned positively monotonic populations (2–3), a low-threshold and high-threshold population (2–12) and between a positively and negatively tuned population (2–12*). The lack of time-symmetry between oppositely tuned populations-2 and 12* is a sign that in this case, the firing rate of the negatively tuned population (12*) drifts for several seconds before affecting the firing rate of the positively tuned population (2).

Trial-to-trial differences in the underlying rate have been shown to give rise to artifacts in cross-correlograms at short

time scales (Brody 1998, 1999). In contrast, we report that the correlation functions are affected on long time scales by the linear increase in variance of firing rate as a function of time that arises from the temporal integration of noise in a quasi-continuous attractor.

The autocorrelograms for three cells in our graded memory network are shown in black, green, and blue in Fig. 3*A*. Strikingly, temporal correlations between spikes persist for many seconds and seem to show a gradual, linear decrease with the time lag. In contrast, a network with robust, discrete states does not show any long-term correlations (Fig. 3*A*, red curve). In fact, the autocorrelograms for all neurons fall to zero during a time scale on the order of the synaptic time constant (100 ms for NMDA receptors in this work; Fig. 3*A*, *inset*).

The autocorrelogram in Fig. 3*A* was calculated by averaging over memory activity after different stimuli. If we compute the correlations separately after each stimulus, we find that only a small subset of stimuli result in such large, long-term persistence in the correlations. This is because, if the stimulus lies on a relatively flat part of a neuron's tuning curve, small network fluctuations do not cause significant changes in the firing rate of that neuron. Tuning curves of neurons in our network are not linear in general, but resemble sigmoidal functions, as seen for the majority of neurons in the experimental data (Miller et al. 2003; Romo et al. 1999). It is the steepest part of the tuning curve that is most sensitive to fluctuations. If the tuning curve of a neuron is linear (unlike those seen in our model), the effect of fluctuations during the delay should be independent of stimulus.

In Fig. 3*B*, we calculated the unnormalized autocorrelogram separately for delay activity after three stimuli, corresponding to the positions X, Y, and Z in Fig. 2*D*. To obtain better statistics, we averaged cross-correlations between 25 neurons in a single population. Because the population fires asynchronously and all neurons within one population receive the same network input, their autocorrelation functions are essentially the same. The average cross-correlation between all pairs of neurons in the same population is identical to the average autocorrelation of those neurons, except at very short time scales (less than ≈25 ms). It is only the δ-function peak at $\tau = 0$ and the dip in the autocorrelation at short τ (because of the refractory period and recovery from reset; Fig. 3*A*) that distinguishes autocorrelations from cross-correlations between cells within a population of our network.

Figure 3*C* shows that the cross-correlogram is greater between neurons with similar tuning curves (such as 2 neurons in the same population, 2–2, or neighboring population, 2–3) than between neurons with very different thresholds (2–12). The cross-correlogram is negative between neurons of opposite tuning (2–12*) and is temporally asymmetric. The asymmetry is a sign that firing rates of one population can drift for several seconds before the change in rate is strong enough to affect the populations with opposite sign of tuning. This is because cross-inhibition is not the dominant feature in our network and suggests that, on a time scale of a few seconds, our system resembles more closely two weakly coupled continuous attractors (where drift in 1 attractor provides input to the other) than a single set of stable points.

In all cases, autocorrelograms and cross-correlograms do not decay exponentially as $\exp(-t/\tau)$ with a characteristic time constant, τ (see the text after *Eq. 24*), but decay more linearly, as is typical for a random walk.

Increasing variance of rate

We found that the variance of single-trial population firing rates (calculated with 100 trials) increases approximately linearly from the stimulus offset for up to 10 s of the mnemonic period (Fig. 4*A*). The nonzero intercept at $t = 0$ reflects trial-to-trial noise in the system's response to the stimulus.

The linear time course with nonzero intercept for the firing rate variance is consistent with a random walk (see *Eqs. A4* and *A5* and *Noise during the stimulus*), with an initial distribution of starting points caused by variability of neuronal response during the stimulus presentation (*Eq. A2*). Note that the variance of neurons in the discrete system remains constant and low (Fig. 4*A*, red curve).

Covariance independent of time lag

Such a random walk also provides an explanation for the observed near-linear behavior of the correlograms reported in Fig. 3. Because random walks are nonstationary processes (the variance increases with time), correlation functions depend separately on the two times of measurement (t_1,$t_1 + \tau$) and not just on their difference, τ (see METHODS). Hence when evaluating correlations as a function of the time lag, τ, by integrating over t_1, the precise measurement interval affects the result. It can be shown mathematically that a linear increase in the variance of rates as a function of time leads to a linear increase in the unnormalized correlogram as a function of the integration interval (Miller 2006; Saleh 1978). In the standard formulation, the integration interval is $(T - |\tau|)$, where τ is the time-lag and T is the total time of measurement. With fixed T, this leads to a linear decrease with τ, as seen in Fig. 3*B*.

In Fig. 4*B*, we tested whether the unnormalized autocorrelogram is truly a function of the measurement interval using an alternative integration window to calculate the temporal average (the sum over k in *Eq. 16*). We calculate the unnormalized autocorrelogram by including spikes over a range of $T' + \tau$ for each value of τ so that the integral over t_1 is over a fixed value, $T' = T - \tau_{max}$, rather than $T - \tau$. Hence the measured correlation interval is independent of τ and proportional to T'. We plotted curves for the average unnormalized autocorrelogram of 100 neurons (25 neurons in each of 4 populations) using increasing values of T' (4, 8, 12, and 16 s for the ascending curves). For a perfect random walk, the curves in Fig. 4*B* would be constant with τ and equally spaced along the y-axis (see METHODS). The curves do have a y-offset that increases monotonically with T', whereas they vary little with τ, so the network's activity is behaving qualitatively like a random walk during the delay.

Power law of power spectra

Furthermore, for a random walk, the power spectrum, $P(\omega_n)$, of the spike train should include a delta-function at the origin and at low frequencies an inverse-square law decay, such that $P(\omega_n) \propto \omega_n^{-\alpha}$ where a equals 2 (Flandrin 1989) and $\omega_n = n\pi/T$. A pure power law decay should appear as a straight line on a log-log plot, with the slope of the line giving the exponent (Fig. 4*C*). We estimated the exponent, α, for several cells by fitting

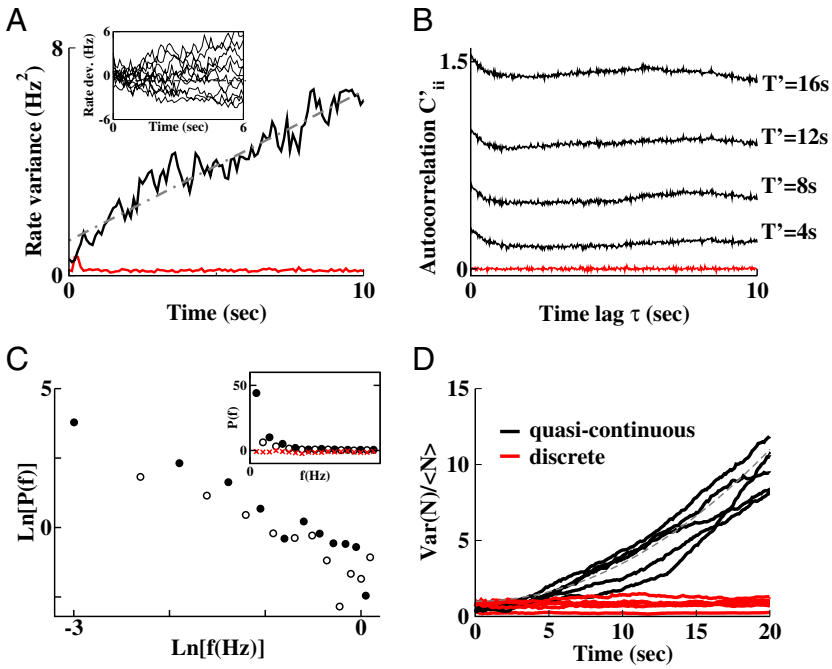

FIG. 4. Random walk behavior after the 14-Hz cue. *A*: trial-to-trial variance of average firing rate in population-2 of the continuous attractor network increases approximately linearly with time (black), but with an offset (at $t = 0$) indicating significant trial-to-trial variation in the responses to the stimulus by the end of the cue. Mean rate for the population following the cue varies between 8 and 11 Hz. Readout population from the discrete attractor shows a low, constant variance (red). *Inset*: each thin line represents the population-average firing rate (with trial-averaged mean rate subtracted) for an individual trial after the end of the stimulus (10 trials are shown of 50 used to calculate the variance for the population in the continuous attractor network). *B*: dependence of the unnormalized autocorrelogram on the measurement interval, T'. We vary the integration window for spikes as a function of τ to maintain a constant range, T', for the first spike in the pair. Hence the window for a pair of spikes is equal to $T' + \tau$. The 4 curves in ascending order are for values of $T' = 4$, 8, 12, and 16 s such that the autocorrelation increases monotonically with T'. For a perfect random walk, curves would be independent of τ and proportional to T'. Results are average from 100 neurons (25 in each of 4 populations) over 50 trials for a single cue (black) and from 25 cells in the readout population of the discrete network (red). *C*: log-log plot of the power spectrum for cells from population-2 of the continuous attractor network, shows decay at low frequencies close to a power-law (close to linear on the log-log plot). Filled circles for frequencies, $f = \omega_n/2\pi$, with odd n, and open circles for even n ($\omega_n = n\pi/T$ with $T = 10$ s). Data are calculated from the spike trains of 25 cells of each population to produce the figure. *Inset*: power spectrum showing low-frequency behavior for cells from 1 population (black circles), with the readout population of the discrete attractor as comparison (red crosses). *D*: Fano factor as a function of time. Dark lines, for neurons in the quasi-continuous network the variance in spike count normalized by the mean spike count (Fano factor) increases with time during the delay after cues that give a large offset in the autocorrelogram; red, neurons in the discrete network have approximately constant Fano factors over the same time interval, with smaller values for cues leading to higher firing rates and more regular interspike intervals. Analysis predicts that a random-walk of the rate leads to a Fano factor that increases quadratically with time (dashed gray line).

the linear region of log-log power spectra. We fitted curves separately through odd and even data points, because these were offset due to noise during the stimulus (see *Eq. A12* and *Noise during the stimulus*). The fits yielded values ranging from $\alpha = 1.6$ to $\alpha = 2.1$ ($\alpha = 1.77$ and 1.66 for the odd and even frequencies, respectively, of population-2, shown in Fig. 4*C*). The fact that α deviates from 2 indicates that our network is not a perfect continuous attractor. Internal dynamics (such as synaptic depression) mean that steady states are not constant, and a lack of perfect tuning means systematic network drift occurs on a slow time scale given by the synaptic time constant divided by the fractional mistuning (Seung et al. 2000a) (see *Eq. 25*). Hence infinitely precise tuning is necessary to generate a perfect continuous attractor (with no systematic drift), whereas the quasi-continuous attractor of our network is sufficient for short-term memory.

Power law of Fano factors

We also considered the variability across trials of spike count in a specific time interval, measured by the Fano factor. It is calculated as the variance in number of spikes divided by the mean spike count. A large Fano factor means large variation in the neuronal activity. Temporal correlations in the spike times give rise to temporal changes of the Fano factor. Fano factors typically vary between 0 (for regular spiking) and 1 (for Poisson processes) when no temporal correlations arise from variations in the firing rate (de Ruyter van Steveninck et al. 1997). The system with discrete attractors (Fig. 2, *C* and *E*) does indeed have a low value of the Fano factor that is constant over time (Fig. 4*D*, red). The Fano factors are near unity for neurons at low firing rates and have values closer to zero for more rapidly firing neurons (in our model, a dominance by synaptic excitation implies that spike trains at a higher firing rate are more regular). In sharp contrast, neurons in the system with a quasi-continuous attractor exhibit Fano factors that increase with time as a result of the internal dynamics of the network. The increase is consistent with calculations for a random walk process, which produces a Fano factor that increases quadratically with time (Miller 2006; Saleh 1978) (see APPENDIX and *Eq. A14*).

In summary, these different characteristics are consistent with each other, showing that our network approximates a continuous attractor and exhibits approximate random walk behavior.

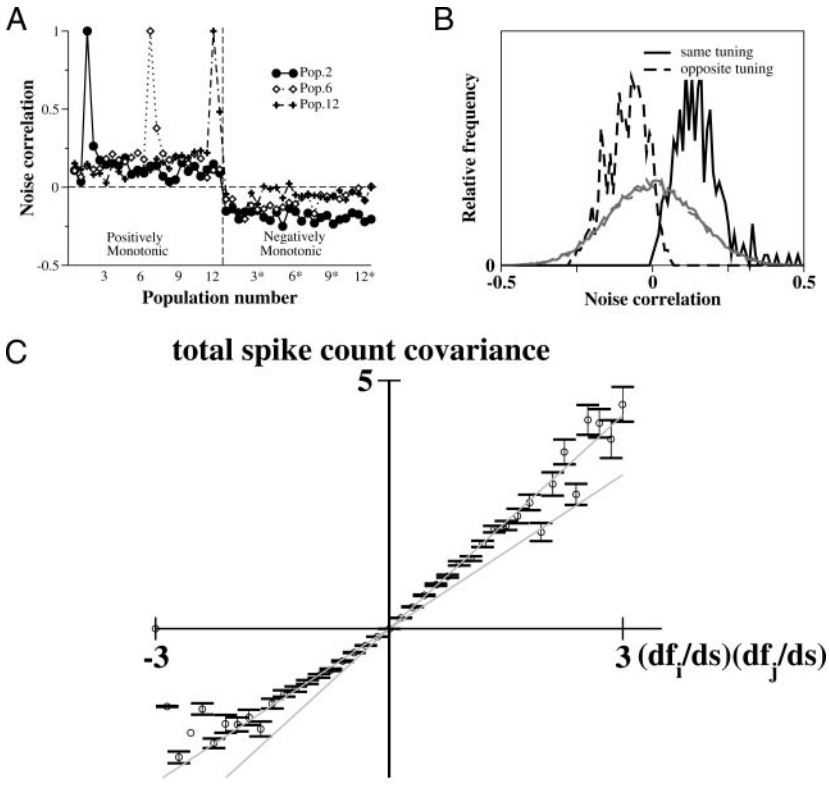

FIG. 5. Noise correlation of persistent activity. *A*: normalized correlations in the spike counts for delay activity after all cues. Correlations of 3 positively monotonic excitatory neurons with 2 neurons from each population are shown. Note that correlations are greatest between cells of the same population (points neighboring the same-cell peak of magnitude 1) or between cells in populations with similar tuning curves, and are negative between oppositely tuned cells. *B*: solid or dashed black curves: distributions of noise correlations in the continuous attractor network between cells of which the 2 tuning curves have the same signs of slope (predominantly positive correlations) or the opposite signs of slope (negative correlations), respectively. Gray: 2 distributions of noise correlations for cells in the discrete attractor network, with means of +0.001 and −0.001 that are barely distinguishable. *C*: covariance in total spike count of pairs of neurons, as a function of the product of the slopes of their tuning curves. Data are binned along the *x*-axis, with mean and error bars of each bin plotted. Data are taken during the delay period, for each cue, and between all pairs of neurons with 2 representatives from each population. We fitted a tanh function to the tuning curve of average rate as a function of stimulus and used the gradient of this curve at each stimulus as d*f*/d*s*. Theory predicts that the 2 quantities are proportional to each other, such that the points would fall about a straight line through the origin. Steeper gray line is a fit through data in the positive quadrant (for neurons with the same sign of tuning curve). Notably, the 2nd gray line, which fits the data in the negative quadrant, is less steep.

Noise during the stimulus

A strong test of a random walk in a line attractor is to fit the variance as a linear function, $A(t + t_0)$, where the value of t_0, representing noise during the stimulus, is the same across all cells (see analysis preceding *Eq. 28*). Therefore to check for consistency, we fit values of t_0 using *Eqs. 28–33* to describe the statistical measures presented in Fig. 4, following a cue of frequency 14 Hz. For example, we fitted curves separately through the odd (open symbols) and even (filled symbols) data points of the power spectra shown in Fig. 4*C*. While both curves should decay with the same power law of 2 for a random walk, initial variance in firing rates at the beginning of the integration window should lead to odd data points (odd n for $\omega_n = n\pi/T$), having a higher amplitude than for even data points. A difference between odd n and even n arises from the cosine contribution in *Eq. A12*. In the standard result, ignoring any initial offset variance (Flandrin 1989), the only oscillations appear from the sine term, which gives zero contribution at all values of ω_n. In contrast, the cosine contribution gives a contribution proportional to $+t_0/T$ for odd n and $-t_0/T$ for even n. Hence we can use this difference in amplitudes to check whether our estimates of t_0 are consistent with estimates from our other statistical measures. In this case, we find that calculations of t_0 range from $t_0 = 0.4$ s to $t_0 = 4.2$ s for different cells.

We found a similarly wide variation in estimates of t_0 from calculations of the variance (Fig. 4*A*), the autocorrelations (Fig. 4*B*), and the Fano factors (Fig. 4*D*). Such a wide variation means that estimates of t_0 do not provide strong evidence for a continuous attractor. However, for population-2, the estimates of t_0 after a 14-Hz cue, using the methods of Fig. 4, *A–D*, were 2.3, 2.7, 3.4, and 2.4 s, respectively. These values are quite comparable and show consistency across different calculations.

Noise correlation

The noise correlation between pairs of neurons, plotted in Fig. 5*A*, is the integral over all time lag of their shuffle-corrected unnormalized cross-correlograms, divided by the geometric mean of the areas under the two shuffle-corrected unnormalized autocorrelograms (Bair et al. 2001; Brody 1999). We selected two neurons from each excitatory population and calculated noise correlations between all pairs of neurons that we selected. In the quasi-continuous attractor network, the largest noise correlations occur between different neurons of the same population (except for the trivial case of noise correlation equal to unity for the same neuron), whereas the noise correlation is typically negative between oppositely tuned neurons (Fig. 5*A*). To quantitatively assess the latter observation, we calculated the histogram of noise correlation for two separate categories of pairs of neuron: pairs of neurons with the same sign of tuning (both positively monotonic or both negatively monotonic) versus pairs of neurons with the opposite sign of tuning (one positively monotonic and the other negatively monotonic). We plotted the distributions of noise correlation for the two categories in Fig. 5*B*. It is clear that noise correlations are positive for neurons with the same sign of tuning and are negative for neurons with the opposite signs of tuning.

In the discrete attractor network, we also expect the sign of noise correlation between two neurons to depend on the relative signs of tuning, because neurons with similar tuning are coupled by excitation, whereas neurons with opposite tuning are coupled by inhibition. We did find slight, but significant, evidence for such a sign-dependence of the noise correlation when measuring across neurons from all populations of the discrete network. Neurons with the same sign of tuning had an average noise correlation of +0.001 (average of 25,000 pairs),

and neurons with opposite tuning had an average of −0.001 (average of 20,000 pairs). The difference is significant, but such a small difference in midpoints of the two histograms (Fig. 5*B*, gray) is masked by the much larger magnitude of the SD (~0.15). Neurons within the same population of the discrete network had a larger noise correlation, averaging 0.007 (cf. with 0.3 for the continuous network).

The analysis after *Eqs. 26* and *27* suggests that, not only should the sign of noise correlation depend on the relative sign of tuning of two neurons, but the covariance in total spike counts should be proportional to the product of gradients of tuning curves (Ben-Yishai et al. 1995; Pouget et al. 1998). To test this prediction, we calculated the covariance in total spike counts between neurons during memory states after different cues and plotted this as a function of the product of gradients of their tuning curves. We binned the results according to the value along the *x*-axis (product of tuning-curve gradients) and plotted the mean with SE in the *y*-axis (total spike-count covariance) in Fig. 5*C*. We noticed that a single linear fit did not match the data well, but rather two separate linear fits through the origin were needed for the two quadrants of data. The linear fit in the negative *x*-*y* quadrant has a smaller slope than the fit through the positive *x*-*y* quadrant. The lower gradient indicates that the magnitude of noise correlation between neurons of opposite tuning is lower (i.e., the correlation is less negative) than expected, given the noise correlations between neurons with the same sign of tuning. This result can be accounted for by the fact that the connections between neurons with the same sign of tuning are stronger than those cross-connections to populations with the opposite sign of tuning (Fig. 1*B*). Hence populations with the same sign of tuning (aligned vertically in Fig. 1*B*) are closely entrained with each other, but the weaker connections across the network (horizontal connections in Fig. 1*B*) allow more independence between fluctuations in the firing rates of oppositely tuned populations. Such covariance in neuronal activity limits the possibility of obtaining more information from noisy neurons by pooling them together (Abbott and Dayan 1999).

Discrete attractor network

In all the results presented for the discrete attractor network, we used stimulus strengths that reliably cause the system to attain one specific stable state. However, if intermediate stimulus strengths are used, trial-to-trial fluctuations during the stimulus can leave the system in different states on different trials. In such an event, the system of discrete attractors possesses long-term noise correlations, which are similar to the continuous attractor network, but with specific, significant differences.

In contrast to the continuous attractor network, whose variance increases with time, the trial-to-trial variance in firing rates is initially large and remains constant with time in the system of discrete attractors. All other results stem from this and resemble equivalent effects that arise from noise during the stimulus (represented by the quantity t_0) in the continuous network. The discrete system has a memory of noise during the stimulus, which lasts for the duration of the trial (in the absence of transitions between discrete states). The amplitudes of these long-term noise correlations in a discrete attractor network are stimulus-dependent, but in an additional way to that shown in Figs. 1*A*, 3*B*, and 5*C* for the continuous attractor. When the stimulus strength is smoothly changed, the network responds from being in one state with near 100% certainty, to 50% one state and 50% the next state, to 100% in the next state and so on through all discrete states (assuming the states are separated far enough compared with noise fluctuations). The variance is minimal when the system is reliably in one state and peaks when it is equally likely to be in two states. Hence the variance and magnitudes of long-term correlation effects all oscillate as a function of cue strength.

To observe these effects in a discrete attractor network, it is most likely that an animal should be trained with a discrete set of stimuli, and the neural responses compared using intermediate stimuli. The large long-term noise correlations at intermediate stimuli would increase the psychometric thresholds around those stimulus values for any behavior relying on a discrete memory network.

DISCUSSION

Our results show that if a network of neurons contains a continuous attractor that acts as a memory store, it will produce correlations that persist on the time scale of any memory held in the network. This is because noise fluctuations along the attractor are not damped, but instead are integrated in time to produce an activity pattern resembling a random walk. The power-law behavior presented here should be a salient characteristic of any system with a continuum of states. For example, in a model of spatial working memory (Camperi and Wang 1998; Compte et al. 2000), the network dynamics maintain the form of a bell-shaped persistent activity pattern but do not prevent random drift of the position of the peak in activity around the network. Hence the location of the peak wanders with time as a random walk, with a variance that increases linearly with time (Compte et al. 2000). Such behavior can lead to autocorrelation and cross-correlation functions that do not decay exponentially with time (Ben-Yishai et al. 1995; Pouget et al. 1998).

Our realization of a continuous attractor depends on recurrent excitatory feedback through synapses within the network. However, recurrent excitation could originate from intrinsic membrane dynamics within a single cell, leading to a range of stable single-cell firing rates (Loewenstein and Sompolinsky 2003). As long as noise is able to change those stable rates, similar power-law behavior should be observable in the spike times of such neurons. The results rely on the ability of noise to change the activity of the network between stable values. A network with a set of discrete stable states, in the presence of strong noise that can change the system from one state to another does possess the same correlations as a continuous attractor (Miller 2006). However, such a network lacks the characteristic robustness of a discrete system. On the other hand, if the noise is weaker than the barrier between different discrete states, the network activity is robust to noise fluctuations. Such robust networks do not show power-law behavior.

Sensory neurons in the visual system can also exhibit unusual fluctuations in their spike times (Baddeley et al. 1997; Teich et al. 1997). In these cases, the power spectra show power-law behavior, proportional to $1/\omega^{\alpha}$ with α between one and two, corresponding to fractional Brownian motion (Mandelbrot and Ness 1968). Similarly, the variance in spike count

increases as a power law, supralinearly with time, to produce Fano factors that increase with time as a power law. Teich suggests the power-law behavior observed in retinal cells under constant illumination (Teich et al. 1997) is the result of internal optimization in the circuitry to encode natural stimuli characterized by scaling statistics. Such power-law statistics within presented natural scenes might directly explain power-law behavior observed in the visual cortex (Baddeley et al. 1997). In contrast to these feedforward networks, our system with a continuous attractor is dominated by its recurrent feedback. It acts as a memory system, maintaining persistent activity after a transient input has withdrawn. It is the behavior of the system in the absence of inputs that leads to the power-law correlations studied in this paper.

Long-term spike correlations give rise to significant noise correlations between neurons in a recurrent network. Such correlations have been seen in neurons associated with the memory of spatial location (Constantinidis and Goldman-Rakic 2002) and with the memory of a vibrational frequency (Machens et al. 2005). Moreover, if two neurons with opposite gradients of their tuning curves are simultaneously measured, one expects to see negative cross-correlations at long time delays and negative noise correlations. Neurons with similar tuning are observed to possess more positive noise correlations than those with dissimilar or opposite tuning curves in spatial working memory tasks (Constantinidis and Goldman-Rakic 2002) and spatial motor response tasks (Lee et al. 1998). Negative noise correlations have been observed between neurons with dissimilar tuning in the spatial motor response task (Lee et al. 1998) as well as between positively tuned and negatively tuned neurons in the vibrational memory task (Machens et al. 2005).

In our analysis of the system as a continuous attractor state, positive correlations between oppositely tuned neurons can only occur if fluctuations in the rate of the whole attractor state occur more often than fluctuations along the attractor. For example, trial-to-trial variations in the concentrations of modulators could cause regional covariances of firing rates to be positive and swamp any negative correlations arising from noise in the network. However, because significant changes in the baseline properties of a large number of neurons would destabilize a continuous attractor, it is difficult to reconcile such a model with experimental data. The cross-inhibition between oppositely tuned neurons in our system means that a random drift in the firing rate of one network causes the firing rate to drift in the opposite direction for the oppositely tuned neurons.

To retrieve a signal encoded in a network with one set of positively monotonic neurons and one set of negatively monotonic neurons, it is necessary to take the difference between the activities of the two sets. This leads to an improved signal-to-noise ratio if oppositely tuned neurons have positively correlated noise, because when taking the difference, any positively correlated noise would cancel (Abbott and Dayan 1999; Romo et al. 2003). This would require oppositely tuned neurons to be either uncoupled (and subject to similar neuromodulation) or to be coupled by excitation. Hence the cross-inhibition suggested by some data (Machens et al. 2005), which we used to help stabilize the mnemonic activity in the network, is not optimal for decoding.

When the dominant fluctuations in a system are along an attractor, the fluctuations in spike counts of different neurons that form the attractor will necessarily be correlated. The diffusion of the activity pattern along the attractor sets a limit on the amount of information available about the original stimulus. Trial-to-trial variability in the firing rate of a neuron that arises from random drift of the network's activity results in a loss of information about the original stimulus, which cannot be regained by pooling together more neurons from the same network subject to the same random drifts in activity (Shadlen and Newsome 1996). In our network, we find that increasing the number of neurons in each population improves the stability of the encoded quantity and reduces the total noise. Because the background noise input remains unchanged, the CV of individual neurons is not greatly reduced, but the noise correlation between cells does decline significantly with increasing network size. Our network has all-to-all connectivity, so the number of synapses per neuron scales with system size, and we make a compensatory reduction in synaptic conductance. Hence the noise caused by recurrent feedback received by a single neuron scales inversely with system size in our model. This decreases the amount of fluctuation along the attractor, reducing the loss of information caused by random drift, as the number of neurons in the system increases.

Experimental observation of a variance in responses that increases approximately linearly with delay would be strong evidence of a random walk process in a memory system. Such behavior has been seen in spatial working memory tasks (White et al. 1994) and is suggested in some monotonically graded memory tasks, such as memory of visual contrast or vernier (see Fig. 1 in Pasternak and Greenlee 2005).

Such a systematic change in the trial-to-trial variance in spikes per bin as a function of time during the task means that the underlying process is nonstationary. Most standard statistical tools assume a stationary process, yet most neuronal spike trains are not stationary, so care is needed with the analysis (Brody 1998). We used the Wigner-Ville power spectrum (Flandrin 1989) (see *Eq. 19*), which in signal theory is an established generalization of the power spectrum to nonstationary processes (Mallat 1999). Calculation of autocorrelations as a function of a single time lag variable, τ, is strictly only valid for stationary processes with constant variance. However, because measurement of two-point correlations (in t_1 and t_2) is experimentally unfeasible with noisy spike trains, some kind of temporal average is required. To reduce the effect of changing variance, we normalized the spike-count covariance functions by the product of SD at different time-points, thus generating temporal correlation coefficients (Papoulis 1984), before averaging them across time (Brody 1999). This method leads to a result bounded between 1 and -1, and it is the most appropriate method for analyzing temporal correlations in a nonstationary process.

In summary, we observed long-term correlations in a network of spiking neurons, which was designed to exhibit a continuous range of persistent activity. The network is sensitive to noise fluctuations, and we contrast its behavior with a more robust network with discrete states where noise cannot change the activity of the network from one discrete state to another. In the discrete network, noise after stimulus offset is rapidly damped and leads to no observable correlations, but in the continuous network, the memory of noise is on the same

time scale as the memory of the stimulus. Memory of noise is a hallmark of random walk behavior, which is defined as the temporal integral of noise. We showed that many properties of our network can be interpreted as approximate random walk behavior, namely linearly increasing variance of firing rates, linearly decaying unnormalized correlograms, power spectra with a power-law decay of exponent close to two, and Fano factors that increase with time. Such behavior, when observed in experimental data, can be considered as evidence for a quasi-continuous attractor.

APPENDIX

Analysis: covariance function for a point renewal process with the rate undergoing a random walk.

For analysis of the effects of random variations in the underlying rate, we assume spikes are emitted randomly, with the probability of a spike in a time from t to $t + \delta t$ being $r(t)\delta t$, where $r(t)$ is the underlying rate at that instance. If the rate function follows a random walk, the spike train of each neuron is a doubly stochastic Poisson point process (Saleh 1978).

The spike covariance $\mathrm{cov}(t_1,t_2)\,\delta t_1\delta t_2$ is given by the probability of spikes in both time intervals t_1 to $t_1 + \delta t$ and t_2 to $t_2 + \delta t$ minus the product of the separate probabilities of a spike in each interval. Hence

$$\mathrm{cov}(t_1,t_2)\delta t_1\delta t_2 = \int_0^\infty \mathrm{d}r_1 P(r_1,t_1) r_1 \delta t_1 \int_0^\infty \mathrm{d}r_2 P(r_2,t_2|r_1t_1) r_2 \delta t_2 - \left[\int_0^\infty \mathrm{d}r_1 P(r_1,t_1) r_1 \delta t_1\right]\left[\int_0^\infty \mathrm{d}r_2 P(r_2,t_2) r_2 \delta t_2\right] \quad (A1)$$

where $P(r_2,t_2|r_1,t_1)$ is the probability density for the underlying rate of r_2 at time t_2 given its value of r_1 at time t_1. For a random walk of the firing rate, beginning at a rate, r_0 at $t = 0$, where r_0 is given by a Gaussian distribution of mean, $\bar{r}_0$, and variance, σ_0^2

$$P(r,t) = \frac{1}{\sqrt{2\pi(At + \sigma_0^2)}} \exp\left[\frac{-(r - r_0)^2}{2(At + \sigma_0^2)}\right] = \frac{1}{\sqrt{2\pi[A(t + t_0)]}} \exp\left\{\frac{-(r - r_0)^2}{2[A(t + t_0)]}\right\} \quad (A2)$$

and

$$P(r_2,t_2|r_1,t_1) = \frac{1}{\sqrt{2\pi A(t_2 - t_1)}} \exp\left[\frac{-(r_2 - r_1)^2}{2A(t_2 - t_1)}\right] \quad (A3)$$

where A is the square of the noise amplitude and we have written $\sigma_0^2 = At_0$ in *Eq. A2* to quantify the initial noise during the stimulus (see Gillespie 1992; Miller 2006). The variance of firing rates increases linearly with time as $A(t + t_0)$. *Equations A2* and *A3* are valid for times such that $At \ll r_0$, because the firing rate cannot be negative.

Substituting *Eqs. A2* and *A3* into *Eq. A1* and solving the integrals leads to

$$\mathrm{cov}(t_1,t_2) = At_1 + \sigma_0^2 = A(t_1 + t_0) \quad (A4)$$

where $0 \le t_1 \le t_2$. Similarly

$$\mathrm{cov}(t_1,t_2) = At_2 + \sigma_0^2 = A(t_2 + t_0) \quad (A5)$$

if $0 \le t_2 \le t_1$. We only consider times greater than zero for the present analysis, assuming no spikes occur before $t = 0$, and hence $\mathrm{cov}(t_1,t_2) = 0$ for $t_1 < 0$ or $t_2 < 0$.

Autocorrelation function for a random walk

The standard calculation of the autocorrelation function for a time lag of t includes all spikes in a total measurement interval, T, assuming no spikes occur before $t = 0$ and after $t = T$. Hence the covariance function is zero for $t_1 > T$ or $t_2 > T$, and the unnormalized autocorrelation function is calculated as

$$C_{xx}(\tau) = \frac{1}{(T - \tau)} \int_0^{T-\tau} \mathrm{d}t_1 \mathrm{cov}(t_1,t_1 + \tau) \quad (A6)$$

where the above form is written for $\tau > 0$ so that when the time of the first spike, t_1, ranges from 0 to $T - t$, the time of the second spike, t_2, ranges from τ to T, and the complete measurement interval from 0 to T is covered by the spike pair. That is, no contribution occurs to the integral for $t_1 > T - \tau$, because in such a case, $t_2 > T$ and the covariance function is zero. Importantly, the measurement interval decreases linearly with τ in the preceding formulation. If we use the result for a random walk of the firing rate, replacing $\mathrm{cov}(t_1,t_2)$ with $A(t_1 + t_0)$ in the preceding equation, we obtain the result

$$C_{xx}(\tau) = A(T + 2t_0 - \tau)/2 \quad (A7)$$

producing a linear decrease with τ. This explains the near linear decay with τ for the time-averaged unnormalized correlation coefficients in Fig. 3*B*.

We used an alternative form for the cross-correlation function when calculating the results presented in Fig. 4*B* to emphasize that, for a pure random walk, the spike covariance is independent of the time lag, τ, but increases with the measurement time. Hence we use values of a measurement interval, T', that are smaller than the total measurement interval of the spike train but are independent of τ. If we have data for spikes from a time of $t = 0$ up to the time $T' + \tau_{max}$, where τ_{max} is the maximum value of time lag used, we can calculate an alternative, unnormalized autocorrelation function as

$$C'_{xx}(\tau) = \frac{1}{T'} \int_0^{T'} \mathrm{d}t_1 \mathrm{cov}(t_1,t_1 + \tau) \quad (A8)$$

which yields

$$C'_{xx}(\tau) = A(T' + 2t_0)/2 \quad (A9)$$

increasing linearly with T' but independent of τ (see Fig. 4*B*).

Power spectrum for a random walk

For a random walk with finite duration, T, and initial variance caused by noise during the stimulus, At_0 (see *Eqs. A4* and *A5*), the covariance function, $\mathrm{cov}(t + \tau/2, t - \tau/2)$, is proportional to $t + t_0 - |\tau|/2$ for $0 \le t + \tau/2 \le T$ and $0 \le t - \tau/2 \le T$ and is zero otherwise. This yields for $t < T/2$

$$W(t,\omega) = \frac{A}{\omega^2}[1 - \cos(2\omega t) + 2\omega t_0 \sin(2\omega t)] \quad (A10)$$

and for $t > T/2$

$$W(t,\omega) = \frac{A}{\omega^2}\{1 - \cos[2\omega(T - t)] + 2\omega(2t - T + t_0)\sin[2\omega(T - t)]\} \quad (A11)$$

Averaging this function over t between 0 and T yields

$$P(\omega) = \frac{2A}{\omega^2}\left\{1 - \frac{\sin(\omega T)}{\omega T} + \frac{t_0}{T}[1 - \cos(\omega T)]\right\} \quad (A12)$$

where the random walk of firing rate is defined with noise amplitude $\sqrt{A}$ such that the variance increases linearly with time as $A(t + t_0)$.

Equation A12 simplifies to $2A/(\omega_{2m}^2)$ for $\omega_{2m} = 2m\pi/T$ and $2A(1 + 2t_0/T)/(\omega_{2m+1}^2)$ for $\omega_{2m+1} = (2m + 1)\pi/T$.

Hence a hallmark of random-walk behavior is a power-law decay of the power spectrum, with an exponent close to two. Any initial trial-to-trial variance, At_0, in the starting point of the random walk leads to an additional oscillating contribution in the power spectrum, so that the curve with odd-integer $n = 2m + 1$ is higher than the curve with even-integer $n = 2m$.

Fano factor for a random walk

If spikes are emitted in a Poisson manner, with probability $r(t)\delta t$ in time t to $t + \delta t$, where the rate varies in time and from trial to trial, but maintains an average $\bar{r}_0$, the Fano factor is given by (Saleh 1978; Miller 2006)

$$F(T) = 1 + \frac{2\int_0^T dt' \int_0^{t'} dt'' Var[r(t'')]}{\bar{r}_0 T} \quad (A13)$$

For a random walk with an initial spread of starting points such that the variance is $\mathrm{Var}[r(t)] = A(t + t_0)$, *Eq. A13* yields

$$F(T) = 1 + \frac{A(T^2 + 3t_0T)}{3\bar{r}_0} \quad (A14)$$

Hence another indication of random walk behavior is a power-law increase in the Fano factor, with any initial variance in the rates contributing a linear term.

ACKNOWLEDGMENTS

The authors thank A. Renart and D. Lee for helpful comments on the manuscript.

GRANTS

This work was supported by National Institute of Mental Health Grants K25MH-064497 to P. Miller and NIMH-062349 to X.-J. Wang.

REFERENCES

Abbott LF and Dayan. The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11: 91–101, 1999.

Aksay E, Baker R, Seung HS, and Tank DW. Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *J Neurophysiol* 84: 1035–1049, 2000.

Averbeck BB and Lee D. Coding and transmission of information by neural ensembles. *Trends Neurosci* 27: 225–230, 2004.

Baddeley R, Abbott LF, Booth MCA, Sengpiel F, Freeman T, Wakeman EA, and Rolls ET. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc R Soc Lond B Biol Sci* 264: 1775–1783, 1997.

Bair W, Zohary E, and Newsome WT. Correlated firing in macaque visual area mt: time scales and relationship to behavior. *J Neurosci* 21: 1676–1697, 2001.

Ben-Yishai R, Lev Bar-Or R, and Sompolinsky H. Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci USA* 92: 3844–3848, 1995.

Brody CD. Slow covariations in neuronal resting potentials can lead to artefactually fast cross-correlations in their spike trains. *J Neurophysiol* 80: 3345–3351, 1998.

Brody CD. Correlations without synchrony. *Neural Comput* 11: 1537–1551, 1999.

Brody CD, Hernández A, Zainos A, Lemus L, and Romo R. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb Cortex* 13: 1196–1207, 2003.

Buzsaki G. Large-scale recording of neuronal ensembles. *Nat Neurosci* 7: 446–451, 2004.

Camperi M and Wang X-J. A model of visuospatial short-term memory in prefrontal cortex: recurrent network and cellular bistability. *J Comput Neurosci* 5: 383–405, 1998.

Compte A, Brunel N, Goldman-Rakic PS, and Wang X-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 10: 910–923, 2000.

Compte A, Constantinidis C, Tegnér J, Raghavachari S, Chafee MV, Goldman-Rakic PS, and Wang X-J. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J Neurophysiol* 90: 3441–3454, 2003.

Constantinidis C and Goldman-Rakic PS. Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J Neurophysiol* 88: 3487–3497, 2002.

de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, and Bialek W. Reproducibility and variability in neural spike trains. *Science* 275: 1805–1808, 1997.

Durstewitz D. Self-organizing neural integrator predicts interval times through climbing activity. *J Neurosci* 23: 5342–5353, 2003.

Egorov AV, Hamam BN, Fransen E, Hasselmo ME, and Alonso AA. Graded persistent activity in entorhinal cortex neurons. *Nature* 420: 173–178, 2002.

Flandrin P. On the spectrum of fractional brownian motions. *IEEE Trans Inf Theor* 35: 197–199, 1989.

Funahashi S, Bruce CJ, and Goldman-Rakic PS. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61: 331–349, 1989.

Gardiner CW. *Handbook of Stochastic Methods* (2nd ed.). Berlin: Springer-Verlag, 1985.

Gawne TJ and Richmond BJ. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13: 2758–2771, 1993.

Gillespie DT. *Markov Processes.* San Diego, CA: Academic Press, 1992.

Goldman MS, Levine JH, Major G, Tank DW, and Seung HS. Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cereb Cortex* 13: 1185–1195, 2003.

Hempel CM, Hartman KH, Wang X-J, Turrigiano GG, and Nelson SB. Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *J Neurophysiol* 83: 3031–3041, 2000.

Koulakov AA, Raghavachari S, Kepecs A, and Lisman JE. Model for a robust neural integrator. *Nat Neurosci* 5: 775–782, 2002.

Lee D, Port NL, Kruse W, and Georgopoulos AP. Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *J Neurosci* 18: 1161–1170, 1998.

Loewenstein Y and Sompolinsky H. Temporal integration by calcium dynamics in a model neuron. *Nat Neurosci* 6: 961–967, 2003.

Machens CK, Romo R, and Brody CD. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* 307: 1121–1124, 2005.

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

Mallat SG. *A Wavelet Tour of Signal Processing* (2nd ed.). San Diego, CA: Academic Press, 1999.

Mandelbrot BB and Ness JW. Fractional brownian motions; fractional noises and applications. *SIAM Rev* 10: 422–437, 1968.

Matveev V and Wang X-J. Implications of all-or-none synaptic transmission and short-term depression beyond vesicle depletion: a computational study. *J Neurosci* 20: 1575–1588, 2000.

Miller P. Analysis of spike statistics in neuronal systems with a continuous attractor or multiple, discrete states. *Neural Comput* 2006.

Miller P, Brody CD, Romo R, and Wang X-J. A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cereb Cortex* 13: 1208–1218, 2003.

Papoulis A. *Probability, Random Variables, and Stochastic Processes,* 2nd ed. New York: McGraw-Hill, 1984.

Pasternak T and Greenlee MW. Working memory in primate sensory systems. *Nat Rev Neurosci* 6: 97–107, 2005.

Perkel DH, Gerstein GL, and Moore GP. Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophys J* 7: 391–418, 1967a.

Perkel DH, Gerstein GL, and Moore GP. Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys J* 7: 419–440, 1967b.

Pouget A, Zhang K, Deneve S, and Latham PE. Statistically efficient estimation using population code. *Neural Comput* 10: 373–401, 1998.

Romo R, Brody CD, Hernández A, and Lemus L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399: 470–474, 1999.

Romo R, Hernández A, Zainos A, and Salinas E. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38: 649–657, 2003.

Saleh, B. *Photoelectron Statistics.* New York: Springer-Verlag, 1978.
Seriès P, Latham PE, and Pouget A. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat Neurosci* 7: 1129–1135, 2004.
Seung HS. How the brain keeps the eyes still. *Proc Natl Acad Sci USA* 93: 13339–13344, 1996.
Seung HS, Lee DD, Reis BY, and Tank DW. The autapse: a simple illustration of short-term analog memory storage by tuned synaptic feedback. *J Comput Neurosci* 9: 171–185, 2000a.
Seung HS, Lee DD, Reis BY, and Tank DW. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26: 259–271, 2000b.
Shadlen MN and Newsome WT. Noise, neural codes and cortical organization. *Curr Opin Neurobiol* 4: 569–579, 1994.
Shadlen MN and Newsome WT. Motion perception: seeing and deciding. *Proc Natl Acad Sci USA* 23: 628–633, 1996.
Shamir M and Sompolinsky H. Nonlinear population codes. *Neural Comput* 16: 1105–1136, 2004.
Singer W and Gray CM. Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18: 555–586, 1995.
Softky WR and Koch C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* 13: 334–350, 1993.
Sompolinsky H, Yoon H, Kang K, and Shamir M. Population coding in neuronal systems with correlated noise. *Phys Rev E* 98: 8095–8100, 2002.
Taube JS and Bassett JP. Persistent neural activity in head direction cells. *Cereb Cortex* 13: 1162–1172, 2003.
Teich MC, Heneghan C, Lowen SB, Ozaki T, and Kaplan E. Fractal nature of the neural spike train in the visual system of the cat. *J Opt Soc Am A* 14: 529–546, 1997.
Tuckwell HC. *Introduction to Theoretical Neurobiology.* Cambridge, MA: Cambridge University Press, 1988.
Varela JA, Sen K, Gibson J, Fost J, Abbott LF, and Nelson SB. A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci* 17: 7926–7940, 1997.
Wang X-J. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J Neurosci* 19: 9587–9603, 1999.
Wang X-J. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24: 455–463, 2001.
White JM, Sparks DL, and Stanford TR. Saccades to remembered target locations: and analysis of systematic and variable errors. *Vision Res* 34: 79–92, 1994.
Zohary E, Shadlen MN, and Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370: 140–143, 1994.