Competition Through Selective Inhibitory Synchrony

Ueli Rutishauser
urut@brain.mpg.de
Department of Neural Systems, Max Planck Institute for Brain Research,
Frankfurt am Main, Hessen 60528, Germany

Jean-Jacques Slotine
jjs@mit.edu
Nonlinear Systems Laboratory, Massachusetts Institute of Technology,
Cambridge, MA 02142, U.S.A.

Rodney J. Douglas
rjd@ini.phys.ethz.ch
Institute of Neuroinformatics, ETH Zurich and University of Zurich,
Zurich 8057, Switzerland

Models of cortical neuronal circuits commonly depend on inhibitory feedback to control gain, provide signal normalization, and selectively amplify signals using winner-take-all (WTA) dynamics. Such models generally assume that excitatory and inhibitory neurons are able to interact easily because their axons and dendrites are colocalized in the same small volume. However, quantitative neuroanatomical studies of the dimensions of axonal and dendritic trees of neurons in the neocortex show that this colocalization assumption is not valid. In this letter, we describe a simple modification to the WTA circuit design that permits the effects of distributed inhibitory neurons to be coupled through synchronization, and so allows a single WTA to be distributed widely in cortical space, well beyond the arborization of any single inhibitory neuron and even across different cortical areas. We prove by nonlinear contraction analysis and demonstrate by simulation that distributed WTA subsystems combined by such inhibitory synchrony are inherently stable. We show analytically that synchronization is substantially faster than winner selection. This circuit mechanism allows networks of independent WTAs to fully or partially compete with other.

1 Introduction

Many models of neuronal computation involve the interaction of a population of excitatory neurons whose outputs drive inhibitory neuron(s), which

in turn provide global negative feedback to the excitatory pool (Amari & Arbib, 1977; Douglas, Koch, Mahowald, Martin, & Suarez, 1995; Hahnloser, Sarpeshkar, Mahowald, Douglas, & Seung, 2000; Yuille & Geiger, 2003; Maass, 2000; Hertz, Krogh, & Palmer, 1991; Rabinovich et al., 2000; Rutishauser, Douglas, & Slotine, 2011; Coultrip, Granger, & Lynch, 1992). Practical implementation of such circuits in biological neural circuits depends on colocalization of the excitatory and inhibitory neurons, an assumption that studies of the extent of axonal and dendritic trees of neurons in the neocortex show is not valid (Katzel, Zemelman, Buetfering, Wolfel, & Miesenbock, 2011; Binzegger, Douglas, & Martin, 2004; Shepherd, Stepanyants, Bureau, Chklovskii, & Svoboda, 2005; Douglas & Martin, 2004). First, a substantial fraction of the axonal arborization of a typical excitatory "spiny" pyramidal neuron extends well beyond the range of the arborization of a typical "smooth" inhibitory neuron, particularly in the populous superficial layers of the neocortex (Yabuta & Callaway, 1998; Binzegger et al., 2004). This spatial arrangement means that excitatory effects can propagate well outside the range of the negative feedback provided by a single inhibitory neuron. Second, the horizontally disposed basket type of inhibitory neuron, a prime candidate for performing normalization, makes multiple synaptic contacts with its excitatory targets, so that even within the range of its axonal arborization, not all members of an excitatory population can be covered by its effect. This connection pattern means that excitatory neurons within some local population must either be partitioned functionally or multiple smooth cells must cooperate to cover the entire population of excitatory cells.

In previous publications, we have shown how winner-take-all (WTA) circuits composed of a small population of excitatory neurons and a single inhibitory neuron can be combined to construct supercircuits that exhibit finite state machine (FSM)–like behavior (Rutishauser & Douglas, 2009; Neftci, Chicca, Indiveri, Cook, & Douglas, 2010). The supercircuits made use of sparse excitatory cross-connections between WTA modules to express the required states of the FSM. These excitatory connections can extend well outside the range of the local WTA connections, and so are consistent with the observed long-range lateral excitatory connections referred to above. On the other hand, we have not previously confronted the question of whether the WTA is necessarily localized to the extent of the smooth-cell arborization, or whether the WTA itself can be well distributed in space within or between cortical lamina, or even between cortical areas. In this letter, we describe a simple modification to the WTA circuit design that couples the effects of distributed inhibitory neurons through synchronization, and so permits a WTA to be widely distributed in cortical space, well beyond the range of the axonal arborization of any single inhibitory neuron, and even across cortical areas. We also demonstrate that such a distributed WTA is inherently stable in its operation.
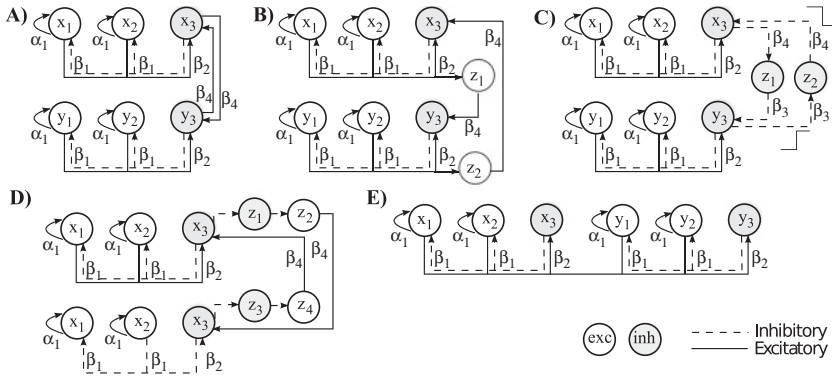
Figure 1: Circuits for distributing WTAs. (A) Illustration of the principal idea: mutual excitation of the inhibitory neurons. (B–E) Biologically plausible versions. (B) Implementation using intermediate excitatory neurons $z_{1,2}$. This circuit will be considered in detail in Figure 2B with more realistic connectivity. (C) Implementation using disinhibition of persistently active units $z_{1,2}$ as illustrated by the step functions. (D) Implementation with disinhibition and long-range excitatory units. (E) Implementation using multiplication of inhibitory neurons. Here, $x_3 = y_3$ at all times. The maximal excitatory projection length is double that of the inhibitory.

2 Results

We have considered a number of circuits that could be used to distribute spatially the WTA behavior (see Figure 1). However, we will describe and analyze only the circuit shown in Figure 2, which we consider to be the most elegant of the distributive mechanisms (notice the similarity to Figure 1B). The key insight is the following: under normal operating conditions, all the participating distributed inhibitory neurons should receive the same summed excitatory input. We achieve this by interposing an excitatory neuron in the negative feedback loop from the excitatory population to its local inhibitory neuron. Instead of the local inhibitory neuron summing over its excitatory population, the interposed neuron performs the summing and passes its result to the inhibitory neuron. This result is also copied to the more distant inhibitory neurons in the spatially distributed WTA. In this way, the inhibitory neuron of each sub-WTA sums over the projections from the interposed excitatory neurons of all other sub-WTAs, including its own one. Thus, each inhibitory neuron is able to provide feedback inhibition to its local sub-WTA that is proportional to the total excitation provided by all excitatory neurons participating in the entire distributed WTA. We will show that functionally, this amounts to a form of synchrony among all the inhibitory units.
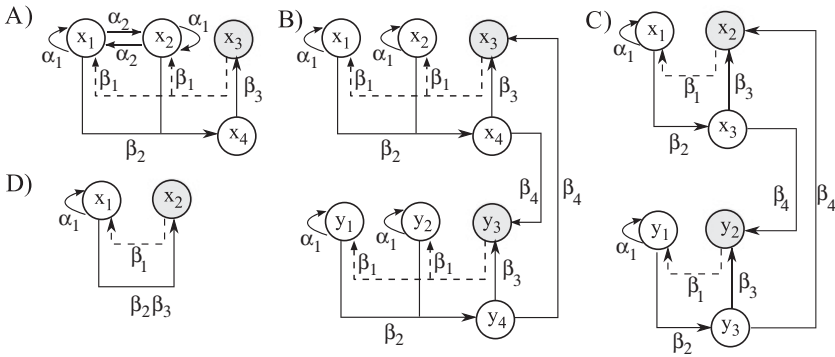
Figure 2: Schematic of connectivity. Gray units and dashed lines are inhibitory, and white units and straight lines are excitatory. (A) Single WTA with two possible winners $x_{1,2}$, inhibitory unit x_3, and intermediate excitatory unit x_4 that carries the average activity of $x_{1,2}$. (B) Coupling of two WTAs to form a single WTA with four possible winners. β_4 are excitatory long-range connections that serve to synchronize the two inhibitory units. (C–D) Reduced versions for theoretical analysis. (C) Merged WTA with one winner each and thus two possible winners x_1 and y_1 in the merged network. (D) Reduced single WTA.

2.1 Connectivity and Dynamics: Single WTA. All of the circuits of Figure 1 can achieve a distributed WTA by merging several independent WTAs, but we consider the circuit shown in Figure 2B to be most feasible, and so our analysis focuses on this one. However, similar reasoning could be applied to all the variants shown. Note that our chosen circuit is similar to that of Figure 1B, but it has a more realistic connectivity pattern in that the summed excitatory activity is projected onto only a single unit, which requires less wiring specificity than Figure 1B.

The dynamics of a single WTA (see Figure 2A), consisting of $1..N-2$ excitatory units, one inhibitory unit x_{N-1}, and one intermediary interconnect excitatory unit x_N, are

$$\tau \dot{x}_i + Gx_i = f(I_i + \alpha x_i - \beta_1 x_{N-1} - T_i),$$

$$\tau \dot{x}_{N-1} + Gx_{N-1} = f(\beta_3 x_N - T_{N-1}),$$

$$\tau \dot{x}_N + Gx_N = f\left(\beta_2 \sum_{j=1}^{N-2} x_j - T_N\right). \qquad (2.1)$$

Each excitatory unit receives recurrent input from itself (α_1) and its neighbors ($\alpha_{2,3,...}$; see Figure 2A). For simplicity, only self-recurrence is considered here ($\alpha = \alpha_1$ and $\alpha_{2,3,...} = 0$), but very similar arguments obtain when

recurrence from neighboring units is included. When the weight matrix $\mathbf{W}$ is used, the dynamics of this system is described as

$$\tau \dot{\mathbf{x}} + G\mathbf{x} = f(\mathbf{W}\mathbf{x} - \mathbf{T} + \mathbf{I}(t)), \qquad (2.2)$$

with

$$\mathbf{W} = \begin{bmatrix} \alpha & 0 & -\beta_1 & 0 \\ 0 & \alpha & -\beta_1 & 0 \\ 0 & 0 & 0 & \beta_3 \\ \beta_2 & \beta_2 & 0 & 0 \end{bmatrix}, \qquad (2.3)$$

where the order of units is $x_{1,2,3,4}$ (i.e., the first column and row is x_1, and the last column and row is x_4). The firing rate activation function $f(x)$ is a nonsaturating rectification nonlinearity $\max(0, x)$. We assume $\tau = 1$ and $G = 1$ throughout unless mentioned otherwise. $\mathbf{T} = [T_1, \ldots, T_{N-1}, T_N]$ is a vector of the constant activation thresholds $T_i \geq 0$.

2.2 Connectivity and Dynamics: Coupled WTA. Two identical single WTAs, each described by weight matrices $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$, can be combined into one distributed WTA that acts functionally as a single WTA by adding a recurrent excitatory feedback loop β_4 between the two WTAs (see Figure 2B). The weight matrix $\mathbf{W}_C$ of the merged system is

$$\mathbf{W}_C = \begin{bmatrix} \mathbf{W}_1 & \mathbf{C}_2 \\ \mathbf{C}_1 & \mathbf{W}_2 \end{bmatrix} \qquad (2.4)$$

with interconnections

$$\mathbf{C}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (2.5)$$

The dynamics of this system are as shown in equation 2.2 using $\mathbf{W} = \mathbf{W}_C$.

2.3 Stability Analysis. The stability analysis, using nonlinear contraction analysis (see the appendix; Lohmiller & Slotine, 1998, 2000; Slotine, 2003; Wang & Slotine, 2005), consists of three steps: (1) demonstrate the contraction of a single WTA, (2) merge two WTAs by demonstrating that inhibitory units synchronize, and (3) demonstrate the contraction of the combined WTAs. We have previously shown how contraction analysis can

be applied to reasoning over the stability and functionality of WTA circuits (Rutishauser et al., 2011). Here, we apply and extend the same methods to this new circuit.

Contraction analysis is based on the Jacobians of the system. For purposes of analysis but without loss of generality, we base this section on a reduced system with only one possible winner for each WTA, as shown in Figure 2C.

The Jacobian of a single system is

$$\tau \mathbf{J} = \begin{bmatrix} l_1\alpha - 1 & l_1 - \beta_1 & 0 \\ 0 & -1 & l_2\beta_3 \\ l_3\beta_2 & 0 & -1 \end{bmatrix}. \tag{2.6}$$

In a stable network, a constant external input to the first unit x_1 will lead to a constant amplitude of x_1 that is either an amplified or a suppressed version of its input.

The activation function $f(x) = \max(x, 0)$ is not continuously differentiable, but it is continuous in both space and time, so that contraction results can still be applied directly (Lohmiller & Slotine, 2000). Furthermore, the activation function is piecewise linear with a derivative of either 0 or 1. We exploit this property by inserting dummy terms l_j, which can be either 0 or 1 according to the derivative of $f(x)$: $l_j = \frac{d}{dx} f(x_j(t))$. In this case, all units are active, and thus $l_1 = l_2 = l_3 = 1$.

A system with Jacobian $\mathbf{J}$ is contracting if

$$\Theta \, \mathbf{J} \, \Theta^{-1} < 0, \tag{2.7}$$

where Θ is a constant transformation into an appropriate metric and $\mathbf{F} = \Theta \, \mathbf{J} \, \Theta^{-1}$ is the generalized Jacobian. If $\mathbf{F}$ is negative definite, the system is said to be contracting. We have shown previously (Rutishauser et al., 2011) how to choose the constant transformation Θ and conditions that guarantee contraction for a WTA circuit where all excitatory units provide direct input to the inhibitory unit (see Figure 1A). In summary, $\Theta = \mathbf{Q}^{-1}$, where $\mathbf{Q}$ is defined based on the eigendecomposition $\mathbf{J} = \mathbf{Q \Lambda Q}^{-1}$. In this case,

$$0 < \alpha < 2\sqrt{\beta_1\beta_2},$$
$$0 < \beta_1\beta_2 < 1, \tag{2.8}$$

guarantee contraction for any such WTA of any size (Rutishauser et al., 2011).

Structurally the two versions of the WTA are equivalent in that a unit was added in the pathway of recurrent inhibition, but no inhibition is added or removed (compare Figure 1A to Figure 2A). Thus, we can apply the same

constraints by replacing the product $\beta_1\beta_2$ with $\beta_1\beta_2\beta_3$ in equations 2.8. This product is equivalent to the inhibitory loop gain. This reduction is verified as follows. Using the notation shown in Figure 2C, assume that $T = 0$ for x_3, where $T > 0$ for the other units. Then,

$$\dot{x}_3 + x_3 = f(\beta_2 x_1),$$
$$\dot{x}_2 + x_2 = f(\beta_3 x_3 - T_2). \tag{2.9}$$

At steady state, $x_2 = f(\beta_2\beta_3 x_1)$, showing that x_3 and x_2 can be merged into a single unit x_2 by providing input of weight $\beta_2\beta_3$ directly to unit x_2 (see Figure 2D). The first key results of this letter are the following limits for contraction of a single such WTA (see Figure 2A):

$$0 < \alpha < 2\sqrt{\beta_1\beta_2\beta_3},$$
$$0 < \beta_1\beta_2\beta_3 < 1. \tag{2.10}$$

2.3.1 *Synchronizing Two WTAs.* Next, we show that connecting two WTAs in the manner illustrated in Figure 2C results in synchronization of the two inhibitory units, which in turn leads to the two WTAs merging into a single WTA. Note that by *synchronization*, we mean that the two variables have the same trajectory, or more generally that their difference is constant (in contrast to other meanings of synchronization, such as in population coding). The approach is to show that adding excitatory connections β_4 of sufficient strength will lead to the activity of the two inhibitory units x_2 and y_2 approaching a constant difference.

The Jacobian of the coupled system as shown in Figure 2C is

$$\mathbf{J}_C = \begin{bmatrix} \mathbf{J}_1 & \mathbf{D}_2 \\ \mathbf{D}_1 & \mathbf{J}_2 \end{bmatrix}, \tag{2.11}$$

with $\mathbf{J}_1 = \mathbf{J}_2 = \mathbf{J}$ (see equation 2.6) and

$$\tau\mathbf{D}_{1,2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & l_{2,5}\beta_4 \\ 0 & 0 & 0 \end{bmatrix}. \tag{2.12}$$

Following Pham and Slotine (2007) and Rutishauser et al. (2011), synchronization occurs exponentially if the following holds:

$$\mathbf{V}\mathbf{J}_C\mathbf{V}^T < \mathbf{0}, \tag{2.13}$$

where $\mathbf{V}$ defines an invariant subset of the system such that $\mathbf{V}\mathbf{x}$ is constant, with $\mathbf{x} = (x_1, x_2, x_3, y_1, y_2, y_3)$. Here we define synchrony as a regime where the differences between the inhibitory units $x_2 - y_2$ and between the interconnect units $x_3 - y_3$ are constant (although not necessarily zero). This results in

$$\mathbf{V} = \begin{bmatrix} 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}, \tag{2.14}$$

which embeds the two conditions.

Condition 2.13 is satisfied if

$$1 < \alpha$$

$$0 < \beta_4 < \beta_3 + 2, \tag{2.15}$$

$$\beta_3 < 2. \tag{2.16}$$

The conditions on the interconnect-weight β_4 guarantee that the dynamics are stable and that the inhibitory units synchronize. As long as $\beta_4 > 0$ is sufficiently small but nonzero, the inhibitory parts of the system will synchronize. Realistically, β_4 needs to be sufficiently large to drive the other inhibitory neuron above threshold and will thus be a function of the threshold T (see Rutishauser et al., 2011, equation 2.51). Here, synchrony is defined as their difference being constant. This in turn shows that the two WTAs have merged into a single WTA since the definition of a WTA is that each excitatory unit receives an equivalent amount of inhibition (during convergence but not necessarily afterward; see the simulations in section 2.6). This is our second key result.

2.3.2 Stability of Pairwise Combinations of WTAs. The final step of the stability analysis is a condition for the coupling strength $\beta_4 > 0$ such that the coupled system as shown in Figure 2C is contracting. The reasoning in this section assumes that the individual subsystems are contracting (as shown above).

The Jacobian of the combined system remains equation 2.11, where $\mathbf{J}_{1,2}$ are the Jacobians of the individual systems and $\mathbf{C}_{1,2}$ are the coupling terms. Rewriting the Jacobian of the second subsystem $\mathbf{J}_2$ by variable permutation $y_2' = y_3$ and $y_3' = y_2$ allows expression of the system in the form of

$$\mathbf{J}_C = \begin{bmatrix} \mathbf{J}_1 & \mathbf{D}_2 \\ \mathbf{D}_1 & \mathbf{J}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{E} \\ \mathbf{E}^T & \mathbf{J}_2 \end{bmatrix}, \tag{2.17}$$

where $\mathbf{E} = \mathbf{D}_1$ (see equation 2.12). This transformation of $\mathbf{J}_2$ is functionally equivalent to the original system (thus, its contraction limits remain), but it allows expression of the connection between the systems in the symmetric

form of equation 2.17.[1] This functionally equivalent system can now be analyzed using the approach that follows.

A matrix of the form

$$\begin{bmatrix} \mathbf{J}_1 & \mathbf{D} \\ \mathbf{D}^T & \mathbf{J}_2 \end{bmatrix}$$

is negative definite if the individual systems $\mathbf{J}_{1,2}$ are negative definite and if $\mathbf{J}_2 < \mathbf{D}^T \mathbf{J}_1^{-1} \mathbf{D}$ (Horn, 1985). Following Slotine (2003) and Rutishauser et al. (2011), this implies that a sufficient condition for contraction is $\sigma^2(\mathbf{D}) < \lambda(\mathbf{J}_1)\lambda(\mathbf{J}_2)$, where $\sigma(\mathbf{D})$ is the largest singular value of $\mathbf{D}$ and equivalent to β_4 in our case (all other elements of $\mathbf{D}$ are zero) and λ is the contraction rate of the individual subsystems. Since the two subsystems are equivalent, the contraction rates are also the same: $\lambda_1 = \lambda_2 = \lambda(\mathbf{J}_{1,2})$. It thus follows that the coupled systems are stable if $\beta_4 < \lambda_1$.

The contraction rate (Slotine, 2003; Rutishauser et al., 2011) of an individual subsystem is the absolute value of the largest eigenvalue of the Hermitian part of $\mathbf{F}$ (also see equation 2.7). It is $\lambda_{1,2} = |(\frac{1}{2}(-2 + \alpha))|$ for our system. Thus the condition for the two coupled systems to be contracting reduces to

$$\beta_4 < 1 - \frac{\alpha}{2}. \tag{2.18}$$

2.3.3 Summary of Boundary Conditions. In summary, the following conditions guarantee the stability of both the single and combined system, as well as hard competition between the two coupled systems (i.e., only one of the WTAs can have a winner). These conditions can be relaxed if $\alpha < 1$, which will permit a soft WTA:

$$1 < \alpha < 2\sqrt{\beta_1 \beta_2 \beta_3},$$
$$0 < \beta_1 \beta_2 \beta_3 < 1,$$
$$0 < \beta_4 < 1 - \frac{\alpha}{2}. \tag{2.19}$$

The lower bound on β_4 is from the synchronization analysis, whereas the upper bound is from the stability analysis. These results illustrate the critical

---

[1]The variable permutation is equivalent to a transformation of $\mathbf{J}_1$ by the permutation matrix $\mathbf{\Theta}$: $\mathbf{J}_2 = \mathbf{\Theta}\mathbf{J}_1\mathbf{\Theta}^{-1}$ with $\mathbf{\Theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

trade-off between having enough strength to ensure functionality, while being weak enough to exclude instability.

2.4 Speed of Winner Selection. How quickly will a system select a winner? For a single WTA, this question is answered by how quickly a single system contracts toward a winner and for a coupled system how quickly the two systems synchronize. One of the key advantages of contraction analysis is that the rate of contraction, and in this case the rate of synchronization, can be calculated explicitly. We will express the contraction and synchronization rate in terms of the time constants τ and its inverse, the decay constant. τ refers to the mean lifetime of exponential decay $x(t) = N_0 e^{-t\frac{1}{\tau}} = N_0 e^{-\lambda t}$. $\lambda = \frac{1}{\tau}$ is the decay constant. Both the contraction and the synchronization rate are expressed in the form of a decay constant λ. For example, the contraction rate of a system of the form $\dot{N} = -\lambda N$ is equivalent to λ.

Physiologically, the time constants τ in our system are experimentally determined membrane time constants that are typically in the range of 10 to 50 ms (Koch, 1998; McCormick, Connors, Lighthall, & Prince, 1985; Koch, Rapp, & Segev, 1996; Brown, Fricke, & Perkel, 1981). For simplicity, we assume that all excitatory and inhibitory units have the same, but different, time constants τ_E and τ_I, respectively. While the exact values depend on the cell type and state of the neurons, it is generally the case that $\tau_I < \tau_E$ due to the smaller cell bodies of inhibitory neurons (McCormick et al., 1985; Koch, 1998).

The bounds, equation 2.19, were calculated assuming equal time constants for all units. However, the same calculations yield to very similar bounds when assuming different time constants for inhibitory and excitatory units (Rutishauser et al., 2011). In this case, the ratio of the time constants $\frac{\tau_I}{\tau_E}$ becomes an additional parameter for the parameter bounds.

2.4.1 Speed of Synchronization. The synchronization rate is equivalent to the contraction rate of the system defined in equation 2.13 (Pham & Slotine, 2007), which is the absolute value of the maximal eigenvalue of the Hermitian part of $\mathbf{V}'(\mathbf{R}\mathbf{J}_C)\mathbf{V}'^T$. Here, the original τ^{-1} is replaced by the diagonal matrix $\mathbf{R}$, with the appropriate τ_E, τ_I terms on the diagonal.[2] The matrix $\mathbf{V}'$ is an orthonormal version of $\mathbf{V}$ as defined in equation 2.14, which here is simply $\mathbf{V}' = \sqrt{2^{-1}}\mathbf{V}$.

The resulting synchronization rate (sync rate) is a function of the weights β_3 (local inhibitory loop) and β_4 (remote inhibitory loop). We assume $\beta_4 \leq \beta_3$, which means that remote connectivity is weaker than local connectivity. However, qualitatively similar results can be found using the

---

[2]For the example of $\mathbf{J}_C$ (see equation 2.11), the diagonal terms are $\tau_E^{-1}, \tau_E^{-1}, \tau_I^{-1}, \tau_E^{-1}, \tau_E^{-1}, \tau_I^{-1}$.

opposite assumption. For $\tau_E = \tau_I$, the sync rate is

$$\lambda_s = \frac{1}{2\tau}(2 - \beta_3 + \beta_4). \tag{2.20}$$

Note the trade-off between local and remote connectivity: stronger remote connectivity increases and stronger local connectivity decreases the speed of synchronization (the larger λ, the quicker the system synchronizes). For approximately equal connectivity strength $\beta_3 \simeq \beta_4$ or in general for $\beta_3, \beta_4 \ll 1$, the sync rate is approximately τ^{-1}.

In general for $\tau_E \neq \tau_I$, the sync rate is $\lambda_s = \frac{\tau_E + \tau_I - \sqrt{\tau_E^2 - 2\tau_E\tau_I + (1+(\beta_3-\beta_4)^2)\tau_I^2}}{2\tau_E\tau_I}$. For example, for $\tau_I = \frac{\tau_E}{2}$, this reduces to $\lambda = \frac{1}{2\tau_I}(3 - \sqrt{1 + 4(\beta_3 - \beta_4)^2})$. Again, for $\beta_3 \simeq \beta_4$, the sync rate is approximately τ_I^{-1}. In conclusion, the sync rate is thus approximately equal to the contraction rate of the inhibitory units. Thus, synchronization occurs very quickly (20–50 ms for typical membrane time constants).

2.4.2 Speed of Contraction. The speed of selecting a winner (the contraction rate) for a single WTA can similarly be calculated based on the absolute value of the maximal eigenvalue of the Hermitian part of $\Theta(\mathbf{RJ})\Theta^{-1}$ (see equation 2.7).

Assuming $\tau = \tau_E = \tau_I$, the contraction rate is

$$\lambda_c = \frac{1}{2\tau}(2 - \alpha). \tag{2.21}$$

Note that the larger α, the longer it takes until the system converges. Qualitatively similar findings result for other ratios of τ_E and τ_I. For a typical value of $\alpha = 1.2$ (see simulations below) and $\tau = 20$ ms, the contraction rate would be 20 s^{-1}. This equals a halfway time (time constant) of $\lambda^{-1} = 50$ ms. For $\alpha = 1.5$, this would increase to 80 ms. The time it takes to find a winner is thus a multiple of the membrane time constant (in this example, 20 ms) and substantially slower than the time it takes to synchronize the network. In conclusion, synchronization is achieved first, which is then followed by winner selection.

2.5 Coupling More Than Two WTAs. So far we have shown how two different WTAs compete with each other after their inhibitory neurons are coupled. Similarly, more than two WTAs can compete with each other by all-to-all coupling of the inhibitory units, that is, every WTA is connected with two β_4 connections from and to every other WTA. Thus, the wiring complexity of this system scales as $O(M^2)$, where M is the number of WTAs in the system (note that M is not the number of units but the number of WTAs). Notice also that the all-to-all coupling concerns only the sparse

long-range excitatory connections, not the internal connectivity of the WTAs themselves.

The same principle can be used to embed hierarchies or sequences of competition. Thus, in a network of such WTAs, some WTAs could be in direct competition with each other, while others are not. Thus, for example, in a network of three WTAs A, B, and C relationships such as A competes with B and B competes with C are possible. In this case, A does not directly compete with C. So if A has a winner, C can also have a winner. If B has a winner, however, neither B nor C can have a winner (see Figures 4D–4F for a demonstration).

Regardless of how many WTAs are combined and whether all compete with all or more selectively, the stability of the aggregated system is guaranteed if the individual subsystems are stable and the coupling strengths β_4 observe the derived bounds. While in themselves combinations of stable modules have no reason to be stable, certain combinations (such as the one we use) of contracting systems are guaranteed to be stable (Slotine & Lohmiller, 2001). This is a key benefit of contraction analysis for the analysis of neural circuits.

2.6 Numerical Simulations. We simulated several cases of the network to illustrate its qualitative behavior. We used Euler integration with $\delta = 0.01$ s. The analytically derived bounds offer a wide range of parameters for which stability as well as function is guaranteed. For the simulations, we chose parameters that verify all bounds discussed.

First, we explored a simple system consisting of two WTAs with two possible winners each (see Figure 3). Parameters were $\alpha = 1.2$, $\beta_1 = 2$, $\beta_2 = 3$, $\beta_3 = \beta_4 = 0.1$, and $T = 0$. We found that any of the four possible winners can compete with each other regardless of whether they reside on the first or second WTA (Figures 3B–3D show an example). The inhibitory units quickly synchronized (see Figure 3C) their activity and reached the same steady-state amplitude (because $\beta_3 = \beta_4$).[3]

Second, we simulated a system with three WTAs using the same parameters (see Figure 4). For all-to-all coupling, all three WTAs directly compete with each other (see Figures 4A and 4B), that is, there can be only one winner across the entire system. Again, the inhibitory units all synchronize quickly during and after convergence (see Figure 4C). We also simulated the same system with more selective connectivity, eliminating competition between WTAs 1 and 3 (see Figure 4D). This arrangement allows either one winner if it is on WTA 2, or two winners if they are on WTAs 1 and 3. If the maximal activity is not on WTA 2, then the network permits two winning states. Otherwise, if the maximal input is on WTA 2, only one winner is

---

[3]If $\beta_3 = \beta_4$, it can be verified that $x_3(t) = y_4(t)$ for all $t > 0$ if the initial values at $t = 0$ are equal. Thus, the two inhibitory neurons become exactly equivalent in this special case.
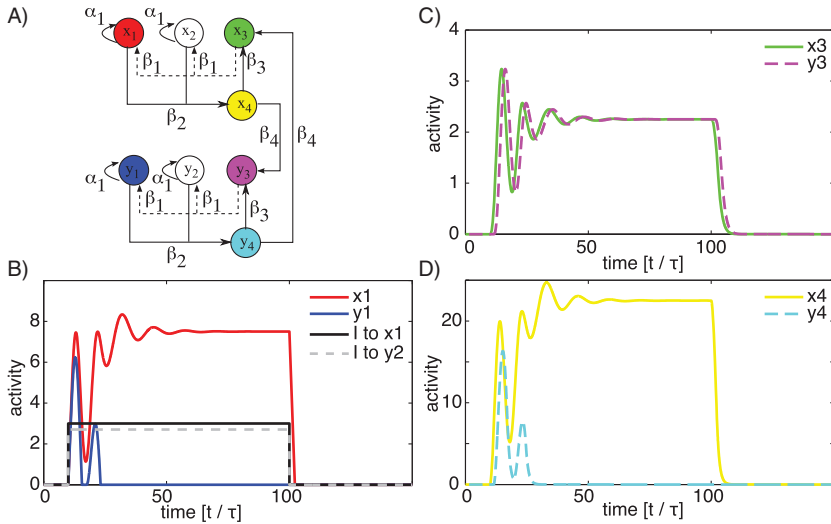
Figure 3: Simulation of merged WTA consisting of two WTAs with two excitatory units each (four possible winners). (A) Illustration of connectivity and notation (color code). (B) Activity as a function of time, for two excitatory units on two different WTAs, together with the external input provided to the same units. Notice how the network selects the winner appropriately. (C) Activity of the inhibitory units in each WTA. Note that y_3 is slightly delayed for plotting purposes. (D) Activity of the two interconnected units. Notice how the output of the losing WTA y_4 descends to zero after the competition has been resolved and the network has contracted. Units of time are multiples of the time constant τ. Notice that the same color indicates the same unit throughout this figure (notation is shown in A).

permitted (see Figure 4E for an illustration). This configuration allows partial competition.

3 Discussion

Neural circuits commonly depend on negative feedback loops. Such recurrent inhibition is a crucial element of microcircuits from a wide range of species and brain structures (Shepherd & Grillner, 2010) and enables populations of neurons to compute nonlinear operations such as competition, decision making, gain control, filtering, and normalization. However, when biologically realistic versions of such circuits are considered, additional factors such as wiring length, specificity, and complexity become pertinent. Here, we are principally concerned with the superficial layers of neocortex where the average distance of intracortical inhibitory connections is
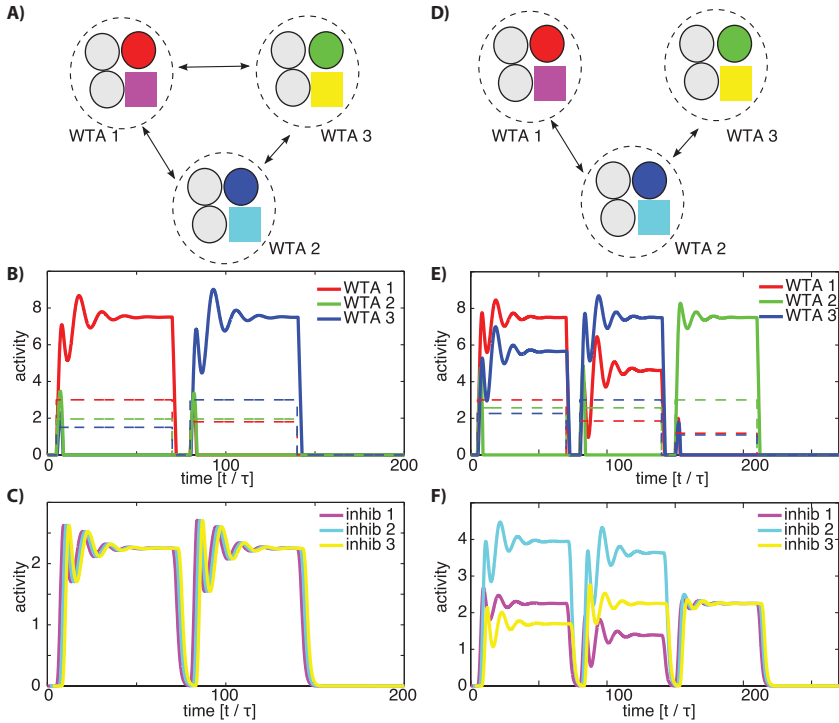
Figure 4: Simulation of merged WTA consisting of three WTAs with three possible winners each. (A–C) Case 1: Pairwise coupling allows all-to-all competition. (A) Illustration of connectivity. Filled circles are excitatory neurons, and filled rectangles inhibitory. The activity of units with colored fills is shown as a function of time. (B) Activity of one excitatory unit for each WTA (bold lines) and the external input to each (dashed lines). Notice that there can be only one winner among the three WTAs. (C) Activity of the inhibitory units, shifted in time to each other slightly for plotting only. (D–F) Case 2: selective coupling, allowing partial competition only between 1 and 2 and 2 and 3 but not 1 and 3. (E) Activity of the excitatory units for different cases. Notice that WTA 1 and 3, but not 2, can have winners simultaneously. Numbers indicate the WTA that the winner belongs to. (F) Activity of the inhibitory units, illustrating synchrony in the presence of different absolute amplitudes. Units of time are multiples of the time constant τ. Notice that the same color indicates the same unit throughout this figure (notation is shown in A).

typically much shorter than the excitatory connections (Bock et al., 2011; Binzegger et al., 2004; Perin, Berger, & Markram, 2011; Katzel et al., 2011; Adesnik & Scanziani, 2010). In contrast, in invertebrates, an inhibitory neuron has been identified that receives input from and projects back to all

Kenyon cells (which are excitatory) (Papadopoulou, Cassenaer, Nowotny, & Laurent, 2011). This neuron has been demonstrated to perform response normalization, making this system a direct experimental demonstration of competition through shared inhibition. No such system has yet been identified in the cortex.

The number of excitatory neurons that can be contacted by an inhibitory neuron thus poses a limit on how many excitatory neurons can compete directly with one another (in terms of numbers and distance). Other models, such as those based on Mexican hat–type inhibitory surrounds (Hertz et al., 1991; Willshaw & von der Malsburg, 1976; Soltani & Koch, 2010), even require that inhibitory connectivity be longer range than the excitatory. These anatomical constraints have been used to argue that models such as the WTA are biologically unrealistic and therefore of limited use.

We have demonstrated here, by theoretical analysis and simulation, that it is possible to extend such circuits by merging several independent circuits functionally through synchronization of their inhibitory interneurons. This extension allows the construction of large, spatially distributed circuits that are composed of small pools of excitatory units that share an inhibitory neuron. We have applied and proved by nonlinear contraction analysis that systems combined in this manner are inherently stable and that arbitrary aggregation by inhibitory synchrony of such subsystems results in a stable system. This composition of subcircuits removes the limits on maximal circuit size imposed by anatomical wiring constraints on inhibitory connectivity because the synchrony between local inhibitory neurons is achieved entirely by excitatory connectivity, which can possibly be long range, permitting competition between excitatory units that are separated by long distances, for example, in different cortical areas. We show that the time necessary to achieve sychronization is much shorter than the time required to select a winner. Thus, synchronization is faster than winner selection, which can thus proceed robustly across long-range connections that enforce synchronization. Further, selective synchronization among some WTAs but not others allows partial competition between some but not other WTAs (see Figure 4). The strength of these long-range connections could be modulated dynamically to enable or disable various competitions between two populations conditional on some other brain state. This modulation could be implemented by a state-dependent routing mechanism (Rutishauser & Douglas, 2009).

There are several possibilities for mapping the abstract units in our model to real physiological neurons. Our units are mean-rate approximations of a small group of neurons. In terms of intracortical inhibition, these would lie anatomically close to each other within superficial layers of neocortex. Since such inhibitory connectivity would have only limited reach, each inhibitory subunit can enforce competition only across a limited number of close-by excitatory units. Competition between different areas is made possible by synchronizing remote populations by long-range excitatory mechanisms

in the way we propose. Direct long-range inhibition, on the other hand, is unlikely both intracortically and subcortically, since all known connections from the thalamus and basal ganglia to cortex are excitatory. Networks such as the LEGION network (Wang & Terman, 1995) assume global inhibitory input to all excitatory units in the network, which, for the reasons we discuss, is unlikely in the case of cortex. It would, however, be possible to implement a feasible version of the global inhibitory input by synchronizing many local inhibitory neurons using the mechanism we describe, resulting in an anatomically realistic version of the LEGION network.

Functionally, the model presented here makes several testable predictions. Consider a sensory area with clearly defined features as possible winners, such as orientations. The model predicts that the inhibitory units would not be tuned to these features, particularly if the number of possible winners is large. This is because the connectivity to the inhibitory units is not feature specific. Experimental studies indicate that this is indeed the case: units that functionally represent different tuning project to the same inhibitory unit, resulting in untuned inhibitory activity (Bock et al., 2011; Fino & Yuste, 2011; Kerlin, Andermann, Berezovskii, & Reid, 2010; Kuhlman, Tring, & Trachtenberg, 2011; Hofer et al., 2011). Second, this model predicts that inhibitory activity between two different areas or parts of the same area can either be highly synchronous or completely decoupled depending on whether the two are competing or functioning independently. This thus predicts that synchrony of inhibitory units should be affected by manipulations that manipulate competition, such as top-down attention.

Our model suggests that synchronized populations of inhibitory neurons are crucial for enforcing competition across several subpopulations of excitatory neurons. It further suggests that the larger the number and spatial distribution of such synchronized inhibitory units, the larger the number of units that compete with each other. Experimentally, synchronized modulation of inhibitory neurons is a common phenomenon that is believed to generate the prominent gamma rhythm triggered by sensory stimulation in many areas (Fries, Nikoli, & Singer, 2007; Whittington, Traub, & Jefferys, 1995; Traub, Whittington, Stanford, & Jefferys, 1996). Recent experiments have used stimulation of inhibitory neurons (Cardin et al., 2009; Sohal, Zhang, Yizhar, & Deisseroth, 2009; Szucs, Huerta, Rabinovich, & Selverston, 2009) to increase or decrease their synchronization with direct observable effects on nearby excitatory neurons such as, for example, increased or decreased amplitude and precision of evoked responses relative to how strongly the inhibitory neurons were synchronizing. Note that our proposal for this function of inhibitory synchrony is distinct and independent from the proposal that gamma-band synchrony serves to increase readout efficacy by making spikes arrive coincidentally from a large number of distributed sources (Tiesinga, Fellous, & Sejnowski, 2008; Singer & Gray, 1995). Here, we propose that an additional function of such synchrony is to allow select populations of excitatory neurons to compete with each other because each receives inhibition at the same time.

Appendix: Contraction Analysis ——————————————————

This appendix provides a short summary of contraction analysis. We have previously published the detailed methods of applying contraction theory to WTA circuits (Rutishauser et al., 2011). Essentially, a nonlinear time-varying dynamic system will be called contracting if arbitrary initial conditions or temporary disturbances are forgotten exponentially fast, that is, if trajectories of the perturbed system return to their unperturbed behavior with an exponential convergence rate. A relatively simple algebraic condition can be given for this stability-like property to be verified, and this property is preserved through basic system combinations and aggregations.

A nonlinear contracting system has the following properties (Lohmiller & Slotine, 1998, 2000; Slotine, 2003; Wang & Slotine, 2005):

- Global exponential convergence and stability are guaranteed.
- Convergence rates can be explicitly computed as eigenvalues of well-defined Hermitian matrices.
- Combinations and aggregations of contracting systems are also contracting.
- Robustness to variations in dynamics can be easily quantified.

Before stating the main contraction theorem, recall first the following properties: The symmetric part of a matrix $\mathbf{A}$ is $\mathbf{A}_H = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{*T})$. A complex square matrix $\mathbf{A}$ is Hermitian if $\mathbf{A}^T = \mathbf{A}^*$, where ^{T} denotes matrix transposition and * complex conjugation. The Hermitian part $\mathbf{A}_H$ of any complex square matrix $\mathbf{A}$ is the Hermitian matrix $\frac{1}{2}(\mathbf{A} + \mathbf{A}^{*T})$. All eigenvalues of a Hermitian matrix are real numbers. A Hermitian matrix $\mathbf{A}$ is said to be positive definite if all its eigenvalues are strictly positive. This condition implies in turn that for any nonzero real or complex vector, $\mathbf{x}$, $\mathbf{x}^{*T}\mathbf{A}\mathbf{x} > 0$. A Hermitian matrix $\mathbf{A}$ is called negative definite if $-\mathbf{A}$ is positive definite.

A Hermitian matrix $\mathbf{A}(\mathbf{x}, t)$ dependent on state or time will be called uniformly positive definite if there exists a strictly positive constant such that for all states $\mathbf{x}$ and all $t \geq 0$, the eigenvalues of $\mathbf{A}(\mathbf{x}, t)$ remain larger than that constant. A similar definition holds for uniform negative definiteness.

Consider now a general dynamical system in $\mathbb{R}^n$,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t), \tag{A.1}$$

with $\mathbf{f}$ a smooth nonlinear function. The central result of contraction analysis, derived in Lohmiller and Slotine (1998) in both real and complex forms, can be stated as a theorem.

Theorem. *Denote by $\frac{\partial f}{\partial x}$ the Jacobian matrix of $\mathbf{f}$ with respect to $\mathbf{x}$. Assume that there exists a complex square matrix $\boldsymbol{\Theta}(x, t)$ such that the Hermitian matrix*

$\boldsymbol{\Theta}(x, t)^{*T}\boldsymbol{\Theta}(x, t)$ *is uniformly positive definite, and the Hermitian part* $\boldsymbol{F}_H$ *of the matrix,*

$$F = \left(\dot{\boldsymbol{\Theta}} + \boldsymbol{\Theta}\frac{\partial f}{\partial x}\right)\boldsymbol{\Theta}^{-1},$$

is uniformly negative definite. Then all system trajectories converge exponentially to a single trajectory, with convergence rate $|sup_{x,t}\lambda_{max}(\boldsymbol{F}_H)| > 0$. *The system is said to be contracting,* $\boldsymbol{F}$ *is called its generalized Jacobian, and* $\boldsymbol{\Theta}(x, t)^{*T}\boldsymbol{\Theta}(x, t)$ *its contraction metric. The contraction rate is the absolute value of the largest eigenvalue (closest to zero, although still negative)* $\lambda = |\lambda_{max}\boldsymbol{F}_H|$.

In the linear time-invariant case, a system is globally contracting if and only if it is strictly stable, and **F** can be chosen as a normal Jordan form of the system, with **Θ** a real matrix defining the coordinate transformation to that form (Lohmiller & Slotine, 1998). Alternatively, if the system is diagonalizable, **F** can be chosen as the diagonal form of the system, with **Θ** a complex matrix diagonalizing the system. In that case, $\mathbf{F}_H$ is a diagonal matrix composed of the real parts of the eigenvalues of the original system matrix.

References

Adesnik, H., & Scanziani, M. (2010). Lateral competition for cortical space by layer-specific horizontal circuits. *Nature*, *464*, 1155–1160.

Amari, S., & Arbib, M. (1977). Competition and cooperation in neural nets. In J. Metzler (Ed.), *Systems neuroscience* (pp. 119–165). San Diego, CA: Academic Press.

Binzegger, T., Douglas, R. J., & Martin, K. A. (2004). A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.*, *24*(39), 8441–8453.

Bock, D. D., Lee, W. C., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., et al. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, *471*, 177–182.

Brown, T. H., Fricke, R. A., & Perkel, D. H. (1981). Passive electrical constants in three classes of hippocampal neurons. *Journal of Neurophysiology*, *46*(6), 812–827.

Cardin, J. A., Carlen, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., et al. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, *459*, 663–667.

Coultrip, R., Granger, R., & Lynch, G. (1992). A cortical model of winner-take-all competition via lateral inhibition. *Neural Networks*, *5*, 47–54.

Douglas, R., Koch, C., Mahowald, M., Martin, K., & Suarez, H. (1995). Recurrent excitation in neocortical circuits. *Science*, *269*(5226), 981–985.

Douglas, R., & Martin, K. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, *27*, 419–451.

Fino, E., & Yuste, R. (2011). Dense inhibitory connectivity in neocortex. *Neuron*, *69*, 1188–1203.

Fries, P., Nikoli, D., & Singer, W. (2007). The gamma cycle. *Trends Neurosci.*, *30*, 309–316.

Hahnloser, R., Sarpeshkar, R., Mahowald, M., Douglas, R., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, *405*(6789), 947–951.

Hertz, J., Krogh, A., & Palmer, R. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.

Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., et al. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, *14*, 1045–1052.

Horn, R. (1985). *Matrix analysis*. Cambridge: Cambridge University Press.

Katzel, D., Zemelman, B. V., Buetfering, C., Wolfel, M., & Miesenbock, G. (2011). The columnar and laminar organization of inhibitory connections to neocortical excitatory cells. *Nature Neuroscience*, *14*, 100–107.

Kerlin, A. M., Andermann, M. L., Berezovskii, V. K., & Reid, R. C. (2010). Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron*, *67*(5), 858–871.

Koch, C. (1998). *Biophysics of computation: Information processing in single neurons*. New York: Oxford University Press.

Koch, C., Rapp, M., & Segev, I. (1996). A brief history of time (constants). *Cerebral Cortex*, *6*(2), 93–101.

Kuhlman, S. J., Tring, E., & Trachtenberg, J. T. (2011). Fast-spiking interneurons have an initial orientation bias that is lost with vision. *Nat. Neurosci.*, *14*, 1121–1123.

Lohmiller, W., & Slotine, J. (1998). On contraction analysis for nonlinear systems. *Automatica*, *34*(6), 683–696.

Lohmiller, W., & Slotine, J. (2000). Nonlinear process control using contraction theory. *AIChE Journal*, *46*(3), 588–596.

Maass, W. (2000). On the computational power of winner-take-all. *Neural Computation*, *12*, 2519–2536.

McCormick, D. A., Connors, B. W., Lighthall, J. W., & Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, *54*(4), 782–806.

Neftci, E., Chicca, E., Indiveri, G., Cook, M., & Douglas, R. J. (2010). State-dependent sensory processing in networks of VLSI spiking neurons. In *IEEE International Symposium on Circuits and Systems*. Piscataway, NJ: IEEE.

Papadopoulou, M., Cassenaer, S., Nowotny, T., & Laurent, G. (2011). Normalization for sparse encoding of odors by a wide-field interneuron. *Science*, *332*, 721–725.

Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci. U.S.A.*, *108*, 5419–5424.

Pham, Q., & Slotine, J. (2007). Stable concurrent synchronization in dynamic system networks. *Neural Networks*, *20*(1), 62–77.

Rabinovich, M. I., Huerta, R., Volkovskii, A., Abarbanel, H.D.I., Stopfer, M., & Laurent, G. (2000). Dynamical coding of sensory information with competitive networks. *Journal of Physiology-Paris*, *94*(5–6), 465–471.

Rutishauser, U., & Douglas, R. (2009). State-dependent computation using coupled recurrent networks. *Neural Computation*, *21*(2), 478–509.

Rutishauser, U., Douglas, R., & Slotine, J. (2011). Collective stability of networks of winner-take-all circuits. *Neural Computation*, *23*(3), 735–773.

Shepherd, G. M., & Grillner, S. (2010). *Handbook of brain microcircuits*. New York: Oxford University Press.

Shepherd, G. M., Stepanyants, A., Bureau, I., Chklovskii, D., & Svoboda, K. (2005). Geometric and functional organization of cortical circuits. *Nat. Neurosci.*, *8*(6), 782–790.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.*, *18*, 555–586.

Slotine, J. (2003). Modular stability tools for distributed computation and control. *International Journal of Adaptive Control and Signal Processing*, *17*, 397–416.

Slotine, J., & Lohmiller, W. (2001). Modularity, evolution, and the binding problem: A view from stability theory. *Neural Networks*, *14*, 137–145.

Sohal, V. S., Zhang, F., Yizhar, O., & Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature*, *459*, 698–702.

Soltani, A., & Koch, C. (2010). Visual saliency computations: Mechanisms, constraints, and the effect of feedback. *Journal of Neuroscience*, *22*, 12831–12843.

Szucs, A., Huerta, R., Rabinovich, M. I., & Selverston, A. I. (2009). Robust microcircuit synchronization by inhibitory connections. *Neuron*, *61*, 439–453.

Tiesinga, P., Fellous, J. M., & Sejnowski, T. J. (2008). Regulation of spike timing in visual cortical circuits. *Nat. Rev. Neurosci.*, *9*, 97–107.

Traub, R. D., Whittington, M. A., Stanford, I. M., & Jefferys, J. G. (1996). A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature*, *383*, 621–624.

Wang, D., & Terman, D. (1995). Locally excitatory globally inhibitory oscillator networks. *IEEE Transactions on Neural Networks*, *6*(1), 283–286.

Wang, W., & Slotine, J. (2005). On partial contraction analysis for coupled nonlinear oscillators. *Biological Cybernetics*, *91*(1), 38–53.

Whittington, M. A., Traub, R. D., & Jefferys, J. G. (1995). Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature*, *373*, 612–615.

Willshaw, D. J., & von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proc. R. Soc. Lond., B, Biol. Sci.*, *194*, 431–445.

Yabuta, N., & Callaway, E. M. (1998). Cytochrome-oxidase blobs and intrinsic horizontal connections of layer 2/3 pyramidal neurons in primate V1. *Vis. Neuroscience*, *15*, 1007–1027.

Yuille, A., & Geiger, D. (2003). Winner-take-all networks. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 1228–1231). Cambridge, MA: MIT Press.