Dynamic functional tuning of nonlinear cortical networks

Martin Stetter
Siemens AG, Corporate Technology, Information & Communications, D-81730 Munich, Germany
(Received 30 September 2005; revised manuscript received 28 December 2005; published 3 March 2006)

The mammalian neocortex is a highly complex and nonlinear dynamic system. One of its most prominent features is an omnipresent spontaneous neuronal activity. Here the possible functional role of this global background for cognitive flexibility is studied in a prototypic mean-field model area. It is demonstrated that the level of global background current efficiently controls the stimulus-response threshold and the stability and properties of short-term memory states. Moreover, it can dynamically gate arbitrary cortical subnetworks, when applied to parts of the area as a weak bias signal. These results suggest a central functional role of the level of background activation: the dynamic functional tuning of neocortical circuits.

DOI: 10.1103/PhysRevE.73.031903 PACS number(s): 87.18.Sn, 05.65.+b, 05.45.−a, 87.18.Bb

I. INTRODUCTION

The neocortex represents the most prominent brain structure of humans and higher mammals. Consisting of about 10^{11} densely interconnected neurons, it represents a highly complex and nonlinear recurrent system. The cortex can be divided into different cortical areas which share a remarkably uniform anatomical structure [1], but are involved in very different functions ranging from basic feature representation to complex associations and mental manipulations.

A general feature of the neocortex present in all cortical areas is a spontaneous background neuronal activity which in total strength usually dominates by far the sparse signal-related activations. During typical cortical operation, about 99% of the neurons are spontaneously active and fire action potentials with about 2–3 Hz, whereas only about 1% of the neurons process signals and fire with some tens of Hz [2–4]. Consequently, the global, noninformative background activity is about tenfold bigger—and consumes tenfold more metabolic energy—than the patterned information-carrying neuronal activity. Likewise, nonspecific input current to each neuron arising from the background synaptic activity is much bigger than specific (information-carrying) synaptic current.

Throughout the last years, a number of models based on realistic networks of spiking neurons [5–7] have been analyzed to explore how, in the visual domain, neural correlates of attention [8–10], context-dependent working memory [11–15], visuomotor association [16], and decision making [17] might arise from recurrent neuronal dynamics. For this class of networks, efficient mean-field formulations have been developed [5,18–21]. Many of these formulations have been found to behave rather similar to each other in terms of stationary spike rates and current-frequency relationships [22].

The observed dominance of the global background activity despite its high metabolic demand suggests that spontaneous activity might play a crucial role for the functional performance of neocortex. In fact, various functional roles for spontaneous activity have been proposed, such as linearizing input-output relationships [2,23], speeding up network responses to transient stimuli [23–25], and varying the stability of persistently active memory states [12]. More recently, the level of spontaneous activity has also been shown to be able to switch an integrate-and-fire neuronal network back and forth between different dynamic states [26].

Here, we study systematically how the level of the global background spontaneous activity affects functional properties of a mean-field model cortical area that are relevant in the context of visual cognition. The model is chosen to prototypically represent the most relevant features of a wide class of realistic mean-field descriptions (cf. Refs. [22,27] and References therein). At the same time, it is kept mathematically simple to allow for the extensive derivation of analytical results.

When local excitation is strong and is balanced by lateral inhibition, we find that the background current drastically affects the system's attractor landscape. When increased from below, the global background current first lowers the network's amplification threshold for small stimuli. As will be defined in more detail in Sec. III C, the amplification threshold is the minimum strength of a stimulus required to drive the stimulated pool's activity into the saturating part of its gain function. Then, a multistability regime appears, where patterned short-term memory states, defined by persistently active pools, co-exist with the state of constant activation across pools. A key finding is that the maximum number of co-existing persistently active neuronal pools systematically increases with the background level. Finally, the constant dc state becomes unstable leading to spontaneous formation of memory states, driven by fluctuations. When the level of global background current is varied selectively for a subnetwork, the functional properties of that subnetwork are changed selectively as well. This is demonstrated by a set of exemplifying simulations, where selective attention, storage in working memory, and the level of vulnerability of activity patterns can be gated by the level of additive biasing current. These findings shed light on the computational mechanisms underlying the recently proposed biased competition and cooperation framework of visual cognitive processing [8,10].

The presented results suggest a central functional role of the global background activity. By varying its level, the functional properties of a cortical area can be dynamically tuned: In the range of hundred milliseconds, its sensitivity to represent and/or its ability to store stimuli by means of persistent activity can be adjusted. This principle might also underly

the control of alertness by global neuromodulatory systems like the dopaminergic system.

II. MODEL

We consider a model cortical area that consists of excitatory and inhibitory neurons. Excitatory neurons of the model area are grouped to N pools that follow a Wilson-Cowan-type dynamics. They drive one common linear inhibitory pool, which in turn spreads inhibition to itself and to all excitatory pools. The excitatory and inhibitory pools are driven by common global background synaptic currents I_0 and I_I, respectively. They are assumed to arise from spontaneous global background activity of other parts of the brain, which are not explicitly modeled in this work, because self-consistent stability of a spontaneous active state with different mean rates has been demonstrated earlier [5]. In addition, each excitatory pool $k=1,\ldots,N$ can receive a small specific stimulating current $I_{s,k}$ from presynaptic neuron assemblies. The set of differential equations for the excitatory and inhibitory ensemble-averaged spike rates, ν_k and ν_I, are given by

$$\tau\frac{d\nu_k}{dt} = -\nu_k + g\left(I_0 + I_{s,k} + w_+\nu_k + \sum_{l\neq k}^{N} w_-\nu_l - w_{EI}\nu_I\right), \quad (1)$$

$$\tau_I\frac{d\nu_I}{dt} = -\nu_I + g_I\left(I_I + \sum_{l=1}^{N} w_{IE}\nu_l - w_{II}\nu_I\right), \quad (2)$$

where $w_+ \geqslant 0$ denotes the strong self-excitation of each excitatory pool, and $w_- \geqslant 0$ the weaker mean lateral excitation strength between pairs of excitatory pools. The lateral excitation strength might actually be structured and store complex patterns as attractors of the network dynamics [28], but for simplicity we assume a constant value w_- here giving rise to relatively simple attractors.

All excitatory pools drive a common inhibitory pool with weights $w_{IE} \geqslant 0$, which spreads inhibition back to itself and all excitatory pools with weights $-w_{II} \leqslant 0$ and $-w_{EI} \leqslant 0$, respectively. The inhibitory pool is kept linear which might be a reasonable first order approximation of observed behavior [29]. The nonlinear activation function $g(I)$ of excitatory pools is assumed differentiable and to consist of an accelerating and a saturating part. For quantitative analysis and simulations, we adopt the easy-to-handle expression $g(I)=0, I<0; =1/2I^2, 0\leqslant I\leqslant 1; =\sqrt{I-3/4}, I>1$ [27], which reasonably well approximates the fast acceleration and slow saturation of current-frequency relationships recently derived for various realistic network models of spiking neurons [22]. It is scaled to have unity maximum slope for convenience. Note, however, that all results of the current study hold for arbitrary functions with sigmoid characteristics. Finally, g_I denotes the gain of the linear inhibitory pool.

Under the adiabatic approximation of the inhibitory population, which is known to respond fast in the cortex [29], Eqs. (1) and (2) can be simplified to

$$\tau\frac{d\nu_k}{dt} = -\nu_k + g\left(I_b + I_{s,k} + W_S\nu_k + W_L/N\sum_{l=1}^{N}\nu_l\right), \quad k=1,\ldots,N, \quad (3)$$

where the net self-excitation $W_S := w_+ - w_- \geqslant 0$, the net lateral excitation $W_L := N(w_- - W_I)$, and the effective global background current $I_b := I_0 - W_I I_I / w_{IE}$ have been introduced. The effective inhibition is given by $W_I := g_I w_{EI} w_{IE}/(1+g_I w_{II})$. The equivalent system Eq. (3) is fully characterized by W_S, W_L, and I_b.

III. RESULTS

A. Stability of dc state

Ultimately we wish to understand how small perturbations I_s on top of a large global background current I_b are amplified and represented by the model area, because this is the regime in which neocortex is found to operate [4]. For this we first analyze the stability of the state of constant activity across pools, referred to as "dc state," $\nu=(\nu_1,\ldots,\nu_N) =: \nu^{FP}\cdot\mathbf{1}$, under constant background input. $\mathbf{1}=(1,\ldots,1)^T$ is the N-dimensional dc vector, the superscript T denotes the transpose of a vector or matrix. The fixed point (FP) condition of Eq. (3) for the dc state reads $\nu^{FP}=g[I_b+(W_S+W_L)\nu^{FP}]$. For $W_S+W_L \geqslant 1/\max[g'(I)]=1$ there exists a regime of bistability for the global dc state, depending on the level of the background current, as illustrated in Fig. 1(a): at the level of I_{b1} there is a single low stable FP, I_{b2} implies bistability of a low and high dc state (separated by an unstable FP), and for I_{b3} only the high state remains. In the bistability regime, a specific input current given to a subset of pools can cause a flip of *all* pools from the low to the high state, caused by the strong lateral excitation $W_L > 1-W_S$. This effect might be a neural correlate of mental associations, where the neural representation of an unconditioned stimulus—giving rise to specific input—autonomously evokes co-activation of neurons that represent associated stimuli—the global flip [27]. It is worth noting that the network's ability to autonomously co-activate pools associated with each other can be dynamically switched on or off by varying the global background current I_b, for example between the values I_{b1} and I_{b2} in Fig. 1(a).

For the regime $W_S+W_L<1$, a single FP of the dc state exists, which deserves further stability analysis. We henceforth consider the representative special case of a balanced network, defined by an equal absolute strength of net self-excitation and the effective lateral inhibition, $W_L=-W_S$. Results remain qualitatively similar when we leave this point, however, as long we stay within the regime specified above. A balanced network is completely characterized by the self-excitation W_S and the global background current I_b. A small perturbation vector ε from the FP state, $\nu=\nu^{FP}\mathbf{1}+\varepsilon$, evolves according to

$$\tau\frac{d\varepsilon}{dt} = -\alpha\varepsilon + \beta(\mathbf{1}^T\varepsilon)\mathbf{1}, \quad \alpha = 1 - g'(I_b)W_S, \quad \beta = \frac{g'(I_b)W_L}{N}. \quad (4)$$

In the regime considered, $\beta<0$ holds and the second term on the right-hand side of Eq. (4) acts as to suppress the growth

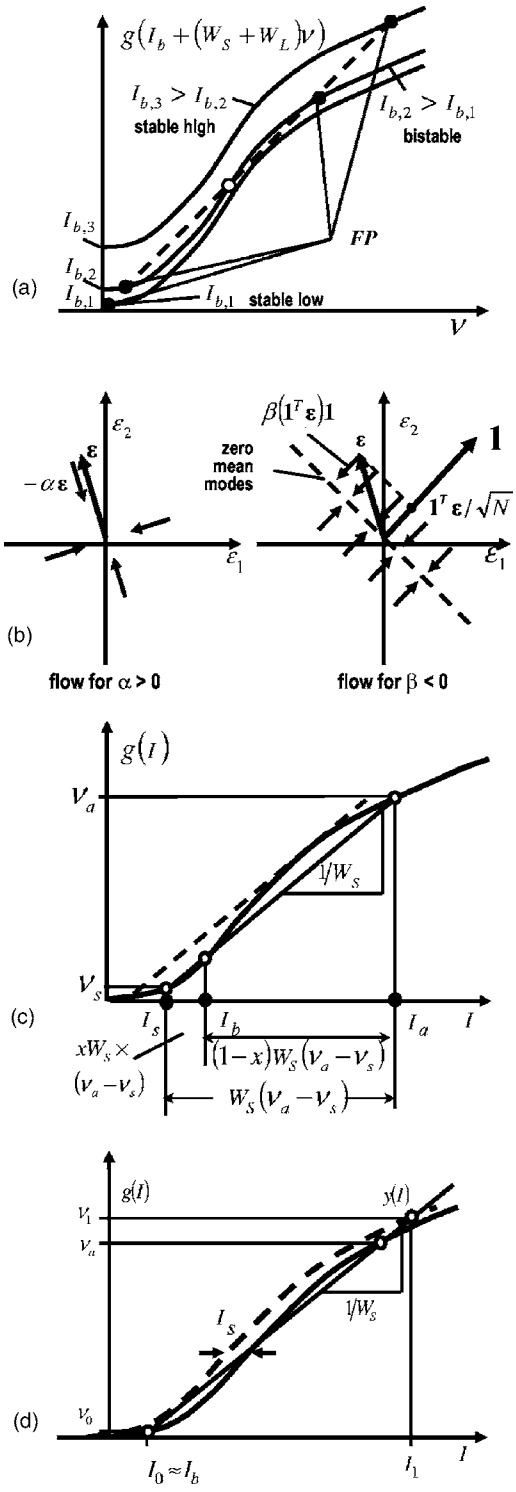

FIG. 1. Schematic illustrations to graphical solutions concerning (a) collective bistability, (b) stability of the state of constant activation, (c) multistability of memory states, and (d) stimulus amplification thresholds. See text for details.

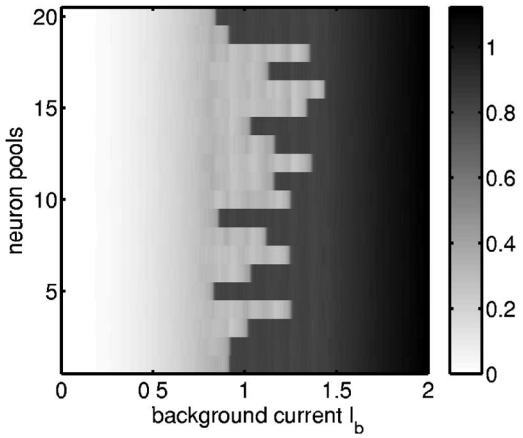

FIG. 2. Gray-level coded activities of a model area with $N=20$ pools (vertical axis) driven by a constant background current I_b only. During the simulation, I_b is continuously increased from 0 to 2 over 10 000 time steps (horizontal axis). To break symmetries, weak white Gaussian noise ($\sigma=0.01$) was superimposed onto the background current. Other parameters: $W_S=-W_L=1.25$, $\tau=5$. When I_b exceeds $0.8=1/W_S$, structured activity patterns spontaneously emerge and die out again for higher values.

of the dc mode [Fig. 1(b), right]. Hence for $\alpha>0$ the dc state is a stable fixed point [Fig. 1(b), left]. However, when α changes sign, the dc state becomes unstable, and any mode orthogonal to the suppressed dc mode, i.e., any mode with zero mean [along dashed line or hyperplane in Fig. 1(b), right] grows. Zero-mean modes represent structured patterns of neuronal activity across pools. Hence, for $\alpha<0$, small fluctuations are spontaneously amplified by the network to form structured activity patterns. We refer to the regime $\alpha<0$ as "spontaneous pattern formation" regime. From the phase boundary condition $g'(I)W_S=1$ and for the example of the gain function g assumed here we obtain the phase boundary current for spontaneous pattern formation, I_{sp}, for $W_S \geqslant 1$ as $I_{sp}=1/W_S$, $0\leqslant I\leqslant 1$; $I_{sp}=(W_S^2+3)/4$, $I>1$. The phase boundary for spontaneous pattern formation, $I_{sp}(W_S)$, is plotted in the summarizing phase diagram Fig. 4 (thick solid).

Figure 2 summarizes the results of a numeric simulation of Eqs. (3) with $N=20$ pools, testing how the stability of the dc state depends on the level of the background current (cf. Sec. III A). For this, the constant background current is slowly increased from zero. When turning up I_b, first small fluctuations are increasingly amplified, but the dc state remains stable. Above the critical value I_{sp}, the area enters the spontaneous pattern formation regime: subsets of pools are spontaneously and discontinuously activated to a high state, whereas the remaining pools are slightly deactivated such as to preserve the mean activity across pools. When I_b further increases, the fraction of activated pools rises, as indicated by Eq. (6), whereas the levels of activity do not change much. Above the second critical background current, the dc state re-stabilizes, but this time at the fully activated level.

The equation for the decay parameter α highlights the functional symmetry between a synaptic efficacy W_S and the background current I_b. Besides W_S also I_b can cause the system to enter or leave the spontaneous pattern formation regime. There is evidence from functional imaging studies that cortical activation without stimulation is a likely neural correlate of imagery and spontaneous thoughts [30], and that imagery and perception draw on roughly the same neural machinery [31]. In light of this, spontaneous and reversible activation of neuronal assemblies are likely to be related with

the spontaneous pop-up of memories and mental images in the brain, a process which might underly the emergence of creative thoughts. Here we demonstrate that this property of a model cortical network can be dynamically enabled or suppressed by adjusting the global background synaptic current driving it.

B. Multistability of memory states

Stable states of structured persistent neuronal activity in the absence of structured stimulation are considered a neural substrate of working memory or short-term memory [32,33]. In analogy, we identify a persistently active pool that is not driven by specific input with a short-term memory stored in the considered network model. We next analyze the existence and multistability properties of structured activity states, referred to as memory states, which contain a subset of persistently active pools. The general fixed point conditions Eq. (3) for a balanced network without the restriction to the dc ansatz are given by $\nu_k = g[I_b + W_S(\nu_k - \bar{\nu})]$, where $\bar{\nu}$ denotes the mean activity across pools. The FP equations are separable in k, i.e., the same FP equation holds for all pools. If $W_S < 1/\max(g') = 1$, there is one stable FP solution, hence only the dc state is stable. For $W_S > 1$, there exists a regime of bistability: in this regime, each pool can be in one of two possible stable states: an active one with spike rate ν_a or a silent one with spike rate ν_s.

We now consider a state where M pools are active and $N-M$ pools are silent, and analyze under which conditions that ansatz solves the resulting FP equations,

$$\nu_a = g(I_a) = g[I_b + W_S(1-x)(\nu_a - \nu_s)],$$

$$\nu_s = g(I_s) = g[I_b - W_S x(\nu_a - \nu_s)]. \tag{5}$$

In Eqs. (5), $x = M/N$ is the fraction of persistently active pools. Figure 1(c) illustrates the graphical solution of the fixed point equations: The ansatz solves Eqs. (5), if there exists a linear function with slope $1/W_S$ which (i) intersects the function $g(I)$ at least two times, (ii) such that the fraction of distances of the intersection currents I_s and I_a from the background current I_b is given by $x/(1-x)$. The resulting FP solutions can be shown to be stable. We now consider the leftmost linear function with the mentioned properties, namely the one that forms the tangential of the saturating part of g [dashed line in Fig. 1(c)]. For the gain function specified above, its intersection currents are given by $I_{a,0} = (W_S^2+3)/4$, $I_{s,0} = \{1-[1-(3-W_S^2)W_S/2]^{1/2}\}/W_S$. In order to obtain stable FP states with at least one pool persistently active, i.e., $x = 1/N$, the background current (besides $I_b \leqslant I_{a,0}$) must fulfill $I_b \geqslant I_{wm} = I_{s,0} + (I_{a,0} - I_{s,0})/N$. Hence we observe a phase boundary $I_{wm}(W_S)$ for the existence of stable working memory states, which is plotted in Fig. 4 (thick dashed). By shifting the background current above or below this boundary, the area's ability to maintain working memories by means of persistent neuronal activity can be dynamically switched on or off. Further, for $I_b \geqslant I_{wm}$, the maximum fraction of simultaneously active pools is given by

$$x_{\max} = (I_b - I_{s,0})/(I_{a,0} - I_{s,0}), \tag{6}$$

digitized to steps of $1/N$.

I_{wm} and I_{sp} outline a regime of multistability in which both the dc state and patterned working memory states coexist. In this regime, the brain area can be brought from a state of constant activation into a patterned working memory state by means of transient external stimulation. Therefore this regime is referred to as *induced* pattern formation regime or working memory regime. This result agrees well with the findings on phase boundaries for multistability found in similar settings [26,34].

It is worth noting, however, that, in addition to the ability of the model area to maintain persistently active memory states, also the maximum number of co-existing persistently active pools (i.e., the number of co-existing memories) can be dynamically adjusted by the global background current I_b. The maximum number of co-existing memories, in turn, can be related to different functions being carried out by the area. For example, when only one memory can exist, we face distractible working memory (found in inferotemporal cortex), which always keeps the most recent input. When several memories can co-exist, we face a network that shows nondistractible working memory (found in prefrontal cortex), which can accumulate memories across time ("folding in time"). Finally, an area that can keep several memories stimulated in the near past, can represent the recent history of a time series of inputs, and can serve as to represent short-term histories of trajectories.

C. Amplification of small specific stimuli

Next it is analyzed how the network responds to a small specific stimulation current I_s. We consider the biologically realistic regime $I_s \ll I_b$ in which the specific input represents a small perturbation which rides on top of the big global and constant background current. In order to represent such a small stimulus by macroscopically detectable neuronal activity, the network must be able to strongly and selectively amplify this small input. Here we provide an approximate analytical solution for the activation threshold of the network for the limit of large number of pools, for $W_S > 1$ and for I_b below the spontaneous pattern formation regime. For the sake of brevity, we consider the special case of only one pool, say, pool number 1, being stimulated when the rest of the network rests in the dc state. Again, all pools except the first one are described by the same equation and, according to our assumption rest in the same, inactive state, $\nu_2 = \cdots = \nu_N =: \nu_0$. For large N, the fixed point equations can be approximated by

$$\nu_1 \approx g(I_1) = g[I_b + W_S(\nu_1 - \nu_0) + I_s], \tag{7}$$

$$\nu_0 \approx g(I_b) = \nu^{FP}. \tag{8}$$

Figure 1(d) illustrates the graphical solution of Eqs. (7) and (8). Whereas the unstimulated pools remain at the spontaneous activity level ν^{FP}, the FP activity of the stimulated pool results from an intersection of the gain function, shifted to the left by I_s, with the line $y(I) = g(I_b) + (I - I_b)/W_S$ [straight

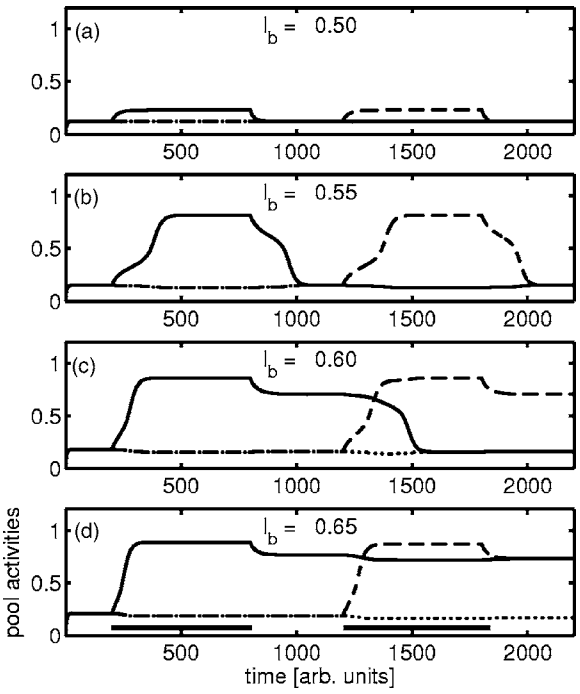

FIG. 3. Responses over time of two stimulated pools (8 and 12, chosen arbitrarily for convenience) and the background activity (dotted line) in a model area for different background currents. Pool 8 stimulus: $I_s=0.05$ in $t\in[200,800]$ (response: solid line). Pool 12 stimulus: $I_s=0.05$ in $t\in[1200,1800]$ (response: dashed line). (a) $I_b=0.50$, no stimulus representation, (b) $I_b=0.55$, stimulus representation, no persistent activity, (c) $I_b=0.60$, persistent activity, only one memory state allowed (distractible working memory) and (d), $I_b=0.65$, working memory, more than one memory state co-exist. Other parameters are $N=20$, $W_S=-W_L=1.25$.

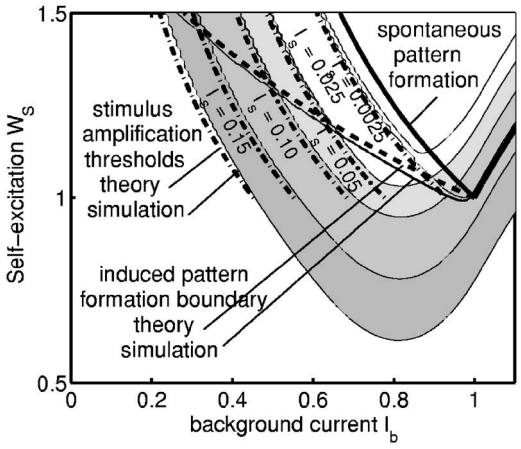

FIG. 4. Phase regimes for a balanced network. Theory: thick solid, $I_{sp}(W_S)$ spontaneous pattern formation boundary; thick dashed, $I_{wm}(W_S)$, induced pattern formation (working memory) boundary; thick dash-dotted, stimulus amplification boundaries for $I_{th}=0.15$, 0.10, 0.05, 0.025, 0.0025; thin solid line, simulation result for I_{wm}, determined by the condition $\nu_a>\nu^{FP}+0.03$. Contour plot: Simulation results for stimulus amplification thresholds (same values as in theory), determined by $\nu_1>\nu^{FP}+0.3$.

line in Fig. 1(d)]. When the stimulus current is steadily increased from zero, ν_1 increases continuously from ν^{FP}, but the pool stays within the lower fixed point, i.e., within the accelerating part of the gain function. However, for a critical threshold current I_{th}, the lower FP for pool 1 vanishes, and its activity discontinuously jumps to a high level in the saturating part of the gain function. This happens when the shifted gain function becomes tangential to the linear function $y(I)$ [illustrated by the dashed curve in Fig. 1(d)]. We define I_{th} as the activation threshold of the network for one stimulus. Please note that only if the pool arrives at the upper FP can it reach the persistently active state ν_a after termination of the stimulus [cf. Fig. 1(d)]. Hence superthreshold stimulation is a necessary condition for subsequent storage in working memory.

For the gain function considered here, the threshold current for activation is given by $I_{th}=W_S(1/W_S-I_b)^2/W_S$, $I_b<1/W_S$, $W_S\geqslant 1$. This demonstrates that also the network's activation threshold (and, related to it, the memory storage threshold) can be dynamically adjusted by the global background current. The larger I_b, the smaller stimuli are amplified and represented by the network. When I_b approaches $1/W_S$ from below, the threshold current becomes infinitely small, and we arrive at the spontaneous pattern formation regime. This is also illustrated in Fig. 4, which displays phase boundaries in terms of I_b and W_S where the amplification threshold becomes $I_{th}=0.15$, 0.1, 0.05, 0.025, and 0.0025 (thick dash-dotted lines).

The dynamic adjustment of stimulus thresholds by the background current can serve as an important neuronal mechanism for dynamically varying the sensitivity of brain systems. As such, it might form a neuronal correlate for the control of alertness and arousal.

Figure 3 shows how the same network as used in Fig. 2 responds to two small specific stimuli and how this response depends on the global background current. For this, pool 8 of the network received a stimulus $I_{s,1}=0.05$ during time steps 200 – 800, and pool 12 received $I_{s,2}=0.05$ during time steps 1200 – 1800. In the plots of Fig. 3, pool 8 activity is shown as a solid line, pool 12 activity as a dashed line, and the average activity of the remaining 18 pools as a dotted line. For the top traces, calculated for $I_b=0.5$, these inputs are both subthreshold, resulting in a small activation [Fig. 3(a)]. When the background current (and only the background current) is changed to 0.55, the same inputs are now superthreshold, resulting in an accelerating amplification and a resulting strong activation [Fig. 3(b)]. When I_b is further increased to 0.6, the first stimulus evokes persistent activity beyond stimulation, i.e., stimulus 1 is stored in short term memory [Fig. 3(c), solid line]. However, the activation of the other pool [Fig. 3(c), dashed line] erases this memory, because for the background current chosen, only one memory at maximum can exist [$Nx_{\max}=1$, cf. Eq. (6)]: we face a distractible working memory. Finally, for even higher background current, both stimuli are simultaneously stored in persistent activity (nondistractible working memory).

Figure 4 summarizes the theoretical (thick lines, described before) and simulated phase boundaries for spontaneous pattern formation, induced pattern formation and stimulus amplification. In the simulations, the induced pattern formation boundary (thin solid line) was determined by the condition,

that persistent activity after stimulation was 0.03 above the background activation. The stimulus amplification thresholds (contour plot, thin lines) were obtained by the condition, that the stimulus response was at least 0.3 above the background activation. These thresholds are arbitrary values, but the results near the analytical curves are largely independent of the exact choice. A comparison of the result shows a very good agreement of the analytical and simulated phase boundaries.

D. Functional gating by subnetwork bias

In light of the dramatic influence of the global background current on a complete network's mode of operation, we next tested whether changing the background current of a subnetwork could selectively change the functional properties of that subnetwork. For this, a model area was stimulated by a small set of specific stimuli, $I_{s,k}$, while a subnetwork of the area received a weak uniform bias current I_{bias} on top of the global background current I_b. Both the stimuli and the bias signals were additive. They are distinguished from each other by their sparsity and strength. Following the biologically observed sparsity of activation, stimuli are considered sparse as well. They punctually target a small fraction of pools, and can be subthreshold or superthreshold. On the other hand, the subnetwork bias is thought to target a larger fraction of pools, to be uniform in strength and to be subthreshold. The biological counterpart of the stimulus current is specific bottom-up input, which carries for example sensory information. The biological counterpart of the bias input is thought to be top-down input from higher representations, which tends to be less specific and modulatory in nature. It is hypothesized to carry back information about hypotheses generated in higher areas and to bias the target area's mode of processing. The mutual interaction of brain areas via bottom-up stimuli and top-down biases has been recently suggested as a powerful framework for the computational neuroscience of visual cognition [8,10,35].

The situation of subnetwork biasing is schematically illustrated in Fig. 5, where, in addition to a constant global background current I_b, the subnetwork formed by pools $1-n$ receives a weak extra additive bias current I_{bias}. When the bias is switched on, the subnetwork inside the dashed box becomes the "biased subnetwork." In the following simulations it is tested, if and how the presence of this subnetwork bias can affect the stimulus response and representation selectively for the biased subnetwork. One example of such an influence of the bias is outlined in Fig. 5, where two pools p and q, one inside and one outside the biased subnetwork, are stimulated by identical currents $I_{s,p}$ and $I_{s,q}$. In the illustrated example, pool p of the biased subnetwork strongly responds to specific stimulation $I_{s,p}$ (highlighted color), whereas the other pool q outside the subnetwork does not respond to its stimulus $I_{s,q}$. In this example, the subnetwork bias acts like a gating variable that enables or prohibits response to stimulation. Note that the definition of a biased subnetwork is dynamic, i.e., by sequentially applying bias currents to different collections of pools, different subnetworks can be dynamically biased (e.g., gated) one after the other as time proceeds.

In the following, a set of exemplifying simulations are presented to demonstrate that such a gating property of a

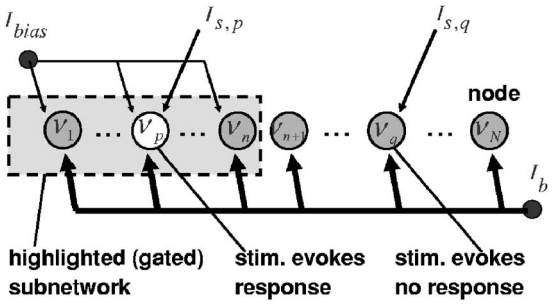

FIG. 5. Schematic outline of a subnetwork bias and its effect. In addition to a constant global background current I_b, a part of the whole network (pools $1-n$ in the illustration) receives a weak extra biasing current I_{bias}. It is tested, how the presence of this subnetwork bias modifies the response to stimulation of the subnetwork as compared to the rest of the network. In the illustrated example, a pool p of the biased subnetwork strongly responds to specific stimulation $I_{s,p}$ (indicated by white color), whereas another pool q outside the subnetwork does not respond to a stimulus $I_{s,q}$ of the same strength.

subnetwork bias actually exists, and that it can highlight different response and representational properties of the biased subnetwork, depending on the value of the global background current.

Figures 6–8 show the results of a set of exemplifying simulations run for a model network with $N=20$ pools, where ten pools receive a weak additive bias current I_{bias}. Shown are the activities of the 20 pools over time (one row for each pool), encoded in the gray level. White means low activity and black indicates high activity. Figure 6(a) displays, for comparison, the response of the network to two small stimuli imposed sequentially on pools $p=8$ and $q=12$ [stimulus durations indicated by the black bars at the bottom of Fig. 6(b)], when the network is in the distractible working

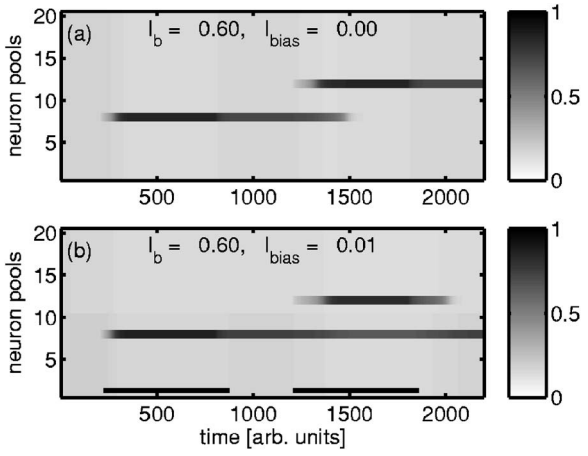

FIG. 6. Gray-level coded responses over time of a network of $N=20$ pools for two sequential stimuli to pools 8 with $I_s=0.05$ in $t\in[200,800]$ and 12 with $I_s=0.05$ in $t\in[1200,1800]$ (black bars at bottom); (a) With uniform global bias $I_b=0.6$: the second stimulus destroys the persistent activity [same as Fig. 3(c), for comparison]. (b) As (a), but pools 1–10 receive a weak extra bias $I_{bias}=0.01$. The persistent activity is stabilized, leading to nondistractible working memory. Other parameters are $W_S=-W_L=1.25$.

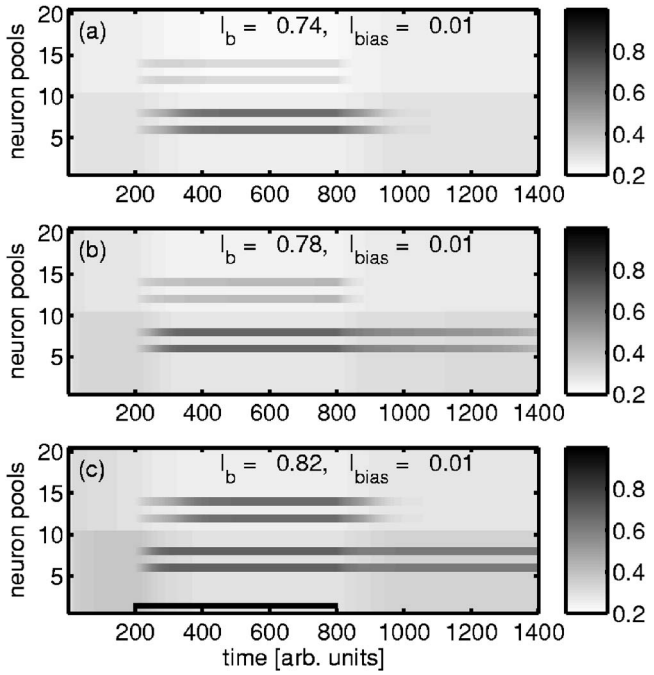

FIG. 7. Gray-level coded responses over time of a network of $N=20$ pools where pools 6, 8, 12, 14 are stimulated by weak input $I_s=0.025$ in $t \in [200,800]$. Pools 1–10 receive a weak extra bias of $I_{bias}=0.01$. (a) For $I_b=0.74$, only stimuli that target the biased subnetwork evoke superthreshold response (attentional spotlight). (b) For $I_b=0.78$, stimuli that target the biased subnetwork are selectively amplified and subsequently stored in working memory (attention-gated working memory). (c) For $I_b=0.82$, all stimuli are represented, but only the ones in the biased subnetwork are stored to working memory (context dependent working memory). Other parameters are $W_S=-W_L=1.1$.

memory regime [same simulation as in Fig. 3(c)]. Pools 8 and 12 are chosen arbitrarily, the numbers are selected for reasons of visualization only. Figure 6(b) shows the network's response under the same conditions, except that pools 1–10 received a weak bias signal $I_{bias}=0.01$, which was kept constant over time and space. It can be seen that the bias, although itself being clearly subthreshold in strength, strongly stabilizes the working memory of pool 8, which keeps its persistent activity despite the transient activation of pool 12. As an effect of the bias, the memory has become nondistractible while still being selective. In this example, the subnetwork bias acts as to stabilize a memory content, the network "concentrates" on a memory aspect.

The simulation runs displayed in Fig. 7 illustrate how the effect of the subnetwork bias depends on the functional mode of the whole network, adjusted by I_b. This time, four pools were stimulated to test, whether the subnetwork bias affects the representation of an individual stimulus or rather the properties of the whole subnetwork [stimulus duration indicated by the black bar in Fig. 7(c)]. Again, pools 1–10 received a weak and uniform bias of $I_{bias}=0.01$. In Fig. 7(a), it can be seen that the bias signal selectively enables the amplification and representation of the stimulated pools within the biased subnetwork (pools 6 and 8), whereas the response of the other stimulated pools (pools 12 and 14) is

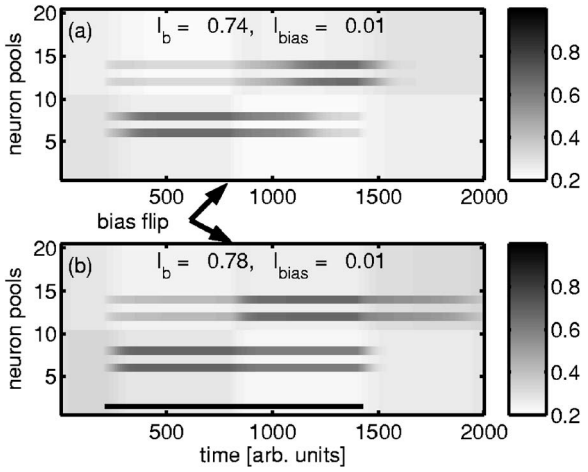

FIG. 8. Gray-level coded responses over time of a network of $N=20$ pools where pools 6, 8, 12, 14 are stimulated by weak input $I_s=0.025$ in $t \in [200,1400]$, while the subnetwork bias flips: For $t<800$, pools 1–10, for $t \geqslant 800$ pools 11–20 receive a weak extra bias of $I_{bias}=0.01$. (a) For $I_b=0.74$, stimuli in the biased subnetwork are highlighted in a reversible way. (b) For $I_b=0.78$, the attentional highlight is no longer reversible, however the last attended stimuli are selectively stored in working memory. Other parameters are $W_S=-W_L=1.1$.

suppressed. Under this condition, the bias signal acts as an attentional spotlight, where only representations within that focus are enabled. In Fig. 7(b), where the global background current is slightly increased, the biased pools are again selected to represent their inputs by superthreshold activity, but now they are persistently active after stimulation. In this situation, the bias signal gates storage of a selected (attended) part of the input stimuli into working memory. When the global background signal is further increased, all stimulated pools become active, but only the biased pools stay active post stimulation and form memory. The context represented by the bias signal determines working memory formation.

The simulations in Fig. 8 explore the dynamical properties of the gating mechanism implemented by the subnetwork bias. Shown is the response of the same network as in Fig. 7 [stimulus durations indicated by the black bar on the bottom of Fig. 8(b)], where in the middle of the stimulus presentation the bias flips from pools 1–10 to pools 11–20. For the situation in Fig. 8(a), first the biased stimuli are selectively amplified, as observed before. When the bias flips, the focus of amplification also changes: the now biased subnetwork represents its stimuli, and the unbiased pools' activities are suppressed. In this mode, the network implements a shifting focus of attention. Figure 8(b) shows the response of the same network with a slightly increased global background current. When the bias flips, the new highlighted pools are enabled, but the old ones are not disabled. In this mode the network implements a trace of attention. The last attended pools stay active and form a working memory.

It is worth noting that in all cases the weak additive signal imposes a strongly multiplicative effect on selective activation and memory. It acts as a powerful gating mechanism which can enable or disable various different representa-

tional and memory properties of the biased subnetwork. This effect can be used to dynamically enable or disable representations in arbitrary subnetworks, and as such can implement an enormous flexibility of representation.

IV. SUMMARY AND CONCLUSIONS

We propose an important functional role for the global spontaneous background activity in the framework of stimulus representation, selective attention and working memory: the dynamic tuning of the functional mode and the response properties of neocortical areas and subnetworks by the level of global spontaneous background input. By analyzing an easily tractable model brain area that is prototypic for a wide class of realistic mean-field formulations we demonstrate, that the level of background input controls the stimulus threshold, the ability to encode stable short-term memories, the maximum number of coactive memories, and the ability to spontaneously activate mental representations. Whereas the configuration of synaptic weights determines the properties of internal representations, the background input can dynamically control the mode of their use, and thereby dynamically tunes the area's function. By means of these mechanisms, the level of spontaneous activity might also underly the control of alertness and arousal.

When the background signal is varied individually for subnetworks by a weak bias, it acts as a strongly nonlinear and flexible functional gating mechanism for that subnetwork. Having in mind the functional tuning properties of a background signal, the strongly nonlinear effect of such a weak additive bias can be understood in a natural way: it drives a subnetwork into a different functional regime. As subnetwork biases can arise from pools of other brain areas, they represent powerful and flexible candidates of context signals. Cortical areas might generate a set of bias signals, by which they selectively and dynamically highlight parts of other brain areas. Such a multiareal system might bear the potential for humanlike cognitive flexibility.

ACKNOWLEDGMENTS

The author gratefully acknowledges support through the German Federal Ministry for Research, BMBF Grant No. 01IBC01A. I wish to thank Bernd Schürmann for fruitful discussions and helpful comments on the manuscript.

[1] D. C. Van Essen, C. H. Anderson, and D. J. Felleman, Science **255**, 419 (1992).
[2] D. J. Amit and M. Tsodyks, Network **3**, 121 (1992).
[3] F. Wilson, S. Scalaidhe, and P. Goldman-Rakic, Proc. Natl. Acad. Sci. U.S.A. **91**, 4009 (1994).
[4] K. W. Koch and J. M. Fuster, Exp. Brain Res. **76**, 292 (1989).
[5] D. J. Amit and N. Brunel, Cereb. Cortex **7**, 237 (1997).
[6] D. J. Amit and N. Brunel, Network Comput. Neural Syst. **8**, 373 (1997).
[7] N. Brunel, J. Comput. Neurosci. **8**, 183 (2000).
[8] E. T. Rolls and G. Deco, *Computational Neuroscience of Vision* (Oxford University Press, Oxford, 2002).
[9] G. Deco and E. T. Rolls, Eur. J. Neurosci. **8**, 2374 (2003).
[10] M. Szabo, R. Almeida, G. Deco, and M. Stetter, Eur. J. Neurosci. **19**, 305 (2004).
[11] X.-J. Wang, J. Neurosci. **19**, 9587 (1999).
[12] N. Brunel and X. J. Wang, J. Comput. Neurosci. **11**, 63 (2001).
[13] P. Del Giudice, S. Fusi, and M. Mattia, J. Physiol. (Paris) **97**, 659 (2003).
[14] G. Deco, E. T. Rolls, and B. Horwitz, J. Cogn Neurosci. **16**, 683 (2004).
[15] R. Almeida, G. Deco, and M. Stetter, Eur. J. Neurosci. **20**, 2789 (2004).
[16] M. Szabo, R. Almeida, G. Deco, and M. Stetter, Neurocomputing **65–66**, 195 (2005).
[17] X.-J. Wang, Neuron **36**, 955 (2002).
[18] H. C. Tuckwell, *Introduction to Theoretical Neurobiology* (Cambridge University Press, Cambridge, England, 1988), Vol. 2.
[19] N. Brunel and S. Sergi, J. Theor. Biol. **195**, 87 (1998).
[20] M. Mattia and P. Del Giudice, Phys. Rev. E **66**, 051917 (2002).
[21] M. Stetter, *Exploration of Cortical Function* (Kluwer Academic Publishers, Dordrecht, Boston, 2002).
[22] G. La Camera, A. Rauch, H.-R. Lüscher, W. Senn, and S. Fusi, Neural Comput. **16**, 2101 (2004).
[23] C. A. van Vreeswijk and H. Sompolinsky, Science **274**, 1724 (1996).
[24] W. Gerstner, Neural Comput. **12**, 43 (2000).
[25] M. C. van Rossum, G. G. Turrigiano, and S. B. Nelson, J. Neurosci. **22**, 1956 (2002).
[26] E. Salinas, Neural Comput. **15**, 1439 (2003).
[27] G. Mongillo, D. J. Amit, and N. Brunel, Eur. J. Neurosci. **18**, 2011 (2003).
[28] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **81**, 3088 (1984).
[29] R. Azouz, C. M. Gray, L. G. Nowak, and D. A. McCormick, Cereb. Cortex **7**, 534 (1997).
[30] S. M. Kosslyn, G. Ganis, and W. L. Thompson, Nat. Rev. Neurosci. **2**, 635 (2001).
[31] G. Ganis, W. L. Thompson, and S. M. Kosslyn, Brain Res. Cognit. Brain Res. **20**, 226 (2004).
[32] J. M. Fuster and G. Alexander, Science **173**, 652 (1971).
[33] E. K. Miller, C. A. Erickson, and R. Desimone, J. Neurosci. **16**, 5154 (1996).
[34] N. Brunel, Cereb. Cortex **13**, 1151 (2003).
[35] G. Deco, R. Almeida, M. Loh, M. Szabo, and M. Stetter, Recent Res. Dev. Phys. **5**, 995 (2004).