Systems/Circuits

Short-Term Plasticity Explains Irregular Persistent Activity in Working Memory Tasks

David Hansel[1] **and German Mato**[2]

[1]Laboratory of Neurophysics and Physiology and Institute of Neuroscience and Cognition, University Paris Descartes, 75270 Paris Cedex 06, France, and [2]Comisión Nacional de Energía Atómica and Consejo Nacional de Investigaciones Cientificas y Técnicas, Centro Atómico Bariloche and Instituto Balseiro, 8400 San Carlos de Bariloche, Argentina

Persistent activity in cortex is the neural correlate of working memory (WM). In persistent activity, spike trains are highly irregular, even more than in baseline. This seemingly innocuous feature challenges our current understanding of the synaptic mechanisms underlying WM. Here we argue that in WM the prefrontal cortex (PFC) operates in a regime of balanced excitation and inhibition and that the observed temporal irregularity reflects this regime. We show that this requires that nonlinearities underlying the persistent activity are primarily in the neuronal interactions between PFC neurons. We also show that short-term synaptic facilitation can be the physiological substrate of these nonlinearities and that the resulting mechanism of balanced persistent activity is robust, in particular with respect to changes in the connectivity. As an example, we put forward a computational model of the PFC circuit involved in oculomotor delayed response task. The novelty of this model is that recurrent excitatory synapses are facilitating. We demonstrate that this model displays direction-selective persistent activity. We find that, even though the memory eventually degrades because of the heterogeneities, it can be stored for several seconds for plausible network size and connectivity. This model accounts for a large number of experimental findings, such as the findings that have shown that firing is more irregular during the persistent state than during baseline, that the neuronal responses are very diverse, and that the preferred directions during cue and delay periods are strongly correlated but tuning widths are not.

Introduction

Working memory (WM), the ability to temporarily hold, integrate, and process information to produce goal-directed behavior, is crucial to higher cognitive functions, such as planning, reasoning, decision-making, and language comprehension (Baddeley, 1986; Fuster, 2008). The persistent activity recorded in neocortex during WM tasks is thought to be the main neuronal correlate of WM (Fuster and Alexander, 1971; Miyashita and Chang, 1988; Goldman-Rakic, 1995). For example, in an oculomotor-delayed response (ODR) task in which a monkey has to remember the location of a stimulus for several seconds to make a saccade in its direction, a significant fraction of the neurons in the prefrontal cortex (PFC) modify their activity persistently and selectively to the cue direction during the delay period (Funahashi et al., 1989, 1990, 1991; Constantinidis et al., 2001; Takeda and Funahashi, 2007). The classical view is that this reflects a multistability in the dynamics of the PFC circuit because sensory inputs are the same in the precue and in the delay periods but neuronal activity is different (Hebb, 1949; Hopfield, 1984; Amit, 1995; Amit and Brunel, 1997; Wang, 2001).

The persistent activity of PFC neurons is highly irregular temporally (Shinomoto et al., 1999; Compte et al., 2003; Shafi et al., 2007). For instance, it has been reported that in ODR tasks, the coefficient of variation of the interspike interval distribution is close to 1, and even >1 for many of the neurons (Compte et al., 2003). Previous modeling works have attempted to account for this irregularity (Brunel, 2000; van Vreeswijk and Sompolinsky, 2004; Renart et al., 2007; Roudi and Latham, 2007; Barbieri and Brunel, 2008). However, in all these models, unless parameters are tuned, the neurons fire much too regularly in memory states. Hence, accounting in a robust way for highly irregular firing and persistent activity remains challenging.

Irregular firing of cortical neurons is naturally explained if one assumes that cortical networks operate in a regime in which excitation is strong and balanced by strong inhibition (Tsodyks and Sejnowski, 1995; van Vreeswijk and Sompolinsky, 1996, 1998; Vogels et al., 2005; Lerchner et al., 2006; Vogels and Abbott, 2009). It is therefore tempting to assume that in WM tasks the PFC network expresses a multistability between balanced states. We argue here that this requires that the nonlinearities underlying the persistence of activity in PFC are primarily in the neuronal interactions and not in the neurons as assumed previously. We also argue that short-term plasticity is a possible substrate for these nonlinearities. Specifically, we propose that the short-term facilitating excitatory synapses recently reported in PFC (Hempel

Received July 19, 2012; revised Oct. 24, 2012; accepted Oct. 26, 2012.

Author contributions: D.H. and G.M. designed research; D.H. and G.M. performed research; D.H. and G.M. analyzed data; D.H. and G.M. wrote the paper.

This research was conducted within the scope of the France-Israel Laboratory of Neuroscience (FILN-LEA) and supported by UniNet EU excellence network, ANR-09-SYSC-002-01 and Project SECyT-ECOS A04B04. We thank S. Funahashi, R. Guetig, Y. Loewenstein, C. Levenes, G. Mongillo, and C. van Vreeswijk for fruitful discussions. We are also thankful to Y. Loewenstein, G. Mongillo, and C. van Vreeswijk for critical reading of our manuscript. We also thank Ecole de Neuroscience Paris Ile-de-France.

Correspondence should be addressed to David Hansel, University Paris Descartes, Laboratory of Neurophysics and Physiology, 45 Rue des Saints-Pères, 75270 Paris, France. E-mail: david.hansel@univ-paris5.fr.

DOI:10.1523/JNEUROSCI.3455-12.2013

 0270-6474/13/330133-17$15.00/0

et al., 2000; Wang et al., 2006) play an essential role in WM. To illustrate these claims, we develop a new network model for spatial WM based on this assumption. We show that the model displays highly irregular spontaneous activity as well as persistent, selective, and highly irregular delay activity. Importantly, it also displays a great deal of the diversity observed in the delay activity of PFC neurons (Funahashi et al., 1989, 1990; Asaad et al., 1998; Romo et al., 1999; Takeda and Funahashi, 2002; Brody et al., 2003; Shafi et al., 2007).

A brief account of this study has been presented in abstract form (Hansel and Mato, 2008).

Materials and Methods

The network models. In this paper we consider two network models made of spiking neurons. One model (Model I) has an unstructured connectivity. The second model (Model II) represents a network in PFC involved in an ODR task and has a ring architecture (Ben-Yishai et al., 1995; Hansel and Sompolinsky, 1996).

Single neuron dynamics. Both models are made up of an excitatory and an inhibitory population of integrate-and-fire neurons. The membrane potential of a neuron follows the dynamics:

$$\tau \frac{dV}{dt} = -V + I^{\mathrm{rec}}(t) + I^{\mathrm{ext}}(t), \tag{1}$$

where τ is the membrane time constant, $I^{\mathrm{rec}}(t)$ is the total recurrent synaptic current the neuron receives from all other cells in the network connected to it and $I^{\mathrm{ext}}(t)$ represents feedforward inputs from outside the network. Whenever the membrane potential of the neuron reaches the threshold, V_T, it fires an action potential and its voltage is reset to V_R. We take $V_T = 20$ mV, $V_R = -3.33$ mV. The time constants of the neurons are $\tau = 20$ and 10 ms for excitatory and inhibitory neurons respectively, in accordance with standard values (Somers et al., 1995).

The external inputs. Simulation of a memory task (Model I) or of the ODR task (Model II) requires a stimulus that depends on time. First, there is a precue period that allows the system to settle in a baseline state. In the second stage, a transient input is applied to simulate the cue, after which the input returns to the previous value. Finally another transient input is applied to erase the memory that has been stored. Thus, the external feedforward input into neuron $i = 1, \ldots, N_\alpha$ in population $\alpha = E, I$ (hereafter neuron i, α) is:

$$I^{\mathrm{ext}}_{i,\alpha}(t) = I^b_\alpha + I^{\mathrm{cue}}_{i,\alpha}(t) + I^{\mathrm{erase}}_{i,\alpha}(t). \tag{2}$$

The first term on the right hand side represents the background input, which is constant in time and depends solely on the target population. The second term represents the transient sensory inputs related to the cue to be memorized. The third term represents a transient input that erases the memory at the end of the delay period. In Model I, $I^\gamma_{i,\alpha}(t)$ (γ = cue, erase) is homogeneous (i.e., $I^\gamma_{i,\alpha}(t) \equiv A^\gamma_\alpha(t)$ does not depend on i). In model II:

$$I^\gamma_{i,\alpha}(t) = A^\gamma_\alpha(t)[1 + \epsilon^\gamma_\alpha \cos(\theta_{i,\alpha} - \theta_\gamma)], \tag{3}$$

where θ_{cue} is the direction in which the cue is presented; $\theta_{i,\alpha} = \frac{2\pi}{N_\alpha} i$ is the direction of the cue for which the sensory input into neuron (i, α) is maximum; and θ_{erase} is the direction of the saccade that is actually performed. An estimate of θ_{erase} is given by the direction of the population vector computed from the activity of the neurons at the end of the delay period (see below). We assume that the angles, $\theta_{i,\alpha}$, are uniformly distributed between 0 and 360°.

For simplicity we take $A^\gamma_\alpha(t) = \hat{A}^\gamma_\alpha H^\gamma(t)$ (γ = cue, erase), where $\hat{A}^\gamma$ is constant, $H^{\mathrm{cue}}(t) = 1$ during the cue period, and $H^{\mathrm{cue}}(t) = 0$ otherwise. Similarly, $H^{\mathrm{erase}}(t) = 1$ when the memory erasing input is present and $H^{\mathrm{erase}}(t) = 0$ otherwise.

Connectivity of the networks. We define the connectivity matrix of the network by $J_{i\alpha,j\beta} = 1$ if neuron (j, β) is connected presynaptically to neuron (i, α) and $J_{i\alpha,j\beta} = 0$ otherwise. In Model I the connectivity is unstructured. Therefore the probability, $P_{i\alpha,j\beta}$, that $J_{i\alpha,j\beta} = 1$ depends only on α and β: $P_{i\alpha,j\beta} = K_{\alpha\beta}/N_\beta$, where N_β is the number of neurons in population $\beta = E, I$ and $K_{\alpha\beta}$ is the average number of connections a neuron in population α receives from population β.

The architecture of Model II is consistent with the columnar functional anatomy of the monkey PFC (Goldman-Rakic, 1987, 1988, 1995; Rao et al., 1999). The probability of connection between two neurons is given by $P_{i\alpha,j\beta} = P_{\alpha\beta}(|\theta_{i\alpha} - \theta_{j\beta}|)$ with:

$$P_{\alpha\beta}(\theta) = C_{\alpha\beta} \exp\left(\frac{[\theta]^2}{2\sigma^2_{\alpha\beta}}\right), \tag{4}$$

where $[\theta] = \min(|\theta|, 2\pi - |\theta|)$ and $C_{\alpha\beta}$ is a normalization that ensures that the total number of inputs a neuron in population α receives from population β is on average $K_{\alpha\beta}$. The range of the interactions is characterized by the parameters $\sigma_{\alpha\beta}$. The resulting network architecture is a probabilistic version of the architecture of the ring model, in which neurons are all connected with probability 1, whereas the strength of their interactions depends on their distance on the ring (Ben-Yishai et al., 1995; Hansel and Sompolinsky, 1996; Compte et al., 2000).

Synaptic interactions. We model the recurrent synaptic input current into neuron (i, α) as:

$$I^{\mathrm{rec}}_{i,\alpha}(t) = \sum_{j\beta n} J_{i\alpha,j\beta} G_{\alpha\beta} u_{j\beta,n} x_{j\beta,n} f_{\alpha\beta}(t - t_{j\beta,n}), \tag{5}$$

where $G_{\alpha\beta}$ is a constant that measures the maximal synaptic current neuron (i, α) receives from neuron (j, β), $t_{j\beta,n}$ is the time of the n-th spike fired by neuron (j, β), $x_{j\beta,n}$ is the amount of synaptic resources available at its synaptic terminals before this spike, and $u_{j\beta,n}$ is the fraction of these resources used by this spike. The dynamics of these two variables are responsible for the short-term plasticity (STP) and we model them as (Markram et al., 1998):

$$u_{j\beta,n+1} = u_{j\beta,n}\exp\left(-\frac{\Delta t_{j\beta,n}}{\tau_f}\right) + U\left(1 - u_{j\beta,n}\exp\left(-\frac{\Delta t_{j\beta,n}}{\tau_f}\right)\right) \tag{6}$$

$$x_{j\beta,n+1} = x_{j\beta,n}(1 - u_{j\beta,n+1})\exp\left(-\frac{\Delta t_{j\beta,n}}{\tau_r}\right) + 1 - \exp\left(-\frac{\Delta t_{j\beta,n}}{\tau_r}\right), \tag{7}$$

where $\Delta t_{j\beta,n}$ is the n-th interspike interval of neuron (j, β); τ_r and τ_f are the recovery and the facilitation time constants of the synapse, and U the maximal utilization parameter.

Finally, the function $f_{\alpha\beta}(t)$ in Equation 5 describes the dynamics of individual postsynaptic currents (PSCs). It is given by $f_{\alpha\beta}(t) = (1/\tau_{\alpha\beta}) \exp(-t/\tau_{\alpha\beta})H(t)$ where $\tau_{\alpha\beta}$ is the synaptic decay time, and $H(t) = 0$ [respectively (resp.) 1] for $t < 0$ (resp. $t > 0$) (Dayan and Abbott, 2001).

Network size, connectivity, and scaling of network parameters with connectivity. The total number of neurons in the network is $N = N_E + N_I$ with $N_E = 0.8N$ and $N_I = 0.2N$. The average total number of synaptic inputs a neuron receives, K, is assumed to be the same in the two populations. We take: $K_{EE} = K_{IE} \equiv K_E = 0.8K$ and $K_{II} = K_{EI} \equiv K_I = 0.2K$. Unless specified otherwise we take $N = 80{,}000$ and $K = 2000$.

The balanced regime is mathematically well defined only in the limit $N \to \infty$, $K \to \infty$ while $K \ll N$, and the strength of the recurrent interactions and the external inputs are scaled as (van Vreeswijk and Sompolinsky, 1998):

$$G_{\alpha\beta} = \frac{g_{\alpha\beta}}{\sqrt{K_\beta}} \tag{8}$$

$$I^b_\alpha = i^b_\alpha \sqrt{K_E} \tag{9}$$

$$\hat{A}^\gamma_\alpha = \hat{a}^\gamma_\alpha \sqrt{K_E}, \tag{10}$$

where $g_{\alpha\beta}$, i^b_α, and $\hat{a}^\gamma_\alpha$ do not depend on K. This scaling ensures that, in a wide range of parameters, the temporal fluctuations in the synaptic in-

Table 1. Parameters of the synaptic interactions for the unstructured and spatial working memory networks

	Unstructured	Spatial WM
g_{EE} (AMPA)	560 mV ms	533.3 mV ms
g_{EE} (NMDA)	560 mV ms	533.3 mV ms
g_{IE} (AMPA)	67.2 mV ms	67.2 mV ms
g_{IE} (NMDA)	7.4 mV ms	7.4 mV ms
g_{EI}	−138.6 mV ms	−138.6 mV ms
g_{II}	−90.6 mV ms	−90.6 mV ms
U	0.03	0.03
τ_r	200 ms	200 ms
τ_f	450 ms	450 ms
σ_{EE}	—	60°
σ_{IE}	—	70°
σ_{EI}	—	60°
σ_{II}	—	60°

puts remain finite and do not depend on the connectivity when the connectivity is large, whereas the time-average excitatory and inhibitory inputs increase and balance each other (van Vreeswijk and Sompolinsky, 1996, 1998). In that limit, even though the excitatory and inhibitory inputs become infinitely large, the temporal mean and SD of the fluctuations of the total inputs remain finite and on the same order of the neuronal threshold.

For finite connectivity, the balance of excitation and inhibition is only approximate. Therefore, to qualify the dynamic regime of the network as balanced, it is important to check the robustness to increasing K. As explained in Results, it is essential to verify that the domain of the parameters in which the multistability between balanced states occurs does not vanish in the limit of large N and large K. To test for this robustness, we performed numerical simulations with a network of size up to $N = 320{,}000$ neurons and connectivities as large as $K = 32{,}000$.

Synaptic parameters. Unless specified otherwise, the parameters of the models used in our simulations are those given in Table 1.

In both models, pyramidal cells form a mixture of fast (AMPA) and slow (NMDA) synapses on other pyramidal cells and interneurons. Both components share the same connectivity matrices $J_{i\alpha,jE}$ but differ in their synaptic strength (g^{AMPA} and g^{NMDA}, respectively) and in the decay time constant of their PSCs. Equation 5 must be interpreted as including the sum over both components. The decay time constant of the excitatory postsynaptic currents are 3 and 50 ms for AMPA and NMDA synapses, respectively. We tested to confirm that taking longer time constants for the NMDA synapses (for instance 80 ms, as reported by Wang et al., 2008) has no impact on the results (data not shown).

The voltage dependence of the NMDA currents was not included in the simulations depicted in this paper. However, we have verified that including voltage dependence while keeping the STP of the recurrent excitation (*EE*) synapse interactions does not qualitatively change the conclusions of our work. For that purpose, we multiplied the synaptic strength, $G_{\alpha E}^{\mathrm{NMDA}}$, by the same voltage-dependent factor as in Compte et al. (2000). Since this factor is always <1, we had to increase G_{EE}^{NMDA} and G_{IE}^{NMDA} by some constant factor equal to 5.

We define R_α as follows: $R_\alpha = g_{\alpha E}^{\mathrm{AMPA}}/(g_{\alpha E}^{\mathrm{NMDA}} + g_{\alpha E}^{\mathrm{AMPA}})$. In most of the simulations we took $R_E = 0.5$ and $R_I = 0.9$. This accounts for the fact that NMDA synapses are more abundant between pyramidal cells than between pyramidal cells and interneurons (Thomson, 1997). All the inhibitory interactions have a decay time constant of 4 ms (Bartos et al., 2001, 2002). We verified that the properties of the model were robust with respect to the values of the excitation and inhibition decay-time constants, as well as with respect to the ratios R_E and R_I.

The recurrent excitation between pyramidal cells displays short-term plasticity. For simplicity, in the simulations described in this paper, we assumed that this is the case for AMPA as well as for NMDA interactions. However, we verified that the properties of the network are essentially the same if only AMPA synapses display STP (provided that the synaptic strength G_{EE}^{AMPA} is increased appropriately to compensate for the absence of facilitation in NMDA synapses). The parameters of the STP were chosen such that the network displays multistability in a broad range of background inputs. The chosen values for the recovery and the facilitation time constants were compatible with the *in vitro* data of Wang et al. (2006). The utilization parameter, U, was in the range of the lower values reported in that study. We assume that excitatory synapses to inhibitory neurons as well as all inhibitory synapses do not display STP. Therefore for these synapses $x_{j\beta,n} = u_{j\beta,n} = 1$.

With the parameters in Table 1, the maximal postsynaptic potentials (PSPs) of the various connections are as follows: 4.3 10^{-2} mV and 0.14 mV for the NMDA and AMPA components of the *EE* synapses, respectively; −2.3 mV for the *EI* synapses, −2.5 mV for the *II* synapses; and 2.5 10^{-2} and 1 mV for the NMDA and AMPA components of the excitatory to the inhibitory neurons, respectively. Note that the PSPs generated by individual excitatory connections are substantially weaker than those generated by inhibitory ones. This is partially compensated for by excitatory connectivity that is larger by a factor of 4 than the inhibitory connectivity, and by greater activity for inhibitory than for excitatory neurons. Moreover, the neurons receive an excitatory tonic input. Altogether, excitation and inhibition inputs balance approximately as we show in the results.

Numerical simulations. Simulations were performed using a second-order Runge–Kutta scheme with a fixed time step, $\delta t = 0.1$ ms, supplemented by an interpolation scheme for the determination of the firing times of the neurons (Hansel et al., 1998).

Characterization of the irregularity in action potential firing. We quantify the irregularity of the discharge of a neuron by the coefficient of variation (CV) of its interspike interval (ISI) distribution defined by:

$$CV = \frac{\langle (\delta_n - \langle \delta_n \rangle)^2 \rangle^{1/2}}{\langle \delta_n \rangle}, \tag{11}$$

where δ_n is n-th ISI of the neuron and $\langle \ldots \rangle$ denotes an average overall number of spikes it has fired.

We also evaluate the coefficient of variation CV_2. CV_2 is computed by comparing each ISI (δ_n) to the following ISI (δ_{n+1}) to evaluate the degree of variability of ISIs in a local manner (Holt et al., 1996). It is defined by:

$$CV_2 = 2\frac{\langle |\delta_n - \delta_{n+1}| \rangle}{\langle \delta_n + \delta_{n+1} \rangle}. \tag{12}$$

For a Poisson spike train, $CV = CV_2 = 1$.

Evaluation of the phase diagrams. To evaluate which regions in the space of parameters display persistent activity, we use the following procedure: In simulating the network, we slowly increase the external input I_E (while keeping the rest of the parameters constant) and monitor the mean and spatial modulation of the network activity. When I_E has reached a predefined value of sufficient size, we continue the simulations while decreasing I_E back to its initial value. This generates a hysteresis curve, which enables us to identify the bistability region for that point in the parameter space. The procedure is repeated for different values of the parameters (e.g., G_{EE}) to obtain a phase diagram.

Simulation of the delayed response task in Model II. At the beginning of a trial, the network is initialized from random initial conditions. After 3 s (representing the fixation period in the experiment), the cue is presented and the related feedforward input occurs for $\Delta t_{\mathrm{cue}} = 0.5$ s. The delay period goes from 3.5 to 6.5 s. The transient input, which erases the memory, begins at $t = 6.5$ s and has a duration of $\Delta t_{\mathrm{erase}} = 1$ s.

Quantification of the single neuron directional selectivity in Model II. Tuning curves were estimated by simulating 20 trials for each of the eight cues from 0 to 315° in intervals of 45°. Using a bootstrap method, we determined whether the task-related activity of a neuron was directionally tuned (Constantinidis et al., 2001). For each neuron we evaluated the quantity $O_{i,\alpha} = [\Sigma_\theta r_\theta^2]^{1/2}$ where r_θ is the firing rate of the neuron during the delay period averaged over the 20 trials with the cue presented in direction θ. We compared the obtained value of $O_{i,\alpha}$ with the one obtained after randomly permuting the angles of each trial before averaging. If the second quantity is smaller than the first one for 99% of the permutations, we consider that the activity is significantly directionally tuned.

We quantified the degree of directional selectivity with the circular variance (CircVar) (Mardia, 1972) defined by CircVar $\equiv 1 - c_1/c_0$ where

$c_k = |\Sigma_\theta\, r_\theta \exp(ik\theta)|$. A broad tuning curve (badly selective response) corresponds to CircVar close to 1 whereas for very sharp tuning CircVar is close to 0.

The tuning curves of a large fraction of the neurons can be well fitted to a von Mises function defined as:

$$r(\theta) = A + B\exp\left(\frac{\cos(\theta - \psi) - 1}{D}\right). \quad (13)$$

We estimated the parameters A, B, ψ, D for each neuron by minimizing the quadratic error of the fit: $E = \Sigma_\theta(r(\theta) - r_\theta)^2/\sigma_\theta^2$, where the sum is over the eight directions of the cue and σ_θ is the trial-to-trial SD of the response. The estimate of the preferred direction (PD) of the neuron is given by PD = $180°\psi/\pi$. The sharpness of the tuning curve [tuning width (TW)] above baseline can be computed from the formula:

$$TW = \frac{180°}{\pi}\cos^{-1}\left(1 + D\log\frac{1 + \exp\left(-\frac{2}{D}\right)}{2}\right). \quad (14)$$

The quality of the fit is estimated by evaluating the χ^2 distribution for 4 degrees of freedom (8 points minus 4 parameters) (Press et al., 1992). This probability characterizes the goodness-of-fit. Bad fits correspond to extremely low values of the probability, q. We consider that the fit is good if $q > 0.001$. To determine the spatial modulation of the network activity and population vector in Model II, let us denote by $f_{j\alpha}(t)$ the firing rate of neuron (j, α) averaged over a time window of 50 ms around time t. To characterize the spatial modulation of the activity of population α at time t we computed:

$$Z_\alpha(t) = \sum_j f_{j\alpha}(t)\exp(i\theta_{j,\alpha})/\sum_j f_{j\alpha}(t) \quad (15)$$

$$= M_\alpha(t)\exp(i\Psi_\alpha(t)), \quad (16)$$

where M_α is the modulus of the complex number Z_α and ψ_α is its argument. Note that the real and imaginary parts of $Z_\alpha(t)$ are the components of the population vector at time t for population α. If the network activity is homogeneous, $M_\alpha(t)$ is ~0, whereas for a very sharply modulated activity profile, $M_\alpha(t)$ is ~1. In our simulations, we always found that $\psi_E(t)$ is approximately equal to $\psi_I(t)$. The preferred direction, θ_{erase}, of the feedforward input that erases the memory trace was taken to be the value of $\psi_E = \psi_I$ at the end of the delay period.

Results

Irregular firing in cortex *in vivo* and balance of excitation and inhibition

Cortical neurons fire irregularly (Burns and Webb, 1976; Softky and Koch, 1993; Bair et al., 1994). The neurons that fire less irregularly are those in primary motor cortices, in supplementary motor cortices, or in association with or in motor areas, such as parietal regions (Maimon and Asaad, 2009; Shinomoto et al., 2009), where the CV of the ISI distributions of the neurons are in the range $CV = 0.5–0.8$. The neurons that fire more irregularly are in sensory areas and in prefrontal cortex where the CVs are ~1.

Remarkably, recent experimental studies in monkeys performing WM tasks have reported that the level of temporal irregularity with which PFC neurons fire during the delay period is comparable to, if not higher than, what is observed in spontaneous activity or during the fixation period (Shinomoto et al., 1999; Compte et al., 2003; Shafi et al., 2007).

The highly irregular activity of cortical neurons *in vivo* has long appeared paradoxical in view of the large number of their synaptic inputs (Softky and Koch, 1992, 1993; Holt et al., 1996). This is because the temporal fluctuations of the postsynaptic current produced by $K \gg 1$ presynaptic afferents firing asynchronously are much smaller, by a factor $1/\sqrt{K}$, than its average. Accordingly, since *in vitro* neurons fire regularly in response to weakly noisy input (Connors et al., 1982), one would expect that firing *in vivo* would be only weakly irregular. A generic solution to this problem posits nearly balanced strong excitatory and inhibitory synaptic inputs such that their temporal fluctuations, although much smaller than their means taken separately, are comparable to the average total input and to the neuronal threshold (van Vreeswijk and Sompolinsky, 1996, 1998). Modeling studies have shown that balanced states emerge in a robust manner without fine tuning of parameters from the collective dynamics of recurrent neuronal networks (van Vreeswijk and Sompolinsky, 1996, 1998, 2004; Amit and Brunel, 1997; Lerchner et al., 2006; Hertz, 2010). The balance mechanism has been applied to account for the high variability of spontaneous activity as well as sensory-evoked neuronal activity in cortex (van Vreeswijk and Sompolinsky, 2004; Lerchner et al., 2006). Can it also provide a natural framework to account for the spiking irregularity observed during WM tasks?

In balanced states, the level of activity of macroscopic ensembles of neuronal populations are largely independent of single-cell intrinsic properties (van Vreeswijk and Sompolinsky, 1998). This can be understood heuristically as follows. Let us consider a network comprising one excitatory population (E) and one inhibitory population (I) (Fig. 1*A*). The state of each population is characterized by its activity, f_α, $\alpha = E, I$. Assuming a stationary state of the network, this activity is related to the total input into the population α, h_α, via $f_\alpha = S_\alpha(h_\alpha)$ where S_α is the sigmoidal input–output transfer function of population α and h_E and h_I are given by (see for instance Dayan and Abbott, 2001):

$$h_E = G_{EE}f_E - G_{EI}f_I + I_E \quad (17)$$

$$h_I = G_{IE}f_E - G_{II}f_I + I_I, \quad (18)$$

where the constants $G_{\alpha\beta}$ measure the efficacy of the interactions between population β and α and I_E and I_I are external inputs to the network.

If the excitation is too strong, the inputs h_E and h_I and therefore the activities of the populations reach the saturation levels of S_E and S_I. Conversely, for overly strong inhibition, h_E and h_I are below the (soft) threshold of S_E and S_I and network activity is very low. An appropriate balance of excitation and inhibition is necessary to prevent the network from being in one of these extreme regimes. This occurs if the activities of the two populations obey the conditions:

$$G_{EE}f_E - G_{EI}f_I + I_E \approx 0 \quad (19)$$

$$G_{IE}f_E - G_{II}f_I + I_I \approx 0, \quad (20)$$

which express the very fact that inhibition balances excitation. These balance conditions do not depend on the input–output transfer functions of the populations. They fully determine the population activities as a function of the external inputs. Since these equations are linear, there is generically a unique solution for given values of I_E and I_I. Hence the network cannot exhibit more than one balanced state. This effective washout at the macroscopic level of the neuronal intrinsic properties is a remarkable feature of the balance regime. As a matter of fact, it can be derived in large networks of randomly connected binary neurons (van Vreeswijk and Sompolinsky, 1998), of randomly connected spiking integrate-and-fire neurons (Renart et al., 2010), or of integrate-and-fire networks (Lerchner et al., 2004; van Vreeswijk and Sompolinsky, 2004).

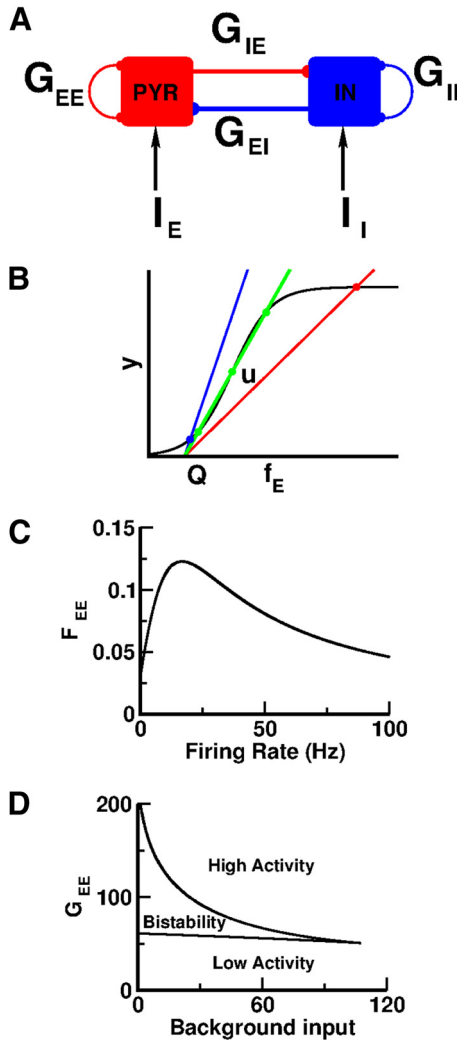

Figure 1. Bistability between balanced states sustained by nonlinear recurrent excitation in a two-population rate model. ***A***, Architecture of the model. ***B***, Graphic solution of the balance equations for the sigmoidal synaptic transduction function, $S_{EE}(f_E)$: Black curve: The function $S_{EE}(f_E)$. Straight lines: $y = (f_E - Q)/G$, $Q = 0.5$; blue: $G = 20$; green: $G = 3.3$; red: $G = 9.3$. The intersections between the straight lines and the black curve correspond to the possible states of the network. For $G = 20$ and $G = 3.3$, the network has only one stable state. For $G = 9.3$, it is bistable and also displays one unstable state (indicated by the symbol *u*). ***C***, Nonlinear input–output transfer function $F_{EE}(f_E) = (a + bf_E)/(1 + cf_E + df_E^2)$ *with* $a = 0.03$, $b = 0.0135$ s, $c = 0.0195$ s, $d = 2.7\ 10^{-3}$ s^2 (top). ***D***, Phase diagram of the network. Background inputs to the *E* and *I* populations are equal ($I_E = I_I$). Parameters: $G_{EI} = 2.5$, $G_{IE} = 9$, $G_{II} = 3$.

According to the classical theory of WM, the selective persistent activity observed during the delay period reflects the coexistence of many collective stable states of the network dynamics in PFC. The argument above implies that for these states to be balanced, other nonlinearities than those present in the input–output transfer functions of the neurons are required. This prompted us to inquire which nonlinearities other than those of single neurons are sufficient to achieve multistability between balanced states.

Bistability between balanced states can be robustly sustained by nonlinear recurrent interactions

The simplest form of persistent activity is exhibited by neural networks that possess two stable states that differ by the level of activity of the neurons. If the network is in one of these states, it remains there until an appropriate perturbation of a macroscopic number of neurons induces a transition to the other state. We will term the state in which the activity is lower as *baseline* and the state in which the activity is elevated as *persistent.*

Bistability between balanced states in a simplified rate model with nonlinear synaptic interactions

It is clear that the linearity of Equations 19 and 20 stems from the assumption that the interactions between the neurons are linear; namely, that the synaptic inputs are proportional to the presynaptic firing rates and linearly sum up.

We first relax this assumption by considering that the recurrent excitatory interactions depend nonlinearly on the activity of the excitatory neurons. This means that we replace G_{EE} in Equations 19 and 20 by a term $G_{EE}F_{EE}(f_E)$, which depends nonlinearly on the activity of the excitatory population. Equations 19 and 20 are therefore replaced by:

$$G_{EE}S_{EE}f_E) - G_{EI}f_I + I_E = 0 \quad (21)$$

$$G_{IE}f_E - G_{II}f_I + I_I = 0, \quad (22)$$

with $S_{EE}(f_E) = f_E F_{EE}(f_E)$. Expressing f_I as a function of f_E using Equation 22 and inserting in Equation 21 we get:

$$\frac{f_E - Q}{G} = S_{EE}(f_E), \quad (23)$$

where

$$G = G_{II}G_{EE}/(G_{IE}G_{EI}) > 0, \quad (24)$$

and $Q = (G_{II}I_E/G_{EI} - I_I)/G_{IE}$. Equation 23 determines f_E as a function of the model parameters. Note that this equation is formally equivalent to the one that determines the firing rate of a population of excitatory neurons with an input–output transfer function, S_{EE}, coupled recurrently with linear interactions of strength *G*, receiving an external input *GQ*. It can be solved graphically: its solutions are given by the intersections of the straight line $y = (f_E - Q)/G$ with the curve $y = S_{EE}(f_E)$.

Of particular interest is the case in which S_{EE} has a sigmoidal shape (Fig. 1*B*). Then, for *G* small (Fig. 1*B*, blue line) or *G* large (red line) only one solution exists (blue and red points, respectively). For intermediate *G* (green line), three solutions coexist (green points). One corresponds to a low activity state and another to a high activity state. In the third solution (point *u*) the activity is at an intermediate level. Stability analysis reveals that the low and high activity states are stable whereas the intermediate state is unstable. Therefore, a network with such nonlinear recurrent excitatory interactions can display bistability between two balanced states.

As an example, we consider the function F_{EE} plotted in Figure 1*C*. We numerically solved Equation 23 for different values of the external input and the strength of the recurrent excitation. The resulting phase diagram is plotted in Figure 1*D*. It shows that there is a large domain in the parameter space where the network displays bistability. In this domain, the balanced conditions, Equations 21 and 22, are fulfilled. Hence, the bistability is between balanced states.

Similar analyses can be performed when nonlinearities are present in *II*, *EI*, or *IE* interactions. This shows that bistability between balanced states can also occur if the *II* interactions are nonlinear with a sigmoidal transfer function. However, nonlinearities only in *EI* or *IE* interactions are not sufficient to sustain bistability of balanced states (results not shown).

Nonlinearities induced by facilitating recurrent excitatory synapses can sustain bistability between balanced states in a spiking network

The analysis above provides insights into the possibility of achieving bistability between balanced states in large neuronal networks in which excitatory recurrent synaptic currents are sigmoidal functions of the firing rates of the presynaptic neurons. Synapses exhibiting STP with facilitation at a low presynaptic firing rate display input–output transfer functions that exhibit this feature. This suggests that STP may underlie bistability of balanced states. Let us note that it has been previously found that synapses with STP can give rise to bistability in a fully connected network of excitatory integrate-and-fire neurons, although in this case no irregular firing is observed (Hempel et al., 2000).

We investigated this hypothesis in a network consisting of two large populations of integrate-and-fire neurons with random and unstructured connectivity (see Materials and Methods for details). The recurrent excitatory synapses are endowed with STP described according to the model of Markram et al. (1998). The efficacy of an *EE* connection, $G_{EE}ux$, is the product of the maximal efficacy G_{EE}, the amount of available synaptic resources x, and the utilization fraction of resources u. At each presynaptic spike, the synapse depresses due to depletion of neurotransmitter and it also facilitates due to calcium influx. As a result, the variable x is reduced by a quantity ux (depression) and the fraction u increases (facilitation). Between spikes, u relaxes to its baseline level, U, and x recovers to 1, with time constants τ_f and τ_r, respectively. The parameters we use for the STP are given in Table 1. Figure 2*A* depicts the steady-state input–output transfer function for these parameters when the presynaptic spike train has Poisson statistics or when it is periodic. In both cases, the shape is similar to the one in Figure 1*C*. Note that the shape of the input–output transfer function depends only weakly on the spike statistics.

We performed extensive numerical simulations to study the dependence of the network steady states on the model parameters. Figure 2*B* shows the phase diagram of the model as a function of the strength of the recurrent excitation and the background inputs. All other parameters of the model are given in Tables 1 and 2. It is qualitatively similar to the phase diagram of our nonlinear rate model (Fig. 1*D*). It displays a wide region of bistability between a low (baseline) and an elevated (persistent) activity state. In this region, the network prepared in the baseline state remains in that state. However, a transient input of appropriate intensity and duration induces a switch of the network to the activity-elevated state. The network persists in that state until another appropriate transient input switches it back to baseline (Fig. 2*C*).

The balance regime is characterized by excitatory and inhibitory inputs into neurons that are much larger than the neuronal threshold, whereas the temporal mean and temporal fluctuations of the total (net) inputs are comparable to the threshold. Figure 3*A* shows the excitatory (red), inhibitory (blue), and total synaptic currents (black) to an excitatory neuron in the network in the baseline and in the persistent states. In both situations, the time average of the excitatory current into this neuron is much larger than the threshold. However, it is compensated for to a large extent by a strong inhibition. This results in a total input whose temporal mean is below threshold at a distance comparable to the amplitude of the input temporal fluctuations. As a result, in baseline as well as during the delay period, the action potentials this neuron fires are driven by the temporal fluctuations. The resulting spike trains are highly irregular in both epochs.

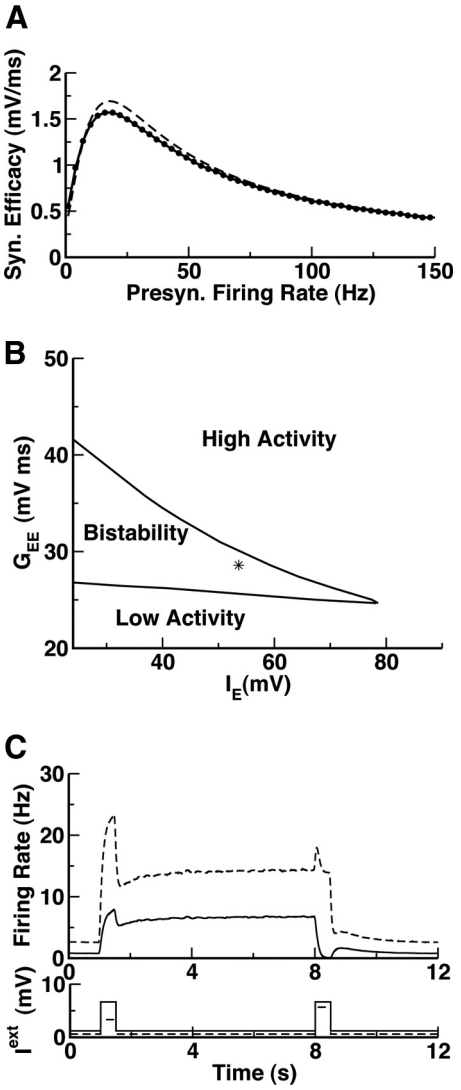

Figure 2. Bistability between balanced states induced by STP in recurrent excitation in a two-population network of integrate-and-fire neurons. Parameters as in Table 1. We keep the relationship $I_E^b = I_I^b$. ***A***, The transduction function of the recurrent excitatory synapses facilitates (resp. depresses) at a low (resp. high) presynaptic firing rate. This function was computed by simulating the model synapse (Eqs. 6, 7 in Materials and Methods) and the stationary value (after 5 s of simulation) of the product ux averaged over 100 trials. Dashed line, Periodic input. Dots, Input with Poisson statistics. ***B***, Phase diagram of the network ($G_{EI} = 2.6$). The star indicates the parameters used in ***C***. ***C***, Top, The population average activity of the excitatory (solid line, low activity state: $f_E = 1.25$ Hz; high activity state: $f_E = 3.9$ Hz) and inhibitory (dashed line, low activity state: $f_I = 4.1$ Hz; high activity state: $f_I = 9.32$ Hz) populations. Bottom, External inputs (background plus transient inputs).

The histograms plotted in Figure 3*B* show that these features are not specific to this particular neuron. For all the neurons, the mean excitatory and inhibitory currents are much larger than the threshold in baseline as well as in the persistent state. However, in both states the mean net input is comparable, in absolute value, to the threshold and the input fluctuations. It is in general below threshold, but the fluctuations are large enough to bring the membrane potential of the neurons above threshold. This is clear from the histogram of the mean inputs plus 1.5 SDs plotted in green in Figure 3*B*. The distribution of the membrane potentials in the two populations can be seen in Figure 3C. We can see that even if the firing rate is higher in the delay state than in baseline, the membrane potentials tend to be smaller in the second case

Table 2. Parameters of the external current for the unstructured and spatial working memory networks

	Unstructured	Spatial WM
i_E^b	1.66 mV	1.66 mV
$\hat{a}_E^{\text{cue}}$	4.66 mV	2.4 mV
ϵ_E^{cue}	0	0.17
$\hat{a}_E^{\text{erase}}$	2.66 mV	5.2 mV
$\epsilon_E^{\text{erase}}$	0	0.23
i_I^b	0.83 mV	0.83 mV
$\hat{a}_I^{\text{cue}}$	2.33 mV	1 mV
ϵ_I^{cue}	0	0
$\hat{a}_I^{\text{erase}}$	1.83 mV	3.7 mV
$\epsilon_I^{\text{erase}}$	0	0.28

because the mean total input decreases. The activity of the neurons is highly irregular in both states. This is depicted in Figure 3*D*, where the spike rasters of a subset of neurons are plotted. This is also confirmed by Figure 3*E*, which plots the *CV* of the ISI of 2000 neurons as a function of their averaged firing rates.

Interestingly, in the persistent state, the mean net inputs into the neurons are more negative than during baseline. However, the resulting mean hyperpolarization of the neurons is compensated for by an increase in their input temporal fluctuations in such a way that the neuronal activity is larger in the persistent states than in baseline. The firing is more irregular in the persistent state: the histogram of the *CV* of the ISI distributions is shifted toward values larger than during baseline (Fig. 3*E*). This is a consequence of the increase in the input temporal fluctuations.

Note that our model does not incorporate the voltage dependence of NMDA synapses (Jahr and Stevens, 1990). This is another nonlinearity that will not be washed out in the balanced state. We have checked that these nonlinearities do not affect the qualitative behavior of the model when STP is present, but they are incapable by themselves of generating persistent activity in a balanced state. This is because as the firing rate increases the mean total input decreases and the mean membrane potential decreases also (Fig. 3*C*). The increase in the firing rate is allowed only by the increase in the fluctuations, but the NMDA synapse filters those fluctuations and is dominated by the mean voltage. Therefore the voltage-dependent NMDA synapse will not become potentiated as firing rate increases, and cannot provide a suitable substrate for WM.

Robustness of the bistability regime with respect to connectivity changes

We investigated the robustness of the bistability regime in our model with respect to changes in connectivity K by simulating the network with different values of K while the strength of the interactions and the external inputs are scaled according to Equations 8, 9, and 10 (van Vreeswijk and Sompolinsky, 1996, 1998, 2004). This scaling guarantees that the temporal fluctuations in the total inputs into the neurons remain similar while increasing K.

Changing the connectivity from $K = 2000$ to $K = 4000$ has some effect in the phase diagram of the network as indicated by the comparison of the solid and dashed lines in Figure 4*A* (left). The lower boundary of the bistable region moves slightly upward but at the same time the upper boundary also moves in the same direction. The latter move is larger than the former. Hence, in fact, the bistable region is slightly larger for $K = 4000$. This suggests that the phase diagram remains essentially the same when K increases. Figure 4*A* (right) plots the critical value of the background current on the boundary of the bistable region for $G_{EE} = 1.6$ V ms for

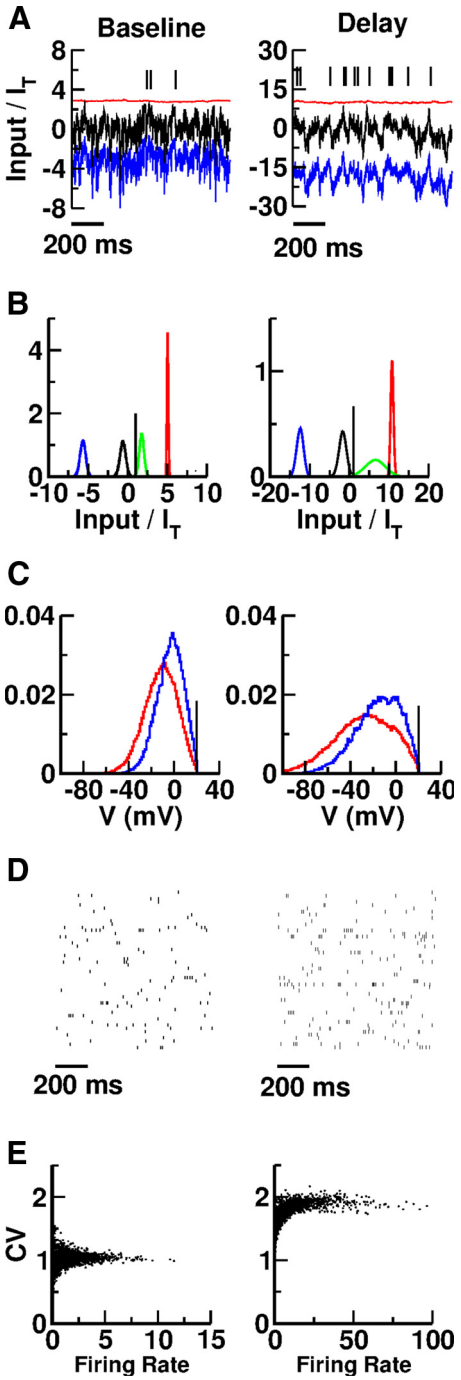

Figure 3. Inhibition balances excitation in the baseline as well as in the persistent state. ***A***, Input currents into an excitatory neuron. Red, Excitatory input (recurrent plus background). Blue, Inhibitory input. Black, Total input (excitatory plus inhibitory). The firing rate and the *CV* of the ISI histogram of this neuron are as follows: baseline: $f = 1.4$ Hz, $CV = 1.2$; delay: $f = 16.6$ Hz, $CV = 1.9$. ***B***, Population histograms of the inputs into the neurons normalized to the firing threshold. Blue, Time-averaged inhibitory input. Red, Time-averaged excitatory input. Black, Time-averaged total input. Green, Time-averaged total input plus 1.5 SDs of the total input fluctuations. All neurons are included in the histograms. Vertical line corresponds to threshold. ***C***, Population histograms of the membrane potentials. Red, Excitatory population. Blue, Inhibitory population. The membrane potentials of all the neurons are included in the histograms. The potentials are sampled with a rate of 0.1 Hz. Vertical line corresponds to threshold. ***D***, Spike trains of 200 excitatory neurons during baseline and delay periods. ***E***, *CV* versus firing rates. One thousand neurons in each population are included. The histograms of the *CV*s and the firing rates are plotted in Figure 4. The results plotted in ***C*** and ***D*** were obtained in simulations 100 s long.

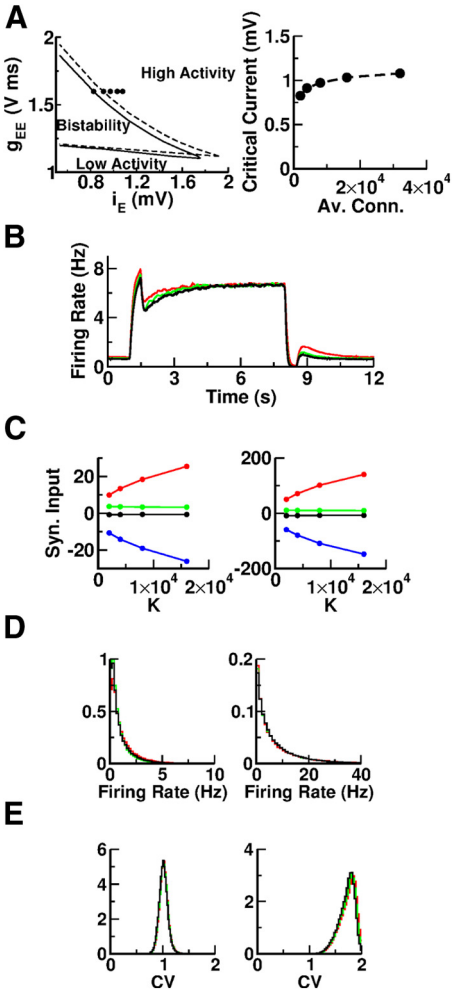

Figure 4. The bistable regime is robust with respect to changes in the average connectivity, *K*. ***A***, Left, Phase diagram for $K = 2000$ (solid line) and $K = 4000$ (dashed line). The dots show the critical value of the background current below which the network displays bistability for $g_{EE} = 1.6$ V ms and $K = 2000$, 4000, 8000, 16,000, and 32,000 (left to right). Right, The critical value of the background current below which the network displays bistability as a function of *K*. For the simulations with $K = 2000$, the network size was $N = 80{,}000$. For the other values of *K*, $N = 160{,}000$. ***B***, The average activity of the excitatory population versus time for different values of *K*. Red, $K = 2000$; green, $K = 4000$; black, $K = 8000$. The network size is $N = 80{,}000$ and the synaptic strength is $g_{EE} = 1.23$ V ms. ***C***, The population averages of the mean excitatory input (red), total inhibitory input (blue), mean net input (black), and of the input fluctuations (green). Left, Baseline. Right, Delay period. ***D***, Histograms of the firing rates for $K = 2000$ (red), $K = 4000$ (green), and $K = 8000$ (black); $N = 80{,}000$. All the neurons in the two populations are included. Left, Baseline. Right, Delay period. The histograms are very similar for all the three values of the connectivity. ***E***, Histograms of the *CV* for $K = 2000$ (red), $K = 4000$ (green), and $K = 8000$ (black); $N = 80{,}000$. *CV*s were estimated from spike trains 100 s long. All the neurons with a firing rate larger than 0.5 Hz in the two populations are included. Left, Baseline. Right, Delay period. The histograms are almost identical for all three values of the connectivity.

$K = 2000$ up to $K = 32{,}000$. The overall variation of the critical current suggests that it saturates as *K* becomes very large.

The properties of the dynamical states of the network for the reference set of parameters (see Tables 1, 2) are also compared in Figure 4 for $K = 2000$, 4000, and 8000. Figure 4*B* confirms the robustness of the bistability with respect to *K*. Indeed, the network is bistable, the population average activities in the two coexisting stable states are the same, and the overall dynamics of activity during the switch on/off are very similar for all the values of *K* tested. Figure 4*C* plots, as a function of *K*, the population average of the mean excitatory currents (red) and mean inhibitory currents (blue) as well as the mean (black) and the fluctuations (green) of the net inputs into one excitatory neuron. The mean excitation and the mean inhibition increase proportionally to $\sqrt{K}$. This contrasts with the mean and the fluctuations of the net inputs, which remain almost constant and on the order of the neuronal threshold. These features indicate that the baseline as well as the elevated activity states are balanced independently of *K*. Finally, Figure 4*D*,*E* plots the histograms of the single neuron firing rates and *CV* for different values of *K*. It is clear that increasing the connectivity has almost no effect on these distributions.

We therefore conclude that the bistability, the excitation and inhibition balance, the irregularity, and the heterogeneity of the neuronal activity are robust in our network with respect to change in connectivity.

Network mechanism underlying visuospatial WM

The classical framework to investigate visuospatial WM in primates is the ODR task schematically represented in Figure 5*A*. In this task, a monkey needs to remember the location of a stimulus for a delay period of several seconds and make a saccade in that direction at the end of the period (Funahashi et al., 1989). Electrophysiological recordings performed in dorsolateral prefrontal cortex of primates has revealed that neurons in this region modify their activity persistently and selectively to the cue direction during the delay period (Funahashi et al., 1989, 1990, 1991; Constantinidis et al., 2001; Takeda and Funahashi, 2007). It is believed that this selective persistent activity is the neural correlate of information on the location of the cue that has to be memorized to perform the saccade at the end of the delay period.

A theoretical framework to account for this selective persistent delay activity is a recurrent network made of identical neurons with the geometry of a ring and a connectivity pattern such that the interaction between two neurons depends solely on their distance on the ring (Camperi and Wang, 1998; Compte et al., 2000). With sufficiently strong and spatially modulated recurrent excitation and appropriate inhibition, the network operates in a regime of multistability between a state in which the activity is homogeneous and a set of states characterized by a bumpy activity profile. The "bump" can be localized at an arbitrary location if the network connectivity is tuned so that it is rotationally invariant. During the cue period, a transient stimulus tuned to a specific location on the ring, corresponding to the direction to be memorized, selects the state in which the bump peaks at that location. After the stimulus is withdrawn, the network remains in this state. Therefore, the network is able to encode the memory of the cue location.

A crucial ingredient in this hypothesis is that neuronal populations respond in a nonlinear fashion to external inputs. This is essential not only to generate the persistence of neuronal activity during the delay but also its bumpy localized profile (i.e., selectivity) (Hansel and Sompolinsky, 1998). In all the models studied so far to account for selective persistent activity in ODR tasks (Compte et al., 2000; Barbieri and Brunel, 2008), the population nonlinearities are induced by nonlinear input–output transfer functions of single neurons. As argued above, the network in these models cannot display baseline as well as memory balanced states because of the washout at the population level of the latter nonlinearities. In the following we show that this becomes possible if the nonlinearities are generated by short-term facilitation in the recurrent excitatory synapses.

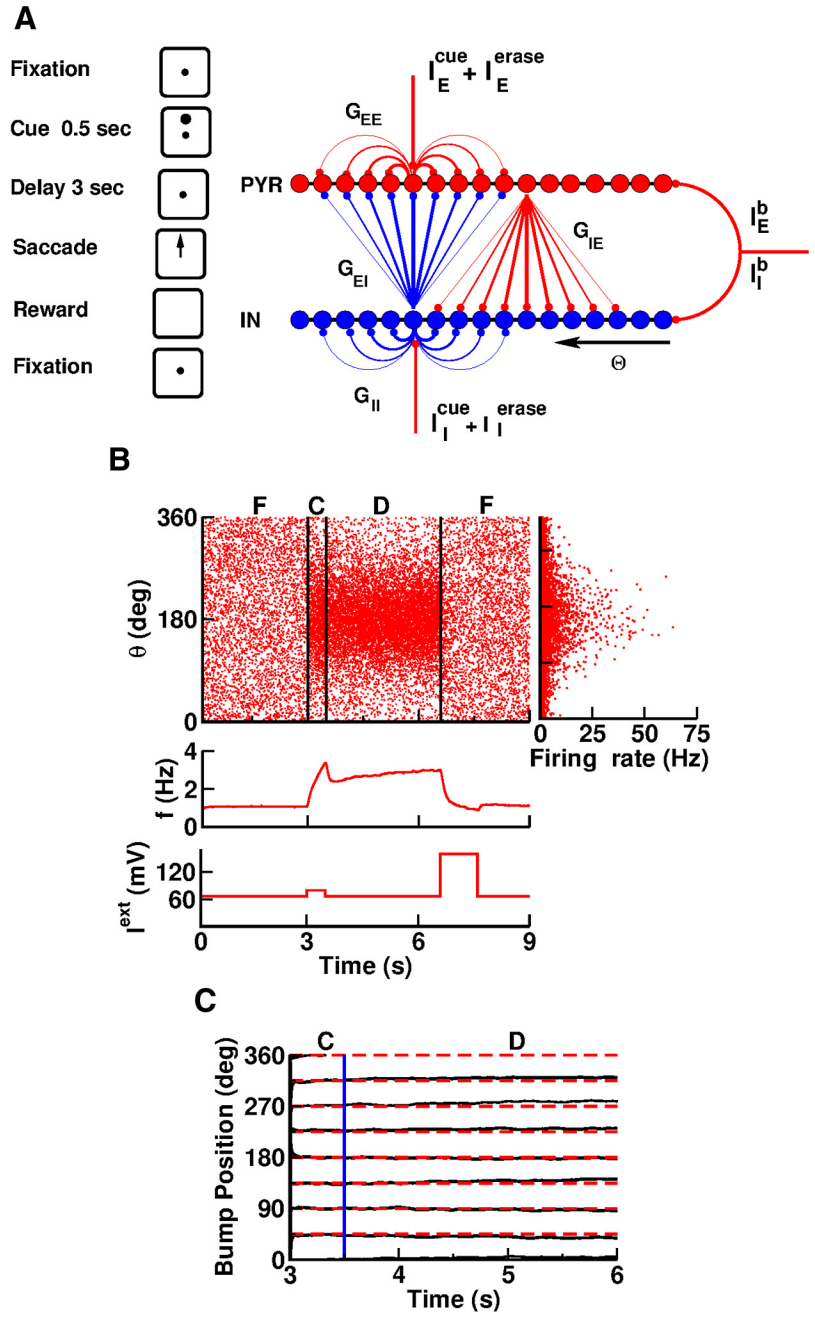

Figure 5. The PFC network model. ***A***, Left, The visuomotor WM task. A trial begins while the monkey fixates the screen center. A visual cue appears for 0.5 s. The monkey must memorize the cue direction during a 3 s delay period while it maintains fixation. At the end of the delay it must make a saccade in the cue direction and restore fixation. Right, Model architecture. Red dots, Excitatory (pyramidal) neurons. Blue dots, Inhibitory interneurons. Neurons are arranged according to their preferred directions θ. The probability of connection decreases as a function of the difference between the preferred directions. The interaction strengths are G_{EE}, G_{IE}, G_{EI}, and G_{II}. The background currents I_E^b, I_I^b are applied uniformly to all the neurons. The transient currents I_E^{cue}, I_I^{cue}, I_E^{erase}, and I_E^{erase} are applied as explained in Materials and Methods. ***B***, One trial of the task simulated in the network. Cue direction, 180°. Fixation, cue, and delay periods are denoted by F, C, and D, respectively. Top, Left, Spike rasters. Ten percent of the excitatory neurons are included. Top, Right, Spatial profile of the time-averaged activity of the neurons during the delay period. Twenty-five percent of the neurons are represented. Middle, Population-averaged activity of excitatory neurons versus time. Bottom, External input to excitatory neurons versus time. ***C***, The position of the bump versus time during cue (C) and delay (D) periods for eight directions of the cue (every 45°, dashed lines). The blue vertical line denotes the beginning of the delay period (see Materials and Methods for details).

Balanced visuospatial WM can be sustained by STP

We consider a large recurrent network of integrate-and-fire neurons with a ring architecture (Fig. 5*A*; Materials and Methods). Each neuron is parametrized by its location, θ, on the ring. The probability of connection between two neurons decreases with their distance according to Equation 4. The neurons receive a homogeneous background input that is constant in time. During the cue period, the neurons receive an additional input. This stimulus depends on the direction of the cue according to Equation 3. It is maximum for $\theta_{cue} = \theta$. Another transient input erases the memory at the end of the delay period. A crucial feature of the model is that excitatory recurrent synapses display STP as in Figure 2*A*. See Materials and Methods for the details of the model.

We studied this network using numerical simulations that mimic the experimental protocol of an ODR task. Figure 5*B* shows a typical trial. At the start, during the precue period, the network activity is low and homogeneous: the network is in its baseline state. The visual cue is presented for 500 ms in direction $\theta_{cue} = 180°$. Subsequently the neurons near $\theta = 180°$ elevate their activity. Upon removal of the cue, the network relaxes to a state with a "bumpy" pattern of activity localized near θ_{cue}. The network activity remains elevated and localized close to $\theta = \theta_{cue}$ (Fig. 5*C*), maintaining memory of the cue direction during the 3 s delay period. Eventually, the transient input at the end of the delay period erases this memory.

For the duration of the turn-on input we have chosen 0.5 s, as this is the typical duration of the cue period in the ODR task experiments (Funahashi et al., 1989). We have also assumed that the cue-related input is tuned to account for the direction selectivity of the cue period activity of PFC neurons. With the parameters we have chosen, the cue-related input, averaged over all directions, is not very different from the background input. The modulation with the direction, ε, is 0.17. So the maximum/minimum of this input are only 17% larger/smaller than the background. In fact, our simulations indicate that the mean cue-related input can be taken as the same as for the background input (results not shown). The most critical parameter is ε, which needs to be at least 0.1. Therefore, the transient input necessary to activate the persistent state does not need to be large compared with background.

As far as we know, nothing is known experimentally about the strength and the duration of the external inputs responsible for turning the switch off. Simulations show that erasing of the memory can be achieved in various ways in our model. In fact, a broad range of stimulus parameters and stimulus durations can achieve switch-off. The most efficient way is a sufficiently strong and long-lasting feedfoward

inhibitory input into the excitatory population, or a strong excitatory input on the inhibitory neurons. However, such stimuli suppresses almost completely the activity in all the neurons. This is not observed in experiments. In the parameters of our reference set, both populations of neurons are excited by the transient stimulus that erases the memory. The excitation of the interneurons induces an inhibition on the pyramidal cells. This inhibition is to some extent balanced by the feedforward excitatory input these neurons receive via the erasing stimulus. As a result, this input reduces their activity and only some of the pyramidal neurons display a transient complete suppression of activity. Note that according to Table 2 with the parameters of the simulations described in the paper, the mean (over directions) of the switch-off-related input is approximately 2 (for excitatory neurons) and 3 (for inhibitory neurons) times larger than the background input. Hence the switch-off input is larger than the background, but only moderately.

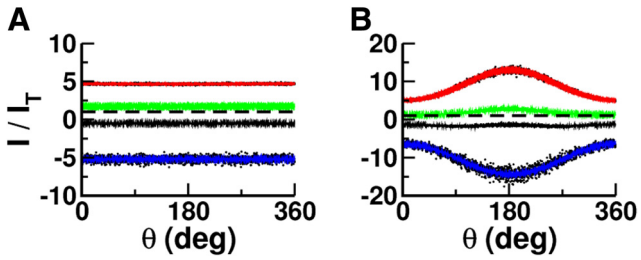

Figure 6. Spatial profiles of the time-averaged inputs into excitatory neurons demonstrating the balance of excitation and inhibition in the spatial WM model. ***A***, Baseline. ***B***, Delay period. Black curves, Top to bottom, Excitatory input, total input, and inhibitory input. Red, Excitatory input averaged with a square filter with a width of 50 neurons to smooth fast spatial fluctuations. Blue, Inhibitory input averaged with the same filter. Green, Total input plus 1.5 SDs of the total synaptic input fluctuations. All the inputs are normalized to the threshold (dashed line).

To show that there is a balance of excitation and inhibition during the precue as well as during the delay period, we computed the time average of each neuron's excitatory, inhibitory, and total synaptic inputs. The spatial profiles of these signals are depicted in Figure 6 for the excitatory neurons. During the precue period, these inputs display heterogeneities on a short spatial scale, due to the randomness of the connectivity. After averaging over these fluctuations, the profile of activity is essentially homogeneous. This is in contrast to the delay period, during which these inputs are modulated, in line with the fact that the mnemonic activity encodes the direction of the cue. Another difference between the two periods is that excitatory and inhibitory inputs into the neurons are larger during the delay than during the precue period. However, during both periods, inhibition approximately balances excitation and action potentials are driven by temporal fluctuations in the inputs (compare green line with threshold). Interestingly, the temporal mean of the total inputs tends to be lower during the delay than during the precue period. This is compatible with the elevation of the activity of a large fraction of the neurons during the delay period because the fluctuations also tend to increase. This is similar to what we found in our simulations of the unstructured network studied above (e.g., Fig. 3).

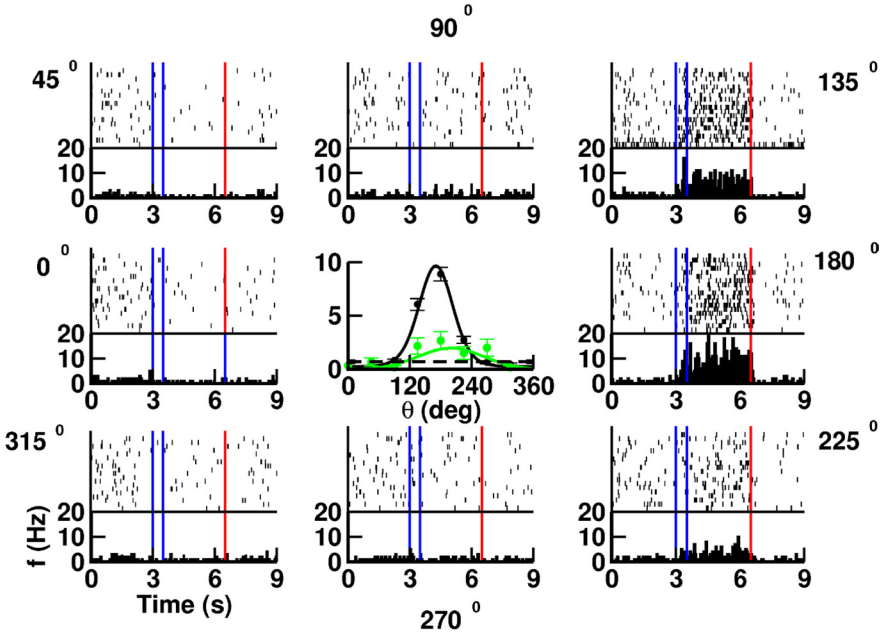

Figure 7. The activity of the neurons during the delay period is selective to the direction of the cue and the spike trains are highly irregular for preferred as well as for nonpreferred cue directions. Rasters and PSTH (bins, 50 ms) for one excitatory cell. Eight cue directions and 20 trials/cue direction. Vertical lines indicate the cue period (blue) and the end of the delay period (red). Center, tuning curves of the neuron for the delay (black; PD, 170°; TW, 38°) and cue (green; PD, 199°; TW, 68°) periods. Note that for this neuron the TWs are different but the PDs are similar. Error bars: SD estimated over the 20 trials. Dashed line, Baseline firing.

The activity of most of the neurons in our network is direction-selective during the cue period as well as during the delay. This is depicted in Figure 7 for one excitatory neuron. Note that its tuning curves during the two periods have slightly different preferred directions (PDs) and different tuning widths (TWs). The discharge pattern of the neuron at baseline is highly irregular ($CV = 0.9$). It is also strongly irregular during the delay whether the cue is presented at PD ($CV = 1.6$) or away from it ($CV = 1.3$). In Figure 8 we show the distribution of CV over the whole population in the precue and the delay periods. For the delay period the CVs were computed separately for preferred (Fig. 8, left) and nonpreferred directions (Fig. 8, right), defined as those directions for which the trial-averaged firing rate was higher (preferred directions) or lower (nonpreferred directions) than the activity of that neuron averaged both over trials and over the eight directions (Compte et al., 2003). These results show that all the neurons in both periods fire in a very irregular manner and that, during the delay, the firing is more irregular when the cue was presented at preferred directions. We also investigated whether slow firing rate nonstationarities can account for part of the irregularity. To do that we computed CV_2, which takes into account difference between adjacent interspike intervals (see Materials and Methods). We found (Fig. 8, dashed lines) that the behavior is now much more similar to a Poisson process. This is also compatible with the results found in Compte et al. (2003). In Figure 8 we also show CV in baseline period versus CV in delay period for all the neurons in the network. We can see that there is no correlation between the two values.

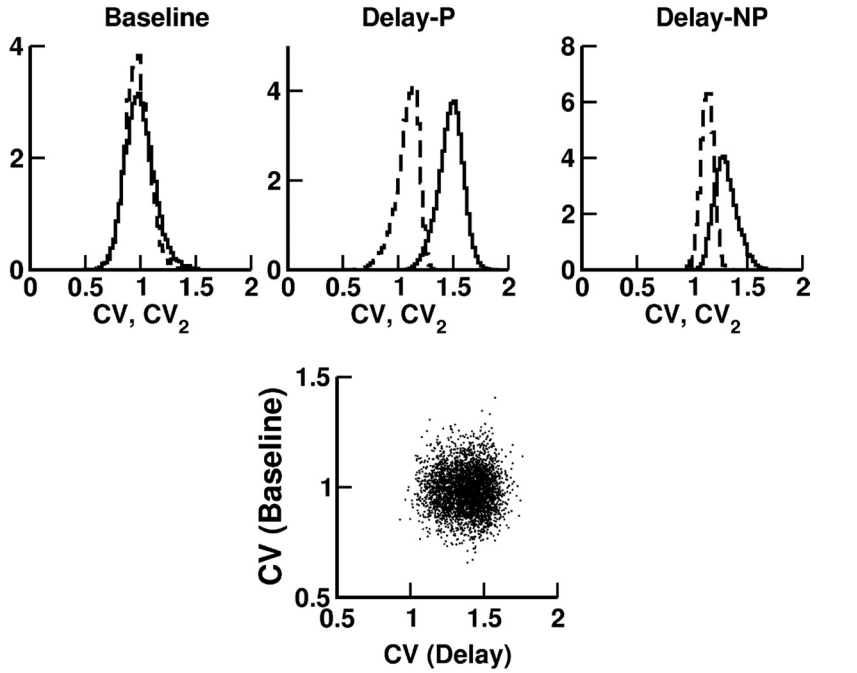

Figure 8. Top, Histograms of the *CV* (solid line) and CV_2 (dashed line) for activities during fixation (baseline), and delay periods. Delay-P, preferred directions. Delay-NP, nonpreferred directions. (see Materials and Methods). For each neuron, *CV* and CV_2 were computed from 20 trials per cue direction. Only directions with average firing rate >2 Hz are included. Bottom, *CV* in baseline versus *CV* in delay for all the neurons.

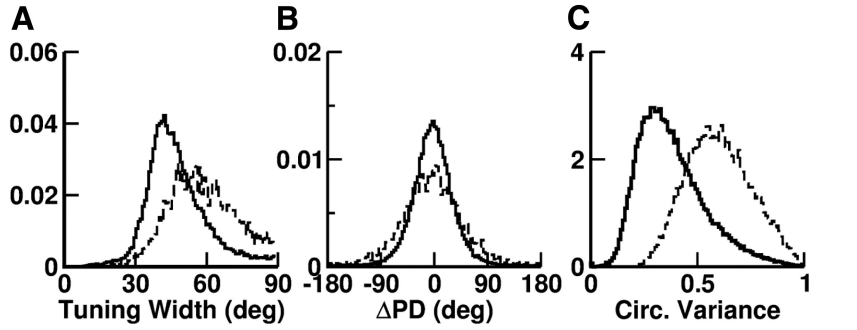

Figure 9. Diversity in the tuning curves. ***A***, Histogram of the tuning width for neurons with tuning curves well fitted with a von Mises function. Solid, Excitatory neurons (TW, 48 ± 14°). Dashed, Inhibitory neurons (TW, 61 ± 16°). ***B***, Histograms of the difference $\Delta PD = PD - \theta$ between the PD of a neuron and its location in the network for these neurons. ***C***, Histograms of the circular variance for all the neurons (solid, excitatory neurons; dashed, inhibitory neurons).

The selectivity properties and the dynamics of the delay activity are diverse

The neurons in our model display a diversity of selectivity properties. Although most neurons (98%) have selective delay activity (selectivity evaluated by the bootstrap method), only half have tuning curves well fitted to a von Mises function (Eq. 13). Moreover the width of the tuning curves of these neurons are broadly distributed (Fig. 9*A*) and the preferred directions are diverse even for nearby neurons (Fig. 9*B*). Neurons with tuning curves badly fitted to a von Mises function also display a broad dispersion in the degree of selectivity as quantified by the circular variance (Mardia, 1972) (Fig. 9*C*). Other aspects of the diversity in tuning curves are depicted in Figure 10*A*. Note in particular that, depending on the neuron, for a cue that is opposite to the preferred direction, the delay activity can be suppressed (also Fig. 7), similar to, or enhanced compared with baseline (Fig. 10*A*, black lines).

The tuning curves of the responses during the cue period are also diverse (Fig. 10*A*, green lines). For instance, for some neurons the optimal response is larger for the cue than for the delay whereas for others the reverse is observed. The simulations also revealed strong correlations between the preferred direction of the delay and the cue responses. In contrast, the tuning widths in the two epochs are only weakly correlated (Fig. 10*B*). Note that in Fig. 10*B* (left) one can see eight faint horizontal stripes associated with the eight equally spaced directions used to evaluate the tuning curves. This is because for very narrow tuning curves, the sampling of the directions is too coarse to obtain a precise estimate of the preferred directions. This effect involves <0.1% of the neurons.

Figure 11 plots the poststimulus time histograms (PSTHs) of the activity for several neurons. It shows that the firing rate dynamics of the neurons during the task are also diverse. In particular, although for many neurons activity remains essentially constant during the delay period (Figs. 1, 3, 11*C*), for others it ramps up (Figs. 2, 4, 11*C*). Some neurons display a phasic response to the cue (Figs. 1, 2, 11*C*), whereas others do not (Figs. 3, 4, 11*C*). In fact, visual inspection of the PSTH for a sample of 200 neurons (data not shown) indicates a phasic cue response at the preferred direction for approximately half of them.

This diversity is also an outcome of the balanced regime in which the network is operating. Because of the randomness in connectivity, the excitatory and the inhibitory inputs fluctuate spatially. Although the spatial fluctuations are much smaller than the spatial average for each of these inputs, they are comparable in size in the total inputs. As a result, the spatial fluctuations in connectivity substantially affect the discharge of the neurons (van Vreeswijk and Sompolinsky, 1996, 1998, 2004) inducing diversity in single neuron activity properties.

The network dynamics are multistable in a broad range of the background inputs

We assessed the robustness of multistability with respect to changes in the background inputs. Figure 12 depicts the bifurcation diagram for the network dynamic states when the background input is varied. This was done by running the dynamics of the network (with the parameters of Table 1) while changing I_E^b (keeping $I_I^b = I_E^b/2$) very slowly. The network was initialized in the low activity state with a small value of I_E^b just above the threshold current of the excitatory neurons, $I_E^b \approx 20$ mV . Figure 12 plots the hysteresis behavior of the spatial average (A) and the spatial modulation (B) of the network activity, as defined in the Materials and Methods section, as a function of I_E^b. By increas-

ing I_E^b slowly, the network remains in a state of low activity up to a value of $I_E^b \approx 80$ mV, beyond which this state no longer exists. The network then settles in a state of elevated activity where it remains while I_E^b keeps increasing. At $I_E^b = 100$ mV the direction of the changes of I_E^b is reversed and it starts to decrease. The network now tracks the state of elevated activity until $I_E^b \approx 24$ mV, when it ceases to exist. This shows that in a broad range of background inputs the network displays multistability between states that differ by their level of activity and by their spatial profile.

Selectivity depends on the range of the inhibition

Importantly, the multistability of the network and the selectivity of the neurons are robust to changes in the connectivity K. This is shown in Figure 13*A*, which plots the distribution of the circular variance for three different values of the connectivity ($K = 2000$, 4000, and 8000). It shows that for larger connectivity the neurons tend to be more selective, but that this effect saturates for large values of K.

Figure 14 plots the bifurcation diagram for different values of the spatial range of the inhibitory interactions. It shows that, even though the range of the inhibition is shorter than the range of the excitation, the network displays multistability. It also shows that the domain of multistability is larger when the inhibition is broader. It can also be seen that in the multistable regime, the spatial average firing rates in the baseline as well as in persistent states depend only weakly on the range of the inhibition (Fig. 14, left). The dependency of the spatial modulation of the activity profile in the persistent state is more pronounced: for broader inhibition the modulation is larger (Fig. 14, right). This corresponds to an increase in the degree of direction selectivity of the neurons during the delay period when inhibition is broader. This is shown in Figure 13*B* and in Table 3. The fraction of neurons with good fit to a circular variance, also given in Table 3, is also sensitive to the range of the inhibitory interactions. The broader the inhibition, the smaller this fraction becomes.

These results indicate that these parameters (range of inhibition and connectivity) can be varied in a very broad range and the network still displays selective persistent activity.

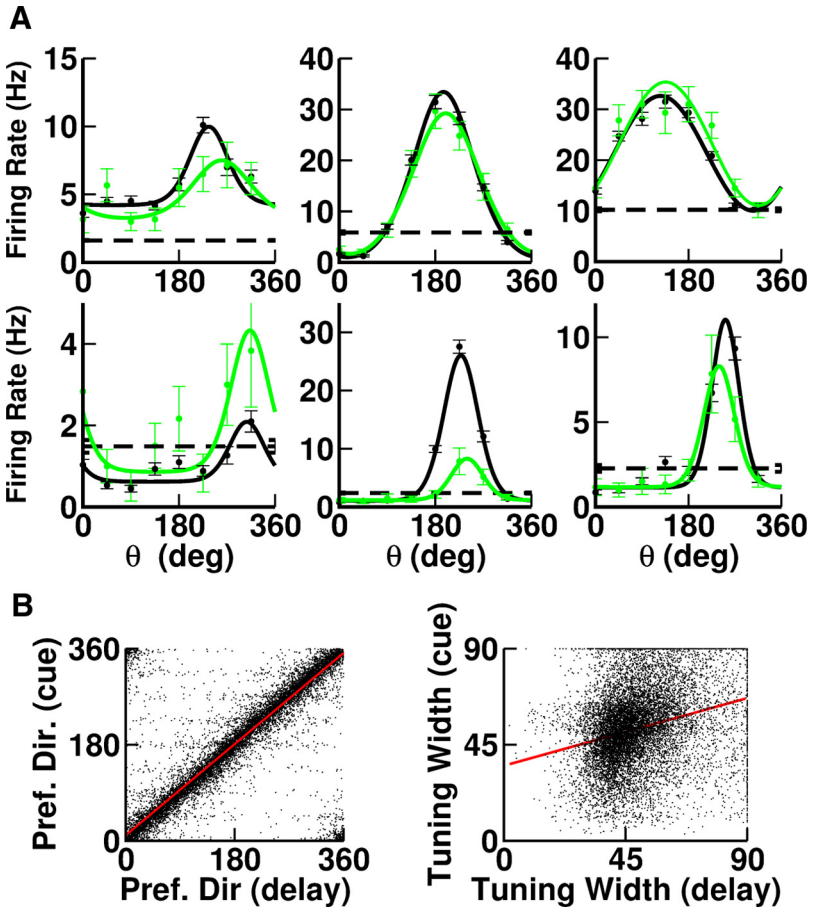

Figure 10. ***A***, Diversity in the shapes of the tuning curves. Black, Delay. Green, Cue. Dashed line, Baseline average firing rate. All the neurons are excitatory except for the last neurons in the second line, which are inhibitory. ***B***, Comparison of the tuning properties of the neurons during the cue and delay periods. Left, The preferred directions are strongly correlated ($R^2 = 0.94$). Right, The tuning widths are weakly correlated ($R^2 = 0.02$). Only neurons with tuning curves well fitted with von Mises functions for the two periods are included.

Encoding of the cue direction in the position of the activity bump

In theory, the existence of a continuous set of persistent attractors is necessary to maintain memory of the cue direction in the network. This continuity is destroyed by very small spatial heterogeneities in the connectivity or in the intrinsic properties of the neurons. In the presence of such heterogeneities, the network state during the delay period drifts toward a discrete attractor of the dynamics that is only very weakly correlated with the cue position (Tsodyks and Sejnowski, 1995; Zhang, 1996; Seung et al., 2000; Renart et al., 2003). Hence, the memory trace encoded in the location of the bump of network activity fades during the delay period at a rate that depends on the velocity of this drift. If the drift is too fast, this trace cannot be conserved for the duration of the delay and the selectivity of the neurons to the cue direction will be impaired. However, if the drift is sufficiently slow, the network can still function properly to encode the position of the cue, provided the delay duration is not overly lengthy.

In our network model, the connectivity is random and hence is heterogeneous. As a result, the dynamics of the network possesses only a small number of attractors, as shown in Figure 15*A* (left). The eight trajectories of the bump plotted in that figure correspond to different directions of presentation of the stimulus during the cue period. After a relatively long time all the trajectories converge toward one of two possible locations. In other words, there are only two persistent attractors. However, if the delay period is not too long, the position of the cue is still strongly correlated at the end of this period (Fig. 15*A*, right).

In fact, the accuracy with which the location of the cue can be memorized decreases with the duration of the delay period. To estimate the rapidity of the memory degradation, we simulated $N_r = 100$ realizations of the network for delay periods of 15 s. For each realization we computed, as a function of time, the location of the bump of activity, $\psi_k(t)$ ($k = 1, \ldots, N_r$), as explained in

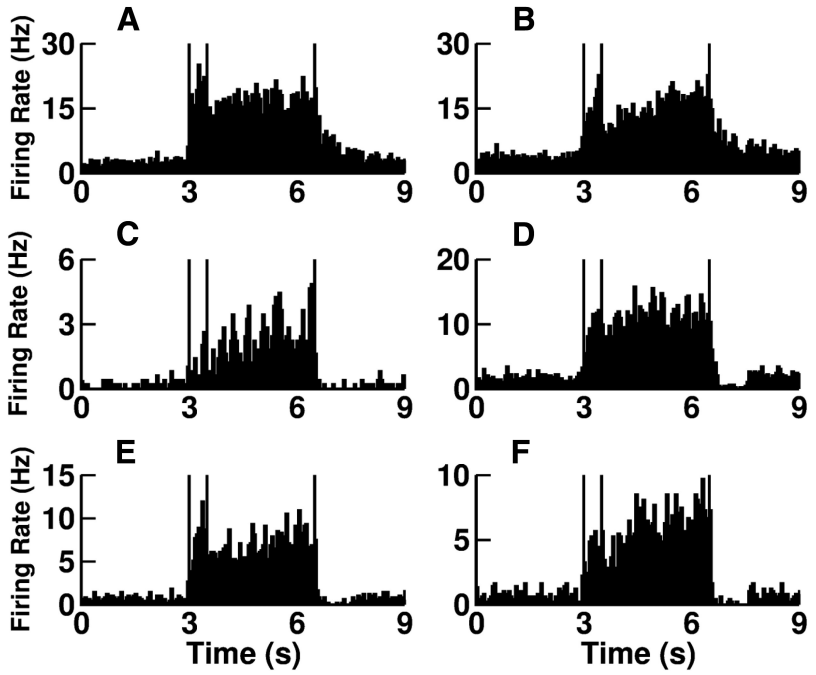

Figure 11. Diversity in the firing rate dynamics during cue, delay, and response periods. Poststimulus histograms (100 trials included) for five neurons. ***A–D***, Different excitatory neurons for a cue direction at their PDs. ***E***, ***F***, One excitatory neuron with PD = 161° and a cue at 135° (***E***) and at 225° (***F***).

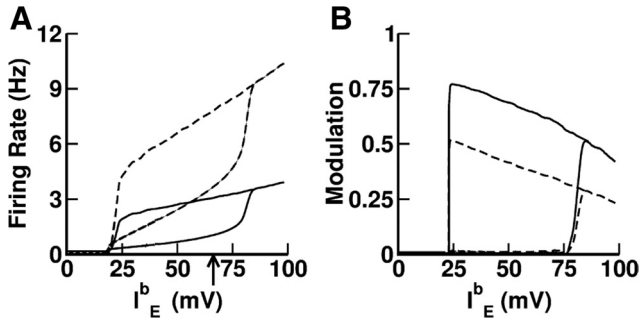

Figure 12. Bifurcation diagram as I_E^b is varied. ***A***, Spatial average activities of excitatory neurons (solid) and inhibitory (dashed). ***B***, Spatial modulation of the activities of the two populations. All parameters are given in Table 1. The arrow indicates the value of the current used in simulations.

Materials and Methods. Defining $\delta\psi_k(t) = \psi_k(t) - \psi_k(0)$ ($t = 0$ at the beginning of the delay period), we evaluated:

$$\Delta(t) = \frac{1}{N}\sum_{k=1}^{N_r} \delta\Psi_k(t), \tag{25}$$

and:

$$\sum(t) = \sqrt{\frac{1}{N}\sum_{k=1}^{N_r} \delta\Psi_k(t)^2}. \tag{26}$$

Clearly $\Delta(t)$ is a small quantity on the order of $1/\sqrt{N_r}$. This is because, after averaging over the realizations of the connectivity, the network displays rotational symmetry. This is in contrast with $\Sigma(t)$, which does not vanish even though N_r is large, as depicted in Figure 15*B*. At short time, the drift is dominated by the effect of the fast noise generated by the network dynamics in the balanced state. Therefore, it is a diffusion process and $\Sigma(t) \propto \sqrt{t}$. For very long times, the dynamics converge toward one of the attractors therefore $\Sigma(t)$ saturates. For intermediate values of t, the drift is dominated by the effect of the heterogeneity in the connectivity. It takes the form of a directed random walk and $\Sigma(t)$ increases linearly with time: $\Sigma(t) \approx Vt$. The rate of this increase, V, is an estimate of the drift velocity and thus of the rapidity of memory deterioration during the delay period. For the parameters of Table 1, we find a drift velocity of approximately 1.8°/s (Fig. 15*B*). Hence, the typical error in encoding the direction of the cue in the memory trace in this network is not >6° if the delay period is 3 s in duration.

The drift velocity depends on the network size, N, and on its connectivity, K. This is depicted in Figure 16, which plots $\log V$ as a function of $\log N$ for two values of K. The best linear fit of the data points reveals that the slope is very close to ½ for these two cases. This means that the velocity of the drift scales is $V \propto 1/\sqrt{N}$ with a prefactor that decreases with the connectivity. This prefactor also depends on other parameters of the network. For instance, we found that it increases with the background input and therefore with the average activity of the network in the persistent state.

The scaling of the drift velocity can be understood in terms of the fluctuations in the neuronal input. For a bump that involves a finite fraction of the network [i.e., a number of neurons which is $O(N)$] and if the fluctuations are uncorrelated, the drift velocity should scale as the fluctuation of the input on the individual neuron divided by $\sqrt{N}$. In Zhang (1996) these fluctuations come from perturbations in connections on the order of $1/\sqrt{N}$. This means that the fluctuation in the total input for a given neuron is on the order of $1/\sqrt{N}$ and the drift velocity is $O(1/\sqrt{N})$. In Renart et al. (2003) the fluctuations come from intrinsic heterogeneity and are order 1, so the drift velocity will be $O(1/\sqrt{N})$. In our case we have $O(K)$ inputs into each neuron, each one of them scaling as $1/\sqrt{K}$[. Therefore the total fluctuations per neuron are on the order of 1 and the drift velocity scales as $O(1\ \sqrt{N})$. The prefactor should remain finite in the limit of very large K because the quenched fluctuations do not vanish in the balanced regime. Proving rigorously this conjecture is an interesting problem that deserves further research.

Itskov et al. (2011) have recently studied a ring model of visuospatial WM with rate-based neuronal dynamics in which the selective delay activity was generated by the nonlinearity of the neurons. In the presence of heterogeneities the dynamics had only a small number of attractors. However, they showed that short-term facilitation in the recurrent interactions can slow down the drift of the bump dramatically. This is because it selectively amplifies synapses from neurons that have been activated by the cue, which tends to pin down the bump at its initial position. A similar effect explains the slowness of the drift in our PFC model. Therefore, facilitation in the recurrent excitatory interactions plays two roles in our mechanism: (1) it induces nonlinearities that sustain the persistence of the

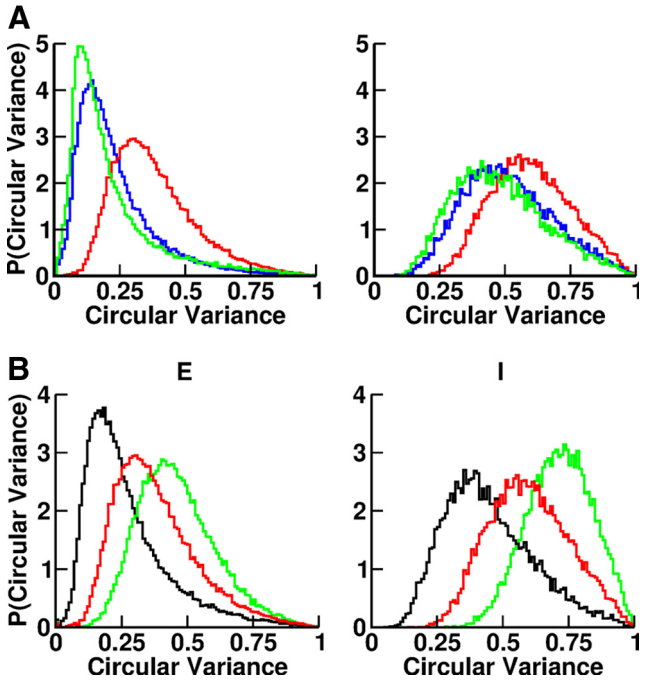

Figure 13. ***A***, Dependence of the width of the tuning curves on the network connectivity. Left, Distribution of the circular variance for the excitatory neurons. Right, Distribution of the circular variance for the inhibitory neurons. $K = 2000$ (red), $K = 4000$ (blue), $K = 8000$ (green). Other parameters as in Tables 1 and 2. ***B***, Dependence of the width of the tuning curves on the range of inhibition. Left, Distribution of circular variance for the excitatory neurons. Right, Distribution of the circular variance for the inhibitory neurons. $\sigma_{II} = \sigma_{EI} = 60°$ (red), $\sigma_{II} = \sigma_{EI} = 80°$ (black), and $\sigma_{II} = \sigma_{EI} = 40°$ (green). All the other parameters (but σ_{II}, σ_{EI}) are given in Table 1.

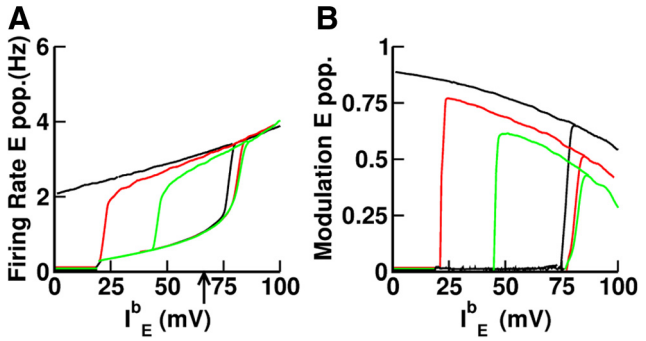

Figure 14. Dependence of the bifurcation diagram on the range of inhibition. ***A***, Spatial average activity of the excitatory neurons for the following: $\sigma_{II} = \sigma_{EI} = 60°$ (red), $\sigma_{II} = \sigma_{EI} = 80°$ (black), and $\sigma_{II} = \sigma_{EI} = 40°$ (green). ***B***, Spatial modulation of the activity of the excitatory neurons. All parameters (but σ_{II}, σ_{EI}) are given in Tables 1 and 2.

Table 3. Tuning width and percentage of neurons with good tuning, $f_{q>0.001}$, as a function of the width of the inhibitory interactions, $\sigma = \sigma_{II} = \sigma_{EI}$

σ	TW_E	TW_I	$f_{q>0.001}$
40°	50 ± 13°	60 ± 14°	75%
60°	48 ± 14°	61 ± 16°	44%
80°	38 ± 13°	49 ± 16°	23%

activity and (2) it slows down the degradation of the memory trace.

In our PFC model, the spatial fluctuations in the connectivity give rise to strong heterogeneities. As a result, the bump of activity that should encode the direction of the cue drifts during the delay period toward a location that is weakly correlated with the stimulus. This leads to an eventual loss of memory. However, we found that the drift is slow. It is on the order of 2°/s in a network with 80,000 neurons and connectivity $K = 2000$ and even slower for a larger network. For plausible size and connectivity one can easily obtain a drift of 0.5–1°/s, which is in line with experimental data (White et al., 1994, Ploner et al., 1998).

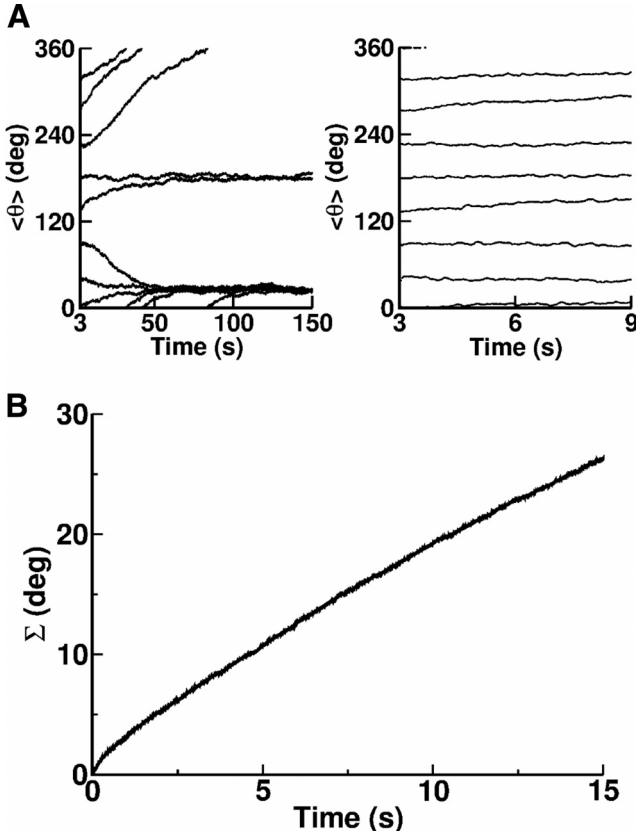

Figure 15. ***A***, The direction of the population vector estimated from the activity of the excitatory neurons versus time. All parameters are as in the reference set. At that time the cue is presented at $t = 3$ s for a duration of 500 ms in one of the eight directions. From bottom to top (in degrees): 0, 45, 90, 135, 180, 225, 270, 315. The total simulation time is 150 s (left) and 10 s (right). ***B***, $\Sigma(t)$ (in degrees) versus time (see definition in Materials and Methods). The size of the network is $N_E = 64{,}000$, $N_I = 16{,}000$. The averaging was performed over $N_r = 100$ realizations of the network. All parameters are as in the reference set.

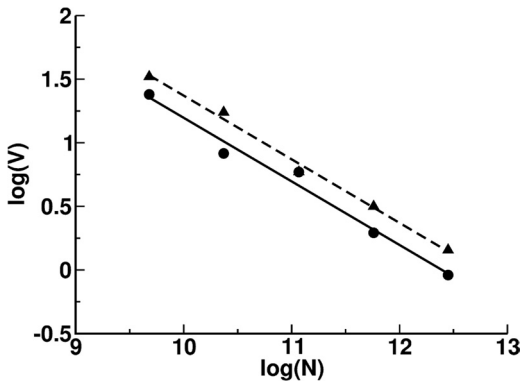

Figure 16. The drift velocity of the bump versus the network size. Circles, Average connectivity as in the reference set. Triangles, Average connectivity smaller by a factor of 2. All other parameters are as in the reference set. Lines correspond to the best linear fit: $y = a_0 + a_1 x$. Solid: $a_0 = 6.193$, $a_1 = -0.499$; correlation coefficient, 0.991. Dashed: $a_0 = 6.372$, $a_1 = -0.5$; correlation coefficient, 0.996.

Discussion

Experimental data support the hypothesis that the cortex operates in states in which excitation is balanced by inhibition (Destexhe et al., 2003; Shu et al., 2003; Haider et al., 2006). Modeling studies (van Vreeswijk et al., 1996, 1998; Lerchner et al., 2004, Vogels et al., 2005; Vogels and Abbott, 2009; Hansel and van Vreeswijk, 2012) have argued that this can explain in a natural way why neurons *in vivo* fire so irregularly in spontaneous as well as in sensory-evoked activity. In the present work, we argued that balance of excitation and inhibition can also explain the high irregularity of neuronal firing in persistent activity but that this requires synaptic nonlinearities. This is because, if the synaptic interactions are linear, neuronal nonlinearities wash out at the population level in balanced states, thus precluding more than one balanced state.

Our first result is that in an unstructured network, nonlinearities induced by short-term facilitation in recurrent excitation are appropriate to generate bistability between balanced states in a very robust way. We then demonstrated that in a network model of PFC, these nonlinearities can also sustain the selectivity of the delay activity as recorded in PFC during ODR tasks. Interestingly, we found in our simulations that the mean input into a neuron is more hyperpolarizing during the delay period than in baseline. Concomitantly, the temporal fluctuations in the input also increase to guarantee that the activity is larger during the delay. This explains why in our model the spike trains are typically more irregular during the delay period than during baseline, in qualitative agreement with the experimental results of Compte et al. (2003). Similar features were found in the mean-field theory of the integrate-and-fire network with stochasticity and short-term plasticity in neuronal interactions studied by Mongillo et al. (2012). However, this effect may be model and parameter dependent and more modeling as well as experimental works are required to further probe its significance.

Another important feature of our PFC model is the diversity across neurons in the neuronal tuning curves and in the dynamics. This occurs although all the neurons (in a given population) are identical. In fact this is another hallmark of the balanced regime in which the network operates. Such diversity is also observed in experimental data (Funahashi et al., 1989, 1990; Takeda and Funahashi, 2007). Remarkably, in our model, the preferred directions of the neurons during the cue and during the delay periods are highly correlated whereas the tuning widths are very weakly correlated in agreement with experimental reports (Funahashi et al., 1990).

An essential property of the mechanism we have proposed is its robustness. Even if the average connectivity, K, is very large, the unstructured network as well as our PFC model display multistability of balanced states in which the activity is driven by temporal fluctuations in synaptic inputs. The temporal irregularities and the heterogeneities in the neuronal firing are very robust features that depend only weakly on K for very large K. This agrees with the work of Mongillo et al. (2012).

This contrasts with what happens in networks with linear synaptic interactions. In this case a coexistence of several states in which the neuronal firing is driven by the temporal fluctuations rather than by the mean inputs can in theory be achieved. However, this requires an increasingly precise tuning of the parameters as the connectivity increases (Renart et al., 2007). This is because the response of populations to external inputs becomes more linear as the connectivity increases. As a result, accounting for both spatiotemporal irregularity and mnemonic activity in models with linear synaptic interactions requires fine-tuning of the network size and connectivity (Renart et al., 2007), the coding level (van Vreeswijk and Sompolinsky, 2004; Lundqvist et al., 2010), synaptic efficacies (Roudi and Latham, 2007), and on the level of fast noise (Barbieri and Brunel, 2008). Durstewitz and Gabriel (2007) argued that a network with voltage-dependent NMDA synapses can significantly display enhanced variability when it operated in a chaotic close-to-bifurcation regime. This presumably also needs fine-tuning of the parameters in the large size limit to insure the right membrane potential distribution.

Facilitation has been found *in vitro* in synapses in a population of pyramidal neurons in PFC (Hempel et al., 2000; Wang et al., 2006). Wang et al. (2006) fitted the dynamics of these synapses to the same model of STP that we used here (Tsodyks and Markram, 1997) and found a broad diversity for parameters τ_f, τ_r, and U. The conditions under which STP gives rise to multistablity in balanced networks have been already investigated by Mongillo et al. (2012) in the infinite connectivity limit for a network of integrate-and-fire neurons. The conditions we found in our models are qualitatively similar. In the case of the rate model (Eqs. 21, 22), they can be summarized as follows. On the one hand, the effective synaptic strength $[F_{EE}(f_E)]$ must be nonmonotonic (i.e., synaptic transmission must be facilitating at low firing rates). This can be obtained if U is sufficiently small and $\tau_r < \tau_f$. If U increases, the region displaying facilitation becomes smaller, but that can be compensated by decreasing τ_r. Another necessary condition is that the maximum value of F_{EE} must be larger than $1/G$ (Eq. 24). This quantity depends on the couplings but it can be made small enough by taking a large value of G_{EE}. Our simulations of the integrate-and-fire model agree with those conditions. They show that the range of the background input in which our PFC model displays persistent selective activity decreases when U increases when all other parameters are kept constant. However, if one also changes G_{EE}, keeping UG_{EE} constant (or, alternatively, decreasing τ_r), there is a selective delay activity for U as large as 0.12. The STP parameters τ_f, τ_r can be changed by $\pm 20\%$ without losing multistability even if all the other parameters are kept constant. These numbers are similar to those reported in the phase diagrams computed by Mongillo et al. (2012). However, if the STP parameters are changed too much, the network model loses its selective delay activity. This is compatible with recent experimental studies that have identified altered STP in PFC circuits as a neural substrate underlying impairment in WM (Fénelon et al., 2011; Arguello and Gogos, 2012).

Mongillo et al. (2008) proposed a mechanism for WM based on STP that differs from the one we have investigated here. It does not require reverberating activity. It relies on the fact that synapses displaying STP keep a transient mark of changes in network activity over time scales of several hundred milliseconds. Mongillo et al. (2008) argued that the mark left by the cue period activity can underlie WM for a duration of <1 s. Similarly, in our mechanism, there should be a mark of persistent activity outlasting the end of the delay period. This prediction can be examined experimentally (e.g., by comparing the activity of the neurons in the postsaccadic and precue periods in an ODR task). STP dynamics can be also expressed in the gradual building up of the activity at the beginning of the delay period. This should be observed preferentially in neurons for which the phasic response during the cue period is not too strong. This effect can be seen in the examples of Figure 11*C*,*F*. Another interesting prediction of our model is that the strong heterogeneity in the neuronal properties manifests itself in an uncorrelated way

between the different periods. For instance, the CV values between the two periods do not correlate (Fig. 8). Similarly, no correlation can be found for tuning widths between cue and delay periods (Fig. 11).

To conclude, the highly irregular activity observed both during fixation and delay periods of WM tasks challenges network mechanisms of WM. We argued in favor of a new mechanism in which nonlinearities in synaptic interactions play a central role in sustaining multistability in neocortical networks. We claimed that the short-term facilitation observed in populations of pyramidal neurons in PFC is a plausible physiological substrate for these nonlinearities. We illustrated this in a model of visuospatial WM that accounts for the first time in a comprehensive and robust way for the observed persistence selectivity, temporal irregularity, and diversity of delay activity of neurons in PFC during ODR tasks. This framework is general and can also be applied to other WM tasks.

References

Amit DJ, Brunel N (1997) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cereb Cortex 7:237–252. CrossRef Medline

Amit DJ (1995) Modeling brain functions: the world of attractor neural networks. Cambridge, UK: Cambridge UP.

Arguello PA, Gogos JA (2012) Genetic and cognitive windows into circuit mechanisms of psychiatric disease. Trends Neurosci 35:3–13. CrossRef Medline

Asaad WF, Rainer G, Miller EK (1998) Neural activity in the primate prefrontal cortex during associative learning. Neuron 21:1399–1407. CrossRef Medline

Baddeley AD (1986) Working memory. Oxford, UK: Oxford UP.

Bair W, Koch C, Newsome W, Britten K (1994) Power spectrum analysis of bursting cells in area MT in the behaving monkey. J Neurosci 14:2870–2892. Medline

Barbieri F, Brunel N (2008) Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? Front Neurosci 2:114–122. CrossRef Medline

Bartos M, Vida I, Frotscher M, Geiger JR, Jonas P (2001) Rapid signaling at inhibitory synapses in a dentate gyrus interneuron network. J Neurosci 21:2687–2698. Medline

Bartos M, Vida I, Frotscher M, Meyer A, Monyer H, Geiger JR, Jonas P (2002) Fast synaptic inhibition promotes synchronized gamma oscillations in hippocampal interneuron networks. Proc Natl Acad Sci U S A 99:13222–13227. CrossRef Medline

Ben-Yishai R, Lev Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. Proc Natl Acad Sci U S A 92:3844–3848. CrossRef Medline

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

Brunel N (2000) Persistent activity and the single-cell frequency-current curve in a cortical network model. Network 11:261–280. CrossRef Medline

Burns BD, Webb AC (1976) The spontaneous activity of neurones in the cat's cerebral cortex. Proc R Soc Lond B Biol Sci 194:211–223. CrossRef Medline

Camperi M, Wang XJ (1998) A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. J Comput Neurosci 5:383–405. CrossRef Medline

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

Compte A, Constantinidis C, Tegner J, Raghavachari S, Chafee MV, Goldman-Rakic PS, Wang, XJ (2003) Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. J Neurophysiol 90:3441–3454. CrossRef Medline

Connors BW, Gutnick MJ, Prince DA (1982) Electrophysiological properties of neocortical neurons in vitro. J Neurophysiol 48:1302–1320. Medline

Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001) The sensory nature of mnemonic representation in the primate prefrontal cortex. Nat Neurosci 4:311–316. CrossRef Medline

Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT.

Destexhe A, Rudolph M, Paré D (2003) The high-conductance state of neocortical neurons in vivo. Nat Rev Neurosci 4:739–751. CrossRef Medline

Durstewitz D, Gabriel T (2007) Dynamical basis of irregular spiking in NMDA-driven prefrontal cortex neurons. Cereb Cortex 17:894–908. Medline

Fénelon K, Mukai J, Xu B, Hsu PK, Drew LJ, Karayiorgou M, Fischbach GD, Macdermott AB, Gogos JA (2011) Deciency of Dgcr8, a gene disrupted by the 22q11.2 microdeletion, results in altered short-term plasticity in the prefrontal cortex. Proc Natl Acad Sci U S A 108:4447–4452. CrossRef Medline

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

Funahashi S, Bruce CJ, Goldman-Rakic PS (1990) Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. J Neurophysiol 63:814–831. Medline

Funahashi S, Bruce CJ, Goldman-Rakic PS (1991) Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 65:1464–1483. Medline

Fuster JM (2008) The prefrontal cortex. Fourth edition. London, UK: Elsevier.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654. CrossRef Medline

Goldman-Rakic PS (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Handbook of physiology. The nervous system. Higher functions of the brain, pp 373–417. Bethesda, MD: American Physiological Society.

Goldman-Rakic PS (1988) Topography of cognition. Parallel distributed networks in primate association cortex. Annu Rev Neurosci 11:137–156. CrossRef Medline

Goldman-Rakic PS (1995) Cellular basis of working memory. Neuron 14:477–485. CrossRef Medline

Haider B, Duque A, Hasenstaub AR, McCormick DA (2006) Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. J Neurosci 26:4535–4545. CrossRef Medline

Hansel D, Mato G (2008) Balanced excitation and inhibition and short-term plasticity: a new paradigm for working memory. Abstract Soc Neurosci. Washington, 2008.

Hansel D, Sompolinsky H (1996) Chaos and synchrony in a model of a hypercolumn in visual cortex. J Comput Neurosci 3:7–34. CrossRef Medline

Hansel D, Sompolinsky H (1998) Modeling feature selectivity in local cortical circuits. In: Methods in neuronal modeling. From synapses to networks. Second edition (Koch C, Segev I, eds). Cambridge, MA: MIT.

Hansel D, van Vreeswijk C (2012) The mechanism of orientation selectivity in primary visual cortex without a functional map. J Neurosci 32:4049–4064. CrossRef Medline

Hansel D, Mato G, Meunier C, Neltner L (1998) On numerical simulations of integrate-and-fire neural networks. Neural Comput 10:467–483. CrossRef Medline

Hebb DO (1949) The organization of behavior. New York: Wiley.

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

Hertz J (2010) Cross-correlations in high-conductance states of a model cortical network. Neural Comput 22:427–447. Medline

Holt GR, Softky WR, Koch C, Douglas RJ (1996) Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. J Neurophysiol 75:1806–1814. Medline

Hopfield JJ (1984) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 81:3088–3092. Medline

Itskov V, Hansel D, Tsodyks M (2011) Short-term facilitation may stabilize parametric working memory trace. Front Comput Neurosci 5:1–19. Medline

Jahr CE, Stevens CF (1990) Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. J Neurosci 10: 3178:3182.

Lerchner A, Ahmadi M, Hertz J (2004) High-conductance states in a mean-field cortical network model. Neurocomputing 58:935–940. CrossRef
Lerchner A, Ursta C, Hertz J, Ahmadi M, Ruffiot P, Enemark S (2006) Response variability in balanced cortical networks. Neural Comput 18:634–659. CrossRef Medline
Lundqvist M, Compte A, Lansner A (2010) Bistable, irregular firing and population oscillations in a modular attractor memory network. PLos Comput Biol 6:e1000803. CrossRef Medline
Maimon G, Assad JA (2009) Beyond Poisson: increased spike-time regularity across primate parietal cortex. Neuron 62:426–440. CrossRef Medline
Mardia KV (1972) Statistics of directional data. London, UK: Academic.
Markram H, Wang Y, Tsodyks M (1998) Differential signaling via the same axon of neorcortical pyramidal neurons. Proc Natl Acad Sci U S A 95:5323–5328. CrossRef Medline
Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331:68–70. CrossRef Medline
Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319:1543–1546. CrossRef Medline
Mongillo G, Hansel D, van Vreeswijk C (2012) Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. Phys Rev Lett 108:158101. CrossRef Medline
Ploner CJ, Gaymard B, Rivaud S, Agid Y, Pierrot-Deseilligny C (1998) Temporal limits of spatial working memory in humans. Eur J Neurosci 10:794–797. CrossRef Medline
Press W, Vetterling W, Teukolsky S, Flannery B (1992) Numerical recipes in FORTRAN 77: the art of scientific computing. Cambridge, UK: Cambridge UP.
Rao SG, Williams GV, Goldman-Rakic PS (1999) Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization. J Neurophysiol 81:1903–1916. Medline
Renart A, Song P, Wang XJ (2003) Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. Neuron 38:473–485. CrossRef Medline
Renart A, Moreno-Bote R, Wang XJ, Parga N (2007) Mean-driven and fluctuation-driven persistent activity in recurrent networks. Neural Comput 19:1–46. CrossRef Medline
Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD (2010) The asynchronous state in cortical circuits. Science 327:587–590. CrossRef Medline
Romo R, Brody CD, Hernández A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. Nature 399:470–473. CrossRef Medline
Roudi Y, Latham PE (2007) A balanced memory network. PLoS Comput Biol 3:e141. CrossRef Medline
Seung HS, Lee DD, Reis BY, Tank DW (2000) Stability of the memory of eye position in a recurrent network of conductance-based model neurons. Neuron 26:259–271. CrossRef Medline
Shafi M, Zhou Y, Quintana J, Chow C, Fuster J, Bodner M (2007) Variability in neuronal activity in primate cortex during working memory tasks. Neuroscience 146:1082–1108. CrossRef Medline
Shinomoto S, Kim H, Shimokawa T, Matsuno N, Funahashi S, Shima K, Fujita I, Tamura H, Doi T, Kawano K, Inaba N, Fukushima K, Kurkin S, Kurata K, Taira M, Tsutsui K, Komatsu H, Ogawa T, Koida K, Tanji J, et al. (2009) Relating neuronal firing patterns to functional differentiation of cerebral cortex. PLoS Comput Biol 5:e1000433. CrossRef Medline
Shinomoto S, Sakai Y, Funahashi S (1999) The Ornstein-Uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex. Neural Comput 11:935–951. CrossRef Medline
Shu Y, Hasenstaub A, McCormick DA (2003) Turning on and off recurrent balanced cortical activity. Nature 423:288–293. CrossRef Medline
Softky WR, Koch C (1992) Cortical cells should fire regularly, but do not. Neural Comput 4:643–646. CrossRef
Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci 13:334–350. Medline
Somers DC, Nelson SB, Sur M (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. J Neurosci 15:5448–5465. Medline
Takeda K, Funahashi S (2002) Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. J Neurophysiol 87:567–588. Medline
Takeda K, Funahashi S (2007) Relationship between prefrontal task-related activity and information flow during spatial working memory performance. Cortex 43:38–52. CrossRef Medline
Thomson AM (1997) Activity-dependent properties of synaptic transmission at two classes of connections made by rat neocortical pyramidal axons in vitro. J Physiology 502:131–147. CrossRef Medline
Tsodyks MV, Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc Natl Acad Sci U S A 94:719–723. CrossRef Medline
Tsodyks M, Sejnowski T (1995) Rapid switching in balanced cortical network models. Network 6:1–14. CrossRef
Tsodyks M, Sejnowski T (1997) Associative memory and hippocampal place cells. Adv Neural Inf Process Syst 6:81–86.
van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274:1724–1726. CrossRef Medline
van Vreeswijk C, Sompolinsky H (1998) Chaotic balanced state in a model of cortical circuits. Neural Comput 10:1321–1371. CrossRef Medline
van Vreeswijk C, Sompolinsky H (2004) Irregular activity in large networks of neurons. In: Methods and models in neurophysics, lecture notes of the XXX Les Houches Summer School (Chow C, Gutkin B, Hansel D, Meunier C, Dalibard J, eds). London, UK: Elsevier Science and Technology.
Vogels TP, Abbott LF (2009) Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. Nat Neurosci 12:483–491. CrossRef Medline
Vogels TP, Rajan K, Abbott LF (2005) Neural network dynamics. Annual Rev Neurosci 28:357–376. CrossRef
Wang H, Stradtam GG 3rd, Wang XJ, Gao WJ (2008) A specialized NMDA receptor function in layer 5 recurrent microcircuitry of the adult rat prefrontal cortex. Proc Natl Acad Sci U S A 105:16791–16796. CrossRef Medline
Wang XJ (2001) Synaptic reverberations underlying mnemonic persistent activity. Trends Neurosci 24:455–463. Medline
Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS (2006) Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nat Neurosci 9:534–542. CrossRef Medline
White JM, Sparks DL, Stanford TR (1994) Saccades to remembered target locations: an analysis of systematic and variable errors. Vision Res 34:79–92. CrossRef Medline
Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. J Neurosci 16:2112–2126. Medline