7 Winner-Take-All (WTA) Network

Samuel Kaski and Teuvo Kohonen

In the practical SOM algorithms, selection of the winner by arithmetic computation is no problem. However, as the biological neural networks must implement this computation by dynamical components and networks with simple structures, special solutions compatible with the real neurophysiology must then be sought.
The winner index c in the biologically motivated SOM was defined by

$$c = \arg\max_i \{\mathbf{m}_i^{\mathrm{T}}\mathbf{x}\} \ . \tag{59}$$

In modeling, the dot products $\mathbf{m}_i^{\mathrm{T}}\mathbf{x}$ correspond to the total postsynaptic activations I_i of the neurons. They are formed directly at the inputs of the neurons. Therefore, it will remain necessary to study under what conditions a physiologically plausible simple network structure can select the largest of its scalar inputs (activations), i.e., implement the winner selection. Such a circuit is called the "winner-take-all" (WTA) network. For an early approach to this problem, cf. [1,2].
Our analysis is potentially applicable to any network in which the connections coming to each neuron can be grouped into external input, self-feedback, and feedback from the other neurons within the network (Figure 4). We used a neuron model introduced earlier [5], which describes changes in the activity, averaged spiking frequency η, of a cell as a function of the external inputs to the cell, I, and a nonlinear *convex* loss function γ:

$$d\eta/dt = I - \gamma(\eta) \ . \tag{60}$$

The nonlinear loss function represents the resultant of all losses and the effect of the refractory time of the cell. (To be exact, equation 60 holds only when $\eta > 0$ or when the right-hand side is positive, since spiking activity must always be positive.)
In the simplest network that we analyze, the input to neuron i consists of the external input coming from outside of the network, I_i^e, self-feedback from the neuron to itself, $g^+\sigma(\eta_i)$, and the feedback from the other cells, $g^- \sum_k \sigma(\eta_k)$. Here g^+ and g^- are coefficients that determine the strength of the connections, and σ models the combined effects of the transfer functions of the possible interneurons and any saturating nonlinearities on the signals. The dynamical system formed of the neural network can be described with the following set of differential equations:

$$d\eta_i/dt = I_i^e + g^+\sigma(\eta_i) + g^- \sum_k \sigma(\eta_k) - \gamma(\eta_i) \ , \tag{61}$$

$i = 1, \ldots, N$, where N is the number of neurons in the network.
This simple network type had already been analyzed previously [6], but now it turned out that the analysis could be generalized [4] to networks with several types of even nonidentical feedback connections (interneurons). To make the analysis most general the system of differential equations generalized from (61) was dressed mathematically into the form of a certain class of dynamical systems,

$$dy_i/dt = \lambda(y_i)[a_i(y_i) + b(y)] \ , \tag{62}$$

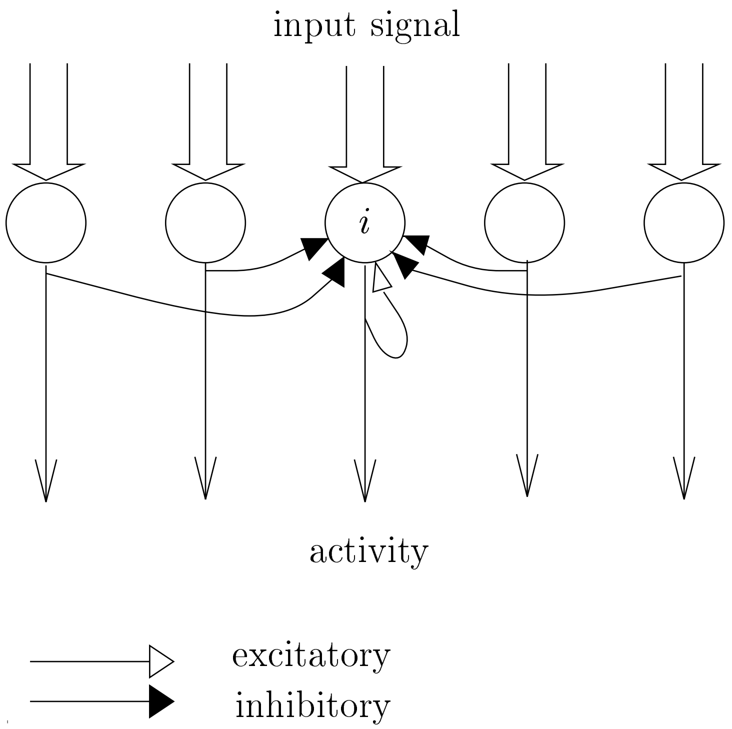


Figure 4: A schematic winner-take-all network. The neurons compete through the negative (inhibitory) feedback connections. The neuron receiving the largest input will be the only neuron that remains active after the initial transient activity. Only the connections coming to neuron i are shown.

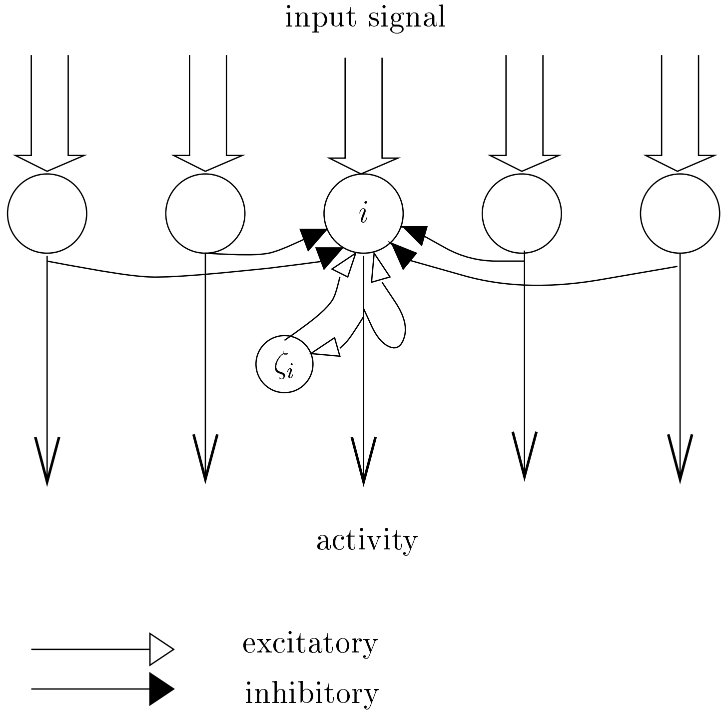
where λ, a_i, and b are certain functions, and y is a vector formed of all the state variables y_i. Convergence properties of these types of systems had already been analyzed in [3].

It was then possible to prove that if certain restrictions are placed on the functions λ, a_i, and b, only one of the state variables y_i remains above a threshold, whereas the rest of them remain below a lower threshold. The lower threshold is zero for the neuron models. This is the essence of any WTA function.

When this more general analysis [4] was applied to the more general neural network models, conditions under which the networks have the WTA property were obtained. The most important conditions concern the external input: one of the neurons must receive the largest input, and all the other inputs must lie within a sensible range that does not depend on the largest input; otherwise also some other neurons may become active. In the beginning of the competition the winner must also be at least as active as the other neurons, which is the case, e.g., in the more complete network model described later in this section. The other, very mild conditions concern the form and steepness of the loss-function γ and the conductance function σ, and the strengths of the feedback connections g^+ and g^-.

The essential novelty in our analyses was their generality. Nevertheless, the models incorporated the common assumption that "sigmoidal"-type nonlinear transfer functions (function σ in 61) are adequate for modeling the effects of interneurons. It was possible, however, to further generalize the analyses by modeling the interneurons explicitly; the system of differential equations (61) then includes one extra equa-

a



b

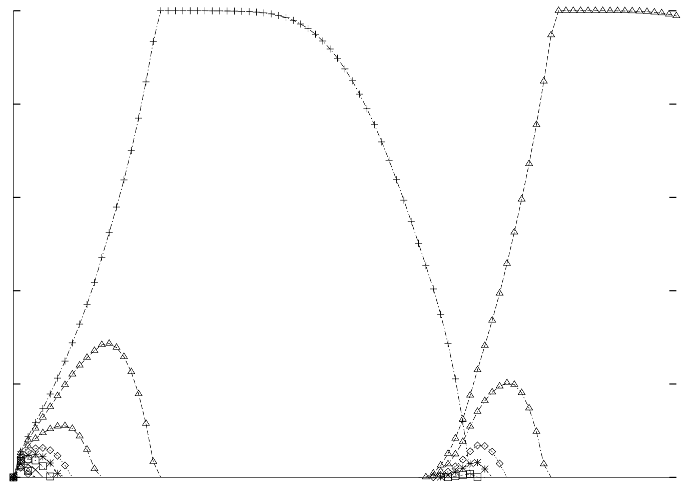


Figure 5: **a** Auxiliary slower inhibitory interneurons (marked with ζ_i in the schematic network) inactivate the active neuron after a brief interval, whereafter the competition may start again. If the inputs have changed meanwhile the previous winner was active, the new winner will be the one receiving the largest input. Otherwise the "runner-up", the neuron receiving the second-largest input will win. **b** A sample period of the activities of the neurons in a 20-neuron network.

tion for each interneuron. It is not possible, however, to guarantee in general that such models converge, but we were able to give general conditions under which the convergent models are WTA networks.

The WTA networks in which the winner *remains* active after it has become active are of course not sufficient models of the activity in physical neural networks. We coined such networks *weak* WTA circuits. In practice a network must be able to follow the changing activity it receives — we called networks in which a new unit becomes active when the inputs change *strong* WTA networks. We have demonstrated that the networks we studied are strong WTA networks if there are certain auxiliary slower interneurons in the network. These neurons provide negative feedback that in effect resets the activity of the winning neuron after a certain period of time (Figure 5).

We may summarize our analyses by concluding that the network structure of the type schematized in Figure 4 has been shown to be very robust in implementing the competition that is a necessary precursor of the Self-Organizing Map, or in fact any competitive learning application.

References

[1] R. L. Didday. *The Simulation and Modeling of Distributed Information Processing in the Frog Visual System.* PhD thesis, Stanford University, 1970.

[2] R. L. Didday. A model of visuomotor mechanisms in the frog optic tectum. *Mathematical Biosciences*, 30:169–180, 1976.

[3] S. Grossberg. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, L11:213–257, 1973.

[4] S. Kaski and T. Kohonen. Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, 7:973-984, 1994.

[5] T. Kohonen. An introduction to neural computing. *Neural Networks*, 1:3–16, 1988.

[6] T. Kohonen. Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, 6:895–905, 1993.