LETTER Communicated by Richard Lippmann

Biophysiologically Plausible Implementations of the Maximum Operation

Angela J. Yu
feraina@gatsby.ucl.ac.uk
Gatsby Computational Neuroscience Unit, University College London, London WC1N 3AR, U.K.

Martin A. Giese
martin.giese@tuebingen.mpg.de
Department for Cognitive Neurology, University Clinic Tübingen, 72076 Tübingen, Germany

Tomaso A. Poggio
tp@ai.mit.edu
CBCL, Massachusetts Institute of Technology, Cambridge, MA 02142, U.S.A.

Visual processing in the cortex can be characterized by a predominantly hierarchical architecture, in which specialized brain regions along the processing pathways extract visual features of increasing complexity, accompanied by greater invariance in stimulus properties such as size and position. Various studies have postulated that a nonlinear pooling function such as the maximum (MAX) operation could be fundamental in achieving such selectivity and invariance. In this article, we are concerned with neurally plausible mechanisms that may be involved in realizing the MAX operation. Different canonical models are proposed, each based on neural mechanisms that have been previously discussed in the context of cortical processing. Through simulations and mathematical analysis, we compare the performance and robustness of these mechanisms. We derive experimentally verifiable predictions for each model and discuss the relevant physiological considerations.

1 Introduction

Neurophysiological experiments have provided evidence that visual processing in the cortex can be characterized by a predominantly hierarchical system, in which areas farther along the ventral pathway are selective for increasingly complex stimulus features, accompanied by increasing invariance with respect to stimulus size and position (Hubel & Wiesel, 1962; Perrett et al., 1991; Logothetis, Pauls, & Poggio, 1995; Tanaka, 1996; Pasupathy & Connor, 1999).

Different mechanisms have been proposed to account for the selectivity and invariance properties of the visual cortex. For instance, one body of theoretical work involves flexible central mechanisms that dynamically adjust stimulus scale and position selectivity according to the input (e.g., the "shifter circuit" presented in Anderson & Van Essen, 1987). However, although there is evidence for such dynamic modulation in a variety of visual areas (Motter, 1994; Connor, Preddie, Gallant, & Van Essen, 1997; Treue & Maunsell, 1996), it functions at a timescale too slow to account for the short latencies found in some object recognition tasks (Thorpe, Fize, & Marlot, 1996). From a different approach, a growing number of models of visual processing call for some form of pooling from feature detectors in an earlier stage of processing. The underlying idea was first postulated by Hubel and Wiesel (1962), to account for complex cell invariance with respect to spatial phase shifts via linear summation of responses of rectifying, phase-sensitive neurons. Fukushima's (1980) more general, hierarchical Neocognitron network achieves invariant responses via a feedforward network of alternating layers of feature detectors and nonlinear pooling neurons. More recently, Riesenhuber and Poggio (1999a, 1999b) have reproduced neurophysiological data from area IT with a hierarchical model that uses a combination of linear summation and maximum (MAX) operations.

Pooling by a MAX operation, as opposed to linear summation, achieves high feature specificity and invariance simultaneously (Riesenhuber & Poggio, 1999b). Suppose the inputs to the system are activity levels of a population of simple cell bar detectors that prefer the same orientation but have receptive fields in different locations. Summing from these detectors gives the same total response if the input is an oriented bar contained in any one of the receptive fields, achieving position invariance. However, the response is even stronger if multiple bars (e.g., a grating) or background clutter is present, causing the pooling neuron to lose selectivity as a bar detector. Taking the maximum of the responses of these feature detectors alleviates this specificity problem, because then the system output is solely determined by the response of its most active afferent. Thus, the MAX operation both preserves feature specificity and achieves invariance in a more robust fashion than linear summation. It is interesting to note that pooling by the MAX operation is computationally equivalent to scanning an image with a template, which has been the basis for many recognition algorithms in computer vision (Riesenhuber & Poggio, 1999b, 2000).

In this study, we are concerned with neurophysiological implementations of the MAX operation. In addition to its proposed involvement in a variety of cortical processes such as object recognition (Riesenhuber & Poggio, 1999b), motion recognition (Giese, 2000), and visual velocity estimation (Grzywacz & Yuille, 1990), the MAX operation is interesting as a basic nonlinear operator that can be implemented by simple, neurophysiologically plausible models. Throughout this article, we define the ideal MAX oper-

ation as a mapping from an input vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ to an output signal z, where

$$z \propto x_m \equiv \max_{1 \le i \le n} x_i.$$

More generally, the operation should achieve the following properties:

- **Selectivity:** The output signal z depends only on the maximum of all the input signals (sometimes referred to as input amplitude in this work), x_m, and not on the other values.
- **Linearity:** The output signal z depends linearly on x_m with a constant gain factor g, that is, $z = gx_m$.

The first property is critical to achieving feature specificity. The second property is important for optimally recovering information about the strength of the maximal input. Of course, biological systems can be expected to implement only an approximation to the ideal MAX operation. In some computation models employing the MAX operation, it has been shown that an approximation to the ideal MAX operation indeed suffices (Riesenhuber & Poggio, 1999b).

Both the computational and implementation aspects of our models derive much inspiration from two closely related areas of earlier work on neural modeling: winner-takes-all (WTA) and gain control networks.

WTA has been widely studied in the neural networks and VLSI literature (Grossberg, 1973; Kohonen, 1995; Lazzaro, Ryckenbusch, Mahowald, & Mead, 1989; Starzyk & Fang, 1993; Hahnloser, Douglas, Mahowald, & Hepp, 2000; Fukai & Tanaka, 1997) and has been used to model cortical functions such as the integration of component motions (Nowlan & Sejnowski, 1995) and attentional selection (Koch & Ullman, 1985; Lee, Itti, Koch, & Braun, 1999). WTA networks select the afferent input with the largest amplitude, but their output is not required to reflect this amplitude and is often nonlinear or even binary. In general, WTA networks convey the identity of the "winner neuron" but not its precise amplitude to the downstream cortical processing areas[1], whereas the MAX operation communicates the amplitude but not necessarily the identity. Gain control circuits make the response of neural detectors independent of stimulus energy or contrast, thus exhibiting some invariance in stimulus size or intensity (Reichardt, Poggio, & Hausen, 1983; Carandini & Heeger, 1994; Wilson & Humanski, 1993; Simoncelli & Heeger, 1998; Salinas & Abbott, 1996). However, typical

[1] There are some exceptions: Yuille and Grzywacz's (1989) divisive feedback network and Hahnloser's (1998) linear threshold VLSI network both output the maximal input under certain conditions, and Lazzaro et al.'s (1989) VLSI network outputs the logarithm of the maximal input. These implementations sit on the definitional boundary between MAX and WTA networks.

gain control circuits fail the linearity requirement because they normalize all input channels regardless of input amplitude.

The remainder of the article is structured as follows. In section 2, we isolate a small number of neural mechanisms that may be involved in realizing the MAX operation and present four highly simplified canonical circuits that implement these mechanisms. In section 3, we present our main simulation results, with the relevant mathematical analysis to be found in the appendix. In section 4, a number of model-specific predictions are presented, along with a discussion on potential experimental paradigms that can be used to verify these predictions. In section 5, we discuss various neuronal implementations of the computational operations involved in each of the models and compare the relative plausibility of these implementations. Finally, section 6 relates our efforts to previous work in neural modeling and makes suggestions for future directions of research. Some of the results have appeared previously in abstract form (Yu, Giese, & Poggio, 2000a, 2000b).

2 Models

The main issues we explore in this work are divisive[2] versus subtractive inhibition,[3] feedforward versus recurrent architecture, and mean firing rate versus integrate-and-fire description.

We focus on four simple canonical neural models that implement the MAX operation with well-studied neural principles and which allow some degree of mathematical analysis (see Figure 1 for schematic diagrams of a generic feedforward and feedback network). All of these models can be described as a three-layer neural circuit with an input layer representing static input signals x_n, a symmetrically connected intermediate layer that transforms the input signals into output signals y_n in a nonlinear fashion, and an output unit z that linearly sums the intermediate-layer activities. In biophysiological terms, the inputs correspond to output signals from earlier stages of sensory processing. If these earlier feature detectors

[2] The physiological mechanisms and functional role of shunting inhibition have been a topic of intensive theoretical and experimental investigations (Naka & Rushton, 1966; Torre & Poggio, 1978; Koch, Poggio, & Torre, 1983; Koch & Poggio, 1987; Ferster & Jagadesh, 1992; Holt & Koch, 1997; Carandini, Heeger, & Movshon, 1997; Doiron, Longtin, Berman, & Maler, 2001; Borg-Graham, Monier, & Fregnac, 1998; Anderson, Carandini, & Ferster, 2000). The latest evidence on the role of shunting inhibition in the visual cortex converges on the observation that inhibitory synaptic conductances can rise significantly and rapidly after stimulus presentation, having a divisive effect on the excitatory synaptic input (Borg-Graham et al., 1998; Anderson et al., 2000).

[3] Intracellular recordings show that half-wave rectification provides a good fit for intracellular recording data (Carandini & Ferster, 2000).

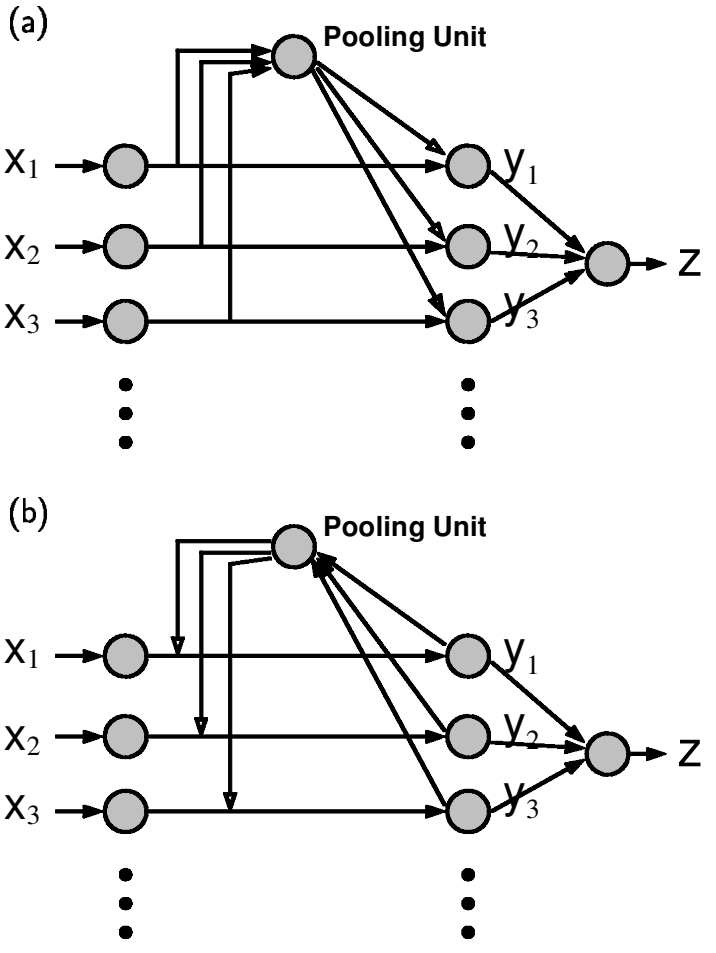

Figure 1: Schematic feedforward and feedback models. (a) Feedforward network. (b) Feedback network. Solid arrow: excitatory inputs. Open arrow: inhibitory inputs. Circles and lines represent computational units and interactions rather than explicit delineation of neurons and their processes. Apparent violations of Dale's law can be resolved via inhibitory interneurons (Li, 2000), as seems to be done in the cortex (White, 1989; Gilbert, 1992; Rockland & Lund, 1983).

collectively have a spectrum of selectivity in a certain feature dimension (e.g., stimulus position), then the output signal would reflect a degree of invariance in this dimension. For example, preliminary data indicate simple and complex cells may play respective roles of x_n and z in implementing the MAX operation in V1 (Lampl, Riesenhuber, Poggio, & Ferster, 2001). A

dedicated summating interneuron receiving inputs from a number of simple cells could inhibit their inputs to intermediate-layer neurons or directly inhibit the dendrites of the complex cell receiving inputs from the simple cells. A dedicated summating unit might not even be necessary if each simple cell (or its downstream interneuron) inhibits the synaptic inputs from the other simple cells to the same complex cell.

The first model is a feedforward network (FFN) with divisive (shunting) inhibition:

$$\begin{aligned} y_n &= \frac{x_n f(x_n)}{c + \sum_k f(x_k)} \\ z &= \sum_n y_n, \end{aligned} \tag{2.1}$$

where $0 < c \ll 1$ is a small, positive constant that determines the baseline activities when the input signal vanishes. $f(x)$ is a positive, monotonically increasing, convex function. In the simulations presented in section 3, $f(x) = x^q$ is used, although the precise form of $f(x)$ is not critical, and we obtained similar results using $f(x) = e^{qx}$.

We also examine a divisive feedback network (DFB) variation of the FFN network:

$$\begin{aligned} \tau \dot{y}_n &= -y_n + \frac{x_n f(y_n)}{c + \sum_k f(y_k)} \\ z &= \sum_n y_n, \end{aligned} \tag{2.2}$$

where f and c are as for the FFN model and τ is the time constant of the dynamical system.

The architectures of the divisive models, FFN and DFB, are similar to those previously proposed for WTA behavior (Grossberg, 1973; Koch & Ullman, 1985; Fukai & Tanaka, 1997), gain control in the fly visual system (Reichardt et al., 1983), and attentional modulation on orientation filters in human vision (Lee et al., 1999).

The third model is a linear threshold network (LIN) with subtractive inhibition, represented as rectified linear inhibitory inputs:

$$\begin{aligned} \tau\, \dot{y}_n &= -y_n - \sum_k w[y_k]_+ + x_n \\ z &= (w+1) \sum_n [y_n]_+, \end{aligned} \tag{2.3}$$

where the constant $w > 0$ specifies the inhibitory synaptic strength. The output gain factor $w + 1$ in equation 2.3 ensures that $z = x_m$ exactly at equilibrium (see section 3.1), but in fact any constant gain factor would

work equally well for the purpose of achieving linearity, as discussed in section 1.

The fourth network is a leaky integrate-and-fire model (SPK), which directly implements the threshold nonlinearity of the LIN network:

$$\begin{aligned} \tau\, \dot{m}_n &= -m_n - w \sum_{k \neq n} y_k + x_n \\ y_n(t) &= \sum_i \delta(t - t_i) \\ z(t) &= \sum_n y_n(t). \end{aligned} \tag{2.4}$$

Unit n fires if m_n exceeds threshold θ. The membrane potential, represented by m_n, is lower-bounded by zero and is also reset to zero after each spike at time t_i. One major difference between the SPK model and the LIN model is that while the dynamic variable y in the LIN model can become negative, m in the SPK model is lower-bounded by zero, and therefore the impact of past history on current activities is more limited in the SPK model. Also, the intermediate-layer units communicate only when one or more of them spike, making their activities more input-bound. As we will see in section 3, these differences result in some fundamentally different response properties.

Network architecture similar to the linear threshold networks, LIN and SPK, has previously been proposed in the context of inhibitory interactions in the limulus retina (Hartline & Ratliff, 1957), orientation tuning in the visual cortex (Ben-Yishai, Lev Bar-Or, & Sompolinsky, 1995), gain fields in the parietal cortex (Salinas & Abbott, 1996), and WTA behavior in analog VLSI circuits (Lazzaro et al., 1989; Hahnloser et al., 2000).

3 Results

In the following, we examine the linearity and selectivity properties (as discussed in section 1) of each network. The responses of intermediate-layer and output-layer units in equilibrium conditions are examined. We also explore the models' abilities to approximate the MAX operation under a range of internal and external parameter values: input distribution, strength of lateral inhibition, number of inputs, initial conditions, and the presence of noise[4]. We also point out circumstances in which z is a more robust reconstruction of x_m than y_m, the latter of which has been used in

[4] In all of our simulations, the inputs are assumed to change much more slowly than network activities and thus are represented as constant as the network response relaxes toward equilibrium.

previous MAX models with similar architecture (Yuille & Grzywacz, 1989). The relevant mathematical analyses can be found in the appendix.

3.1 Network Dynamics.

3.1.1 Divisive Feedforward Network. As with all of the other models we present, the feedforward network has no synaptic delay, thus its "dynamics" is trivial in the sense that the output immediately and fully reflects the system's processing of the input. From equation 2.1, notice that if $f(x)$ is sufficiently convex and therefore sufficiently exaggerates the difference between the maximal input, x_m, and the other inputs, then $f(x_m)$ dominates the sum in the denominator of equation 2.1, resulting in $y_m \approx x_m$ and $y_n \approx 0$, $\forall n \neq m$, giving rise to $z \approx x_m$.

3.1.2 Divisive Feedback Network. For the DFB model, the effect of the nonlinear operation f is similar, except the difference between y_m and y_n is further exaggerated in each iteration until equilibrium is reached. The network dynamics

$$\begin{aligned} y_m &= \frac{x_m y_m^q}{c + \sum_k y_k^q} \\ y_n &= \frac{x_n y_n^q}{c + \sum_k y_k^q}, \quad \forall n \neq m, \end{aligned} \tag{3.1}$$

give rise to the equilibrium relation $y_n/y_m = (x_n/x_m)(y_n/y_m)^q$, which is consistent if $y_n = 0$, $\forall n \neq m$. It is easy to see that the system has a stable equilibrium at $y_n = 0$, $z = y_m \approx x_m$.

3.1.3 Linear Threshold Network. Similarly, in the LIN network, given sufficiently strong lateral inhibition, represented by w in equation 2.3, y_n, $\forall n$, approaches

$$y_n = \sum_k w[y_k]_+ + x_n. \tag{3.2}$$

It is easy to see that $y_m = \frac{1}{(w+1)x_m}$ and $[y_n]_+ = 0$ is a stable equilibrium of this system, giving rise to $z = x_m$. In section A.3, we quantitatively describe the conditions on the parameters that allow the system to reach this equilibrium.

3.1.4 Spiking Network. For the SPK network, since it is deterministic, given similar initial conditions for each m_n, m_m always reaches firing threshold before all the other units, resulting in a strong lateral inhibition of its neighbors in the next time step. More quantitatively, if $w \geq x_m - \theta$, then $[m_n - \theta]_+$ tends toward 0 (y_n is silent), and m_m tends toward x_m, inducing a regular spike train in y_m, and therefore in z. The frequency of this spike train

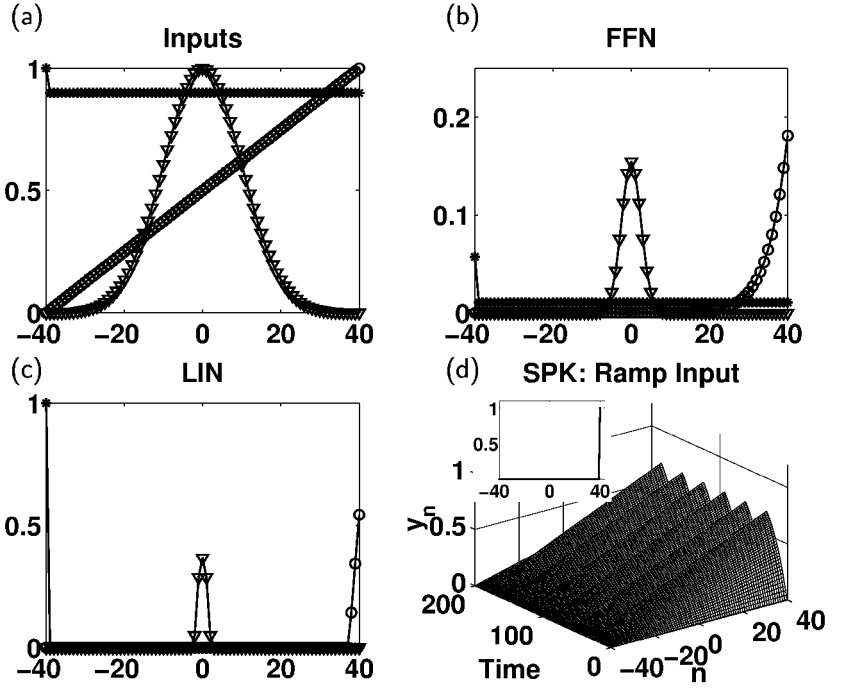

Figure 2: Different input types. (a) Different input types: ∇ gaussian, o ramp, * uniform. $x_m = 1$. Because the network connectivity is symmetric in all models, there is no inherent topology in the output. (b) Corresponding equilibrium intermediate-layer activities, y_n, in the FFN ($q = 6$). z for FFN: Gaussian—0.97, ramp—0.95, uniform—0.91. (c) y_n activities for LIN ($w = 15$). z for LIN: gaussian—1.04, ramp—1.03, uniform—1.00. (d) Evolution of membrane potential m_n in the SPK network in response to "ramp" input: $w = 10$. Inset shows average firing rates (units normalized for comparison).

is roughly proportional to the input magnitude, tempered by the "leakiness" of the integrate-and-fire neuronal model.

3.2 Simulations.

3.2.1 Intermediate-Layer Activities. To illustrate the behavior of $\{y_n\}$ in each of the models, we use several different classes of inputs, with identical amplitude $\alpha = 1$ (see Figure 2a). There are $N = 81$ input and corresponding intermediate-layer units:

- Gaussian: width $\sigma = 10$, centered at $n = 0$
- Ramp: $x_n = \alpha(n/80 + 1/2)$
- Uniform inputs with one "winner": $x_0 = \alpha$, $x_n = 0.90\alpha$, $\forall n \neq 0$

The responses to gaussian and ramp inputs are interesting visual illustrations of how the nonlinear intermediate-layer interactions differentially attenuate the input signals: the gaussian peak is narrower, and the ramp has been transformed into an exponential (see Figures 2b and 2c). The "uniform" inputs represent a worst-case scenario for some of the models (see section A.1) and will be used in later simulations to examine the dependence of z on the magnitude of the submaximal inputs. While LIN completely suppresses the nonmaximal inputs, FFN partially suppresses the nonmaximal inputs. However, $z \approx x_m$ for all types of inputs, for both FFN and LIN, as part of a general pattern where the final output, z, is much more accurate and consistent in reporting x_m than the maximal intermediate-layer activity, y_m. The DFB network responds very well to all the input types, where only $y_m > 0$ in every case (the results are difficult to visualize in the format of Figures 2b and 2c). The SPK network also responds well to all input types, where only m_m ever reaches firing threshold. Figure 2d shows the evolution of membrane potential of the intermediate units in the SPK model, m_n, in response to a ramp input. The inset shows the corresponding average firing rate: only y_m is active.

3.2.2 Dependence on Submaximal Inputs and Inhibitory Strength. Selectivity, or the ability to ignore submaximal inputs $x_n, n \neq m$, is a critical property of the MAX operator, as was discussed in section 1. This property is sensitive to the strength of lateral inhibition, controlled by q in FFN and DFB and by w in LIN and SPK models. We consider the case where all the submaximal inputs, x_n, are identical, since they provide a systematic means of varying the submaximal inputs and a worst-case scenario for some of the models (see the appendix).

Figure 3 shows the dependence of z and y_m on the magnitude of the submaximal inputs x_n, while Figure 4 shows the effect of varying inhibitory strength (q for the divisive models, w for the linear models). For the FFN model, we see that $z = x_m$ for small values of x_n and for $x_n = x_m$, but dips slightly for intermediate values, a phenomenon that concurs with our mathematical analysis (see section A.1). In contrast to z, y_m has a much stronger dependence on x_n, dropping to x_m/N in the limit of $x_n = x_m$. Also, the strength of inhibition, represented in the model by q, has a very small effect on z compared to y_m, which starts deviating from x_m for even small x_n when w is small. A more systematic study of this effect is shown in Figure 4. These results agree with the general observation that z is much more robust than y_m in reproducing x_m in a variety of situations in this and most of the other models studied.

Divisive inhibition applied recurrently, as demonstrated by the DFB plot in Figure 3, significantly improves the system's performance in implementing the ideal MAX operation. z reproduces x_m in equilibrium, independent of x_n and q (see also Figure 4), agreeing with our mathematical analysis (see section A.2).

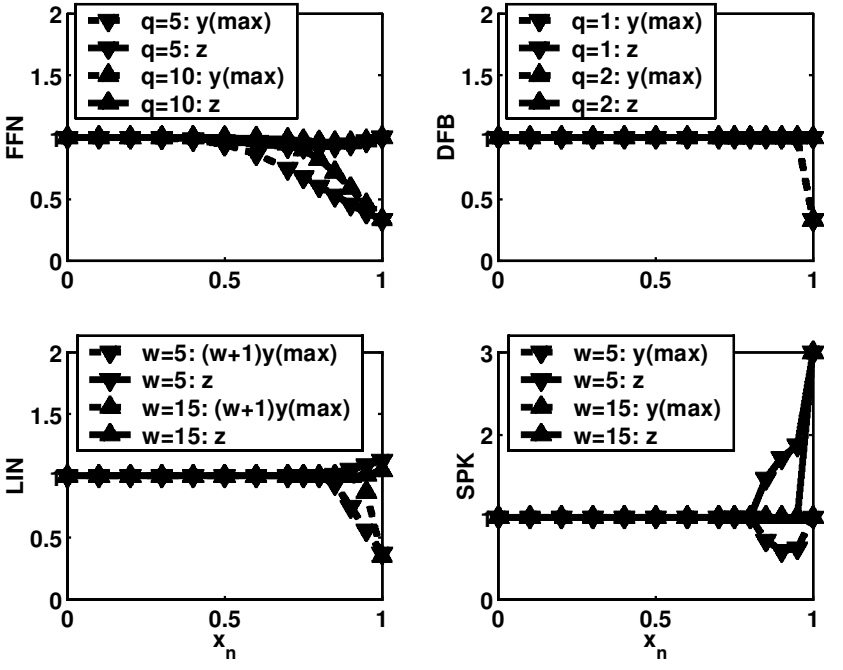

Figure 3: Dependence on nonmaximal input. For each network, the magnitude of $x_n, \forall n \neq m$, was varied. $N = 3, x_m = 1$. The firing rate of the spiking network has been normalized for ease of comparison. z is more robust than y_m in general.

The final output, z, of the LIN model has similar response to variations in x_n as that of FFN, although y_m in this case behaves more robustly with respect to the magnitude of the submaximal inputs x_n (see Figure 3), over a range of lateral inhibition w (see Figure 4).

The SPK model behaves well except when x_n is large (see Figure 3) or w is small (see Figure 4), in which case a spike in a unit m causes its own membrane potential to drop to 0 in the succeeding time step, while the other units receiving inhibition from the spike may be only partially suppressed and be the first to reach firing threshold next. In this case, the firing can alternate between y_m and y_n, and the overall firing rate, z, increases accordingly.

3.2.3 Dependence on Number of Inputs. In some of the previous models proposed for the realization of the MAX operation or WTA, the performance of the models was dependent on the number of inputs (Yuille & Grzywacz, 1989). As the number of inputs increased, the system tended to become increasingly inaccurate in reproducing the maximal input. Our models are relatively immune to this problem (see the appendix). Figure 5 shows the

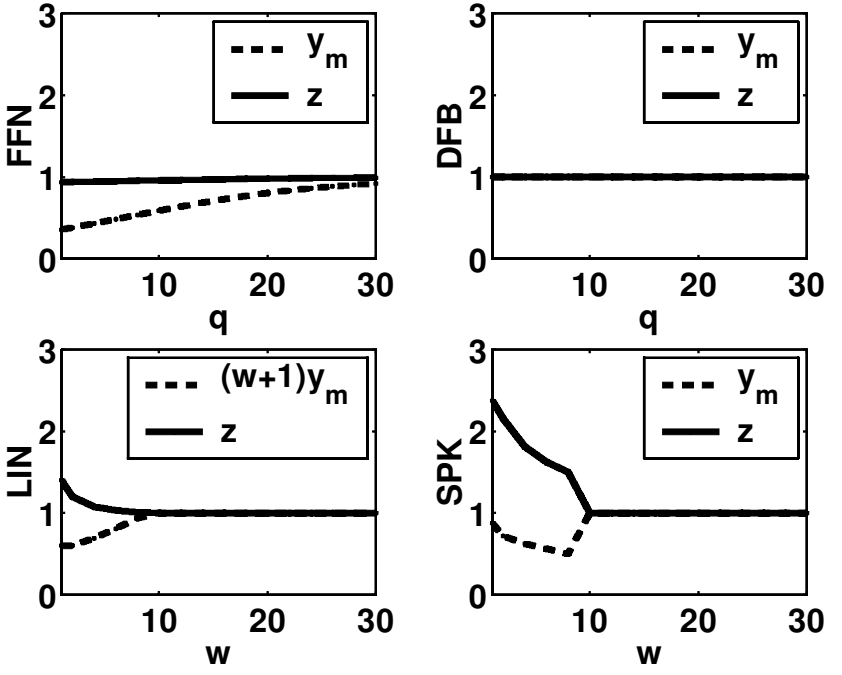

Figure 4: Strength of lateral inhibition. Relationship between output and input amplitude as a function of inhibitory strength, represented by the parameter q or w. Each network is simulated to convergence, as q or w is varied from 2 to 30. In general, z is more robust than y_m. For the DFB model, y_m and z overlap completely. $N = 3$, $x_m = 1$, $x_n = 0.9$.

network responses to number of inputs varying from 2 to 30 (higher values of N result in responses similar to those in the case of $N = 30$). The robustness of z is striking, even though in the FFN model, y_m converges to x_m/N, and in LIN, y_m converges to $w(x_m - x_n)$ (see sections A.1 and A.3 for details).

3.2.4 Dependence on Initial Conditions. The FFN model has no memory of its past states and therefore no dependence on initial conditions. The DFB model has, besides the attractor that gives rise to MAX-operator behavior, other non-MAX attractors. An example is shown in Figure 6b, where the initial strength of one unit allows it subsequently to suppress the other unit and dominate the sum z, even though its input ceases to be the maximum. In section A.2, we examine in more details how this undesirable attractor can arise. The LIN network, in contrast, has a unique equilibrium for z that does not depend on the identity of x_m or the magnitude of x_n, as shown in

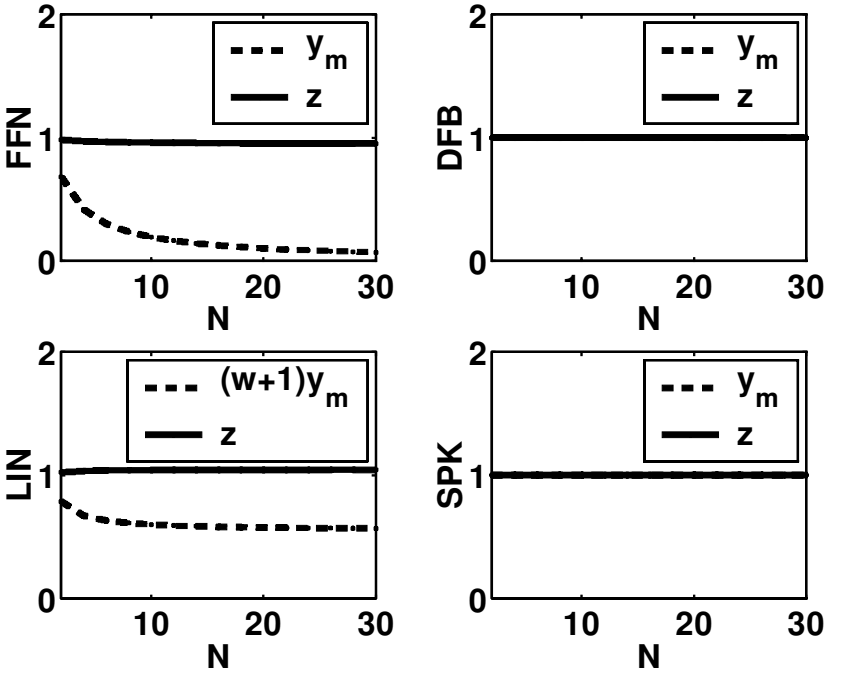

Figure 5: Number of inputs. Dependence of network performance on the number of inputs, N, ranging from 2 to 30. $x_m = 1$, $x_n = 0.9$, $N = 3$. FFN: $q = 15$, DFB: $q = 2$, LIN: $w = 10$, SPK: $w = 20$.

Figure 6. The SPK model has behaviors very similar to LIN: the intermediate-layer activities reflect the relative strength of inputs, but the final output z depends on only the maximal input.

3.2.5 Noise Sensitivity. The reader may well wonder by this point how our deterministic models fare in the presence of noisy inputs. Does the nonlinear amplification of the signal also have an amplifying effect on the noise component of the input? To examine this issue, each model is fed with uniform inputs (with one "winner"), to which gaussian noise, independently generated in each iteration, is added. The results are shown in Figure 7.

While the FFN responds with output noise of magnitude comparable to input noise, the feedback networks actually suppress output noise, in accordance with previous work on recurrent networks (Yuille & Grzywacz, 1989; Salinas & Abbott, 1996): the recurrent interactions have an averaging effect on the noise. Note that the noise behavior of the spiking model is not directly comparable to the other models, as the input noise of the spiking

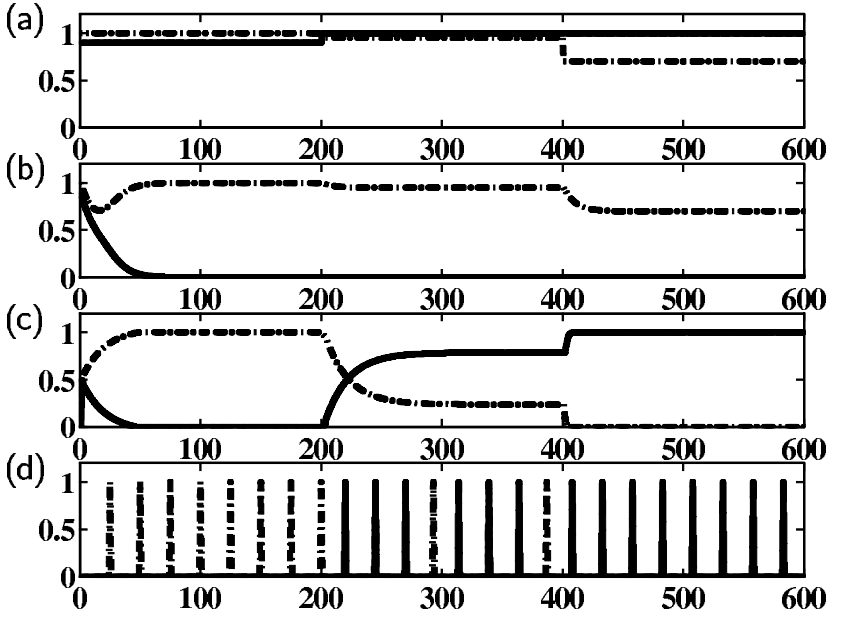

Figure 6: Multistability. Dependence on initial conditions. (a) $N = 2$. Dashed: unit 1. Solid: unit 2. Iterations 1–200: $x_1 = 1, x_2 = 0.9$. Iterations 201–400: $x_1 = 0.95, x_2 = 1$. Iterations 401–600: $x_1 = 0.7, x_2 = 1$. (b) y_1 (dashed) and y_2 (solid) responses of DFB. z depends on initial conditions. (c) y_1 (dashed) and y_2 (solid) responses of LIN. $z = y_1 + y_2$ does not depend on initial conditions. (d) SPK responses similar to LIN's: y_1 (dashed) and y_2 (solid) activities reflect relative strength of inputs, but z is consistent and only a function of the maximal input.

model is measured in terms of membrane potential, while the output noise is measured in terms of firing rate; for all the other models, both the input and output noise are measured in terms of mean firing rate. Overall, it is reassuring that despite the presence of signal amplification in the systems, output noise increases linearly as a function of input noise with acceptably small slopes.

4 Predictions

The results from the simulation studies presented in section 3 give rise to a number of model-specific predictions, which are summarized in Table 1.

Some of the predictions are easier to test than others. Preliminary evidence indicates there are single cells in striate cortex (Sakai & Tanaka, 1997; Lampl et al., 2001) and inferior temporal cortex (Sato, 1989) that respond to visual stimuli in a MAX-like manner. Given such a "MAX" neuron, it is

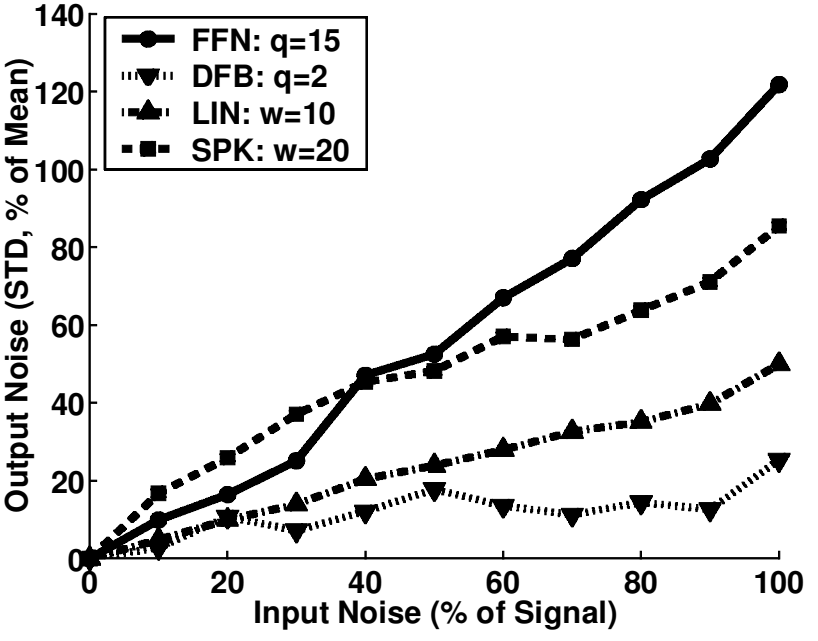

Figure 7: Noise sensitivity. Standard deviation of output for each network over 50 trials, normalized by the expected or noiseless output, as a function of the amount of uncorrelated noise (measured in terms of standard deviation from the *noiseless* input signal) added independently to each unit in each iteration. $N = 3$, $x_m = 1$, $x_n = 0.9$. FFN: $q = 15$, DFB: $q = 2$, LIN: $w = 10$, SPK: $w = 20$.

relatively easy to test whether the neuron exhibits hysteretic behavior. For instance, given the time courses of a MAX neuron's responses to two stimuli separated in space, its combined response should be point-wise maximum of the two responses if it is nonhysteretic, but be dominated by the one that has the shorter-latency and dominant transients if it is hysteretic. In fact,

Table 1: Model-Specific Predictions.

Prediction Model	FFN	DFB	LIN	SPK
Hysteresis	no	no	no	yes
Sensitive to $GABA_A$ blocker	yes	yes	no	no
Sensitive to $GABA_B$ blocker	no	no	yes	yes
Effect of *GABA* blocker on z	none	none	$\uparrow$	$\uparrow$
Sparse y_n activity	no	yes	no	no
Effect of larger N on y_m	$\downarrow$	none	$\downarrow$	none
Noise suppression	no	yes	yes	yes

preliminary data from Lampl et al. (2001) give some support to the latter behavior. In this work, the only model we examined that exhibits nontrivial hysteresis is the DFB model. However, it has been shown previously that if a little self-excitation is imposed on the LIN model, then it would also exhibit hysteresis (Hahnloser, 1998).

In principle, it should also be possible to apply antagonists locally to different inhibitory neurotransmitters, such as $GABA_A$ or $GABA_B$, in order to weaken the proposed inhibitory connections involved in implementing the MAX operation. For example, $GABA_A$ or other inhibitory receptors with reversal potential near resting potential would be critical for implementing the divisive shunting inhibition used in the FFN and DFB models; $GABA_B$ or other inhibitory receptors with very negative reversal potential could be involved in subtractive inhibition involved in the LIN or SPK models. Local application of these blockers therefore might be able to differentiate whether the MAX operation involves divisive or subtractive inhibition, which are most likely to be involved in cellular or network implementations of the MAX operation, respectively. Moreover, the specific effects of suppressing the inhibitory synapses could be different according to our simulation results: when the inhibitory activities are suppressed, LIN and SPK's output should be increased, while FFN and DFB's output should not be significantly affected. Successful implementation of these experiments would provide valuable insight into the underlying mechanisms of the MAX operation, although in practice, interpretation of these experiments might be difficult, as the relevant synapses might live and interact on distal dendrites while recordings will mainly be in the soma.

If experimental data could give some indication as to what x_n and y_n represent, whether they are computed synaptically within the recorded cell or represent input and output to other neurons, more specific predictions might then be tested: sparseness of y_n activation, effect of N on y_m, and even effect of input noise on z.

5 Biophysiological Considerations

Even in the absence of more detailed experimental data, some reasonable conjectures can be made about the relative plausibility of the models based on biophysiological considerations.

The feedforward divisive inhibition model can be implemented in multiple ways. Two particularly attractive implementations emerge if equation 2.1 is broken down in two distinct ways:

$$y_n = x_n * \frac{f(x_n)}{c + \sum_k f(x_k)} \tag{5.1}$$

$$y_n = x_n f(x_n) * \frac{1}{c + \sum_k f(x_k)}. \tag{5.2}$$

Equation 5.1 can be thought of as a multiplication between x_n and the output of a normalizing network of mutually inhibitory neurons that also receive a copy of the input x_n. Such network interactions have been been previously proposed to account for response of V1 neurons receiving afferent LGN inputs and lateral cortical inputs (Carandini & Heeger, 1994). Equation 5.2 represents another possibility: the gating of a nonlinear operation on the input, $g(x_n)$ by an inhibitory pooling neuron computing $\sum_k f(x_k))$, where $g(x_n) \propto x_n f(x_n)$ is necessary for achieving the linearity property of the ideal MAX operation. Such multiplicative gain control interactions have been observed in the parietal cortex (Salinas & Abbott, 1996), insect visual system (Hatsopoulos, Gabbiani, & Laurent, 1995), area LIP (Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990), and the superior colliculus (Van Opstal, Hepp, Suzuki, & Henn, 1995).

The biophysiological considerations of the DFB model are similar to those of the FFN model, involving two main possibilities of decomposing the product $\frac{x_n f(y_n)}{c+\sum_k f(y_k)}$, as described before. The main differences are that for the normalization implementation, the mutually inhibitory units are represented by y_n rather than x_n; similarly, for the gain control implementation, the gain factor now depends on y_n rather than x_n, and the gain-controlled neuron receives inputs from both x_n and y_n and computes $g(x_n, y_n) \propto x_n f(y_n)$. One significant advantage of this recurrent model over the feedforward model is that it works well with little or no nonlinear amplification of the signal by the operator f, as we have seen in the simulation results in section 3 and the mathematical analysis in section A.2. Another computational advantage of this recurrent model is that it performs very efficient noise suppression.

The implementation of the LIN and SPK models is straightforward. Each y_n probably represents the activity of a single neuron, which receives inhibition from its neighbors and excitation from the input. One ambiguity is that the summation can be done by either a dedicated summation neuron (in which case, self-excitation is needed for the SPK model), or the unit could be directly inhibited by its neighbors via dendritic inputs (in which case, self-inhibition is needed for the LIN model).

6 Discussion

In this work we have reviewed and analyzed a number of neural circuits that provide good approximations to the MAX operation, which has been proposed to play a significant role in various processes of the visual system. The MAX operation is interesting because it is an example of a fundamental, nonlinear computational operation that can be realized with neurophysiologically plausible mechanisms. The models were chosen in order to demonstrate different neural principles for the realization of this computational operation. They are simple enough to allow an understanding of

the underlying parametric dependencies and some mathematical analysis. The neural mechanisms on which our models are based are also fundamental in models for contrast gain control and winner-take-all, both of which have been extensively studied in the context of important cortical processes. These processes and the MAX operation may well share similar or overlapping neural substrate.

From our mathematical and simulation analyses, it appears that each model is endowed with a variety of distinct computational properties, although they may be difficult to test in experimental settings. The biophysiological considerations for the various models are also complex and nontrivial. In general, in the absence of more detailed experimental data, the differences in the model behaviors do not lead to immediate support for one model over any other.

This work gives rise to a number of potential directions of future theoretical research. One obvious extension of this work is to analyze neurophysiologically more detailed and more realistic models, which could involve more stochastic descriptions of network dynamics, or biophysiologically more realistic neurons. For instance, software such as Neuron or Genesis could be used to simulate dendritic interactions and shunting conductances explicitly. Another interesting question is how any of these models may be learned through experience or wired up during development. Mechanisms similar to those proposed by Fukushima (1980) to explain learning in the Neocognitron may be explored. The design of these future studies is contingent on crucial experimental data that are not yet available. The simulation and mathematical results from our work give rise to a number of predictions, which can be used to guide future experimental investigations on the MAX operation. Although the experimental demonstration of the different properties predicted by the models is nontrivial, this preliminary theoretical analysis should be a helpful first step for the preparation of more detailed neurophysiological experiments.

Appendix

The mathematical discussions here are intended to help explain and support the various simulation results presented in section 3. We are mainly interested in how well the networks satisfy the selectivity and linearity properties of the MAX operation and how they are affected by internal and external parameters, such as the magnitude of the nonmaximal inputs, the number of inputs, the strength of inhibition, and initial conditions. We also point out why z is a more robust reconstruction of x_m than y_m, the latter of which has been used in previous MAX models with similar architectures (Yuille & Grzywacz, 1989).

A.1 Divisive Feedforward Network. Given equations 2.1, and using a polynomial form of f, $f(x) = x^q$:

$$z = \sum_y y_n \approx \frac{\sum_{n=1}^N x_n x_n^q}{\sum_{k=1}^N x_k^q}$$

$$= x_m \frac{1 + \sum_{n \neq m} r_n^{q+1}}{1 + \sum_{k \neq m} r_k^q}$$

$$= x_m L,$$

where we have defined the ratio $r_n \equiv \frac{x_n}{x_m}$ and $L \equiv \frac{1+\sum_{n\neq m} r_n^{q+1}}{1+\sum_{k\neq m} r_k^q}$. Note $0 \leq L \leq 1$, since $0 \leq r_n \leq 1$. Simple calculus manipulation shows that L reaches minimum (and z deviates maximally from x_m) when the ratios are

$$r_j = \frac{q(1 + \sum_{k\neq m} r_k^{q+1})}{(q+1)(1 + \sum_{k\neq m} r_k^q)},$$

a quantity independent of j. In other words, the selectivity of the model, that is, the ability of z to ignore the submaximal inputs, is most challenged when x_n are identical. Identical nonmaximal inputs therefore provide a worst-case scenario for the FFN model, where z can be expressed as

$$z = x_m \frac{1 + (N-1)r^{q+1}}{1 + (N-1)r^q}$$

and y_m as

$$y_m = \frac{x_m}{1 + (N-1)r^q}.$$

Note that the dependence of z on N is minimal compared to that of y_m, since z has an extra $O(N)$ term in the denominator to balance out the numerator (see Figure 5a). Notice also that the dependence of z on r is not monotonic: $z \approx 1$ when $r \approx 0$ or $r \approx 1$, but is somewhat less for intermediate values of r, consistent with the simulation results in Figure 3a.

A.2 Divisive Feedback Network. As we discussed in section 3.1, there is an attractor at $z = x_m$, a phenomenon that is independent of q, N, or $\{r_n\}$. This is reflected in the robustness of the model's output in Figures 3b, 4b, and 5b. However, there also exist other attractors that do not approximate the MAX operation, such as the ones shown in Figure 6b. A closer look at the system of equations 3.1 makes the reason apparent: this set of equations is stable if any $y_n \approx x_n$, and all the other $y_k = 0$, $k \neq n$. In particular, it means

that if y_n has been the active unit, then it may continue to dominate even if its input becomes smaller than those of its neighbors. Finally, let us note that under the special condition where all inputs are equal, $x_m = x_n$, then there is a unique attractor where $z = x_n$ and each $y_n = z/N$, as long as the initial conditions are not such that $y_n = 0, \forall n$ (the system also has a degenerate, unstable equilibrium at $y_n = 0, \forall n$).

A.3 Linear Threshold Network. The attractor giving rise to the ideal MAX operation, as described in section 3.1, requires $[y_n]_+ = 0, \forall n \neq m$. Now we examine conditions under which this cannot be fulfilled. First, note that given equation 2.3, subtracting the expression for y_n from y_m, gives the convergence

$$y_m - y_n = \beta_n \equiv x_m - x_n.$$

Thus, if $\beta_n < x_m/(w+1)$, then $y_m - y_n < x_m/(w+1)$, and it cannot be such that $y_m = x_m/(w+1)$ and $y_n \leq 0$. In this case, Equation 3.2 still holds, and we have the following:

$$\begin{aligned}\sum_n [y_n]_+ &= \sum_{j:\, y_j>0} y_j \\ &= \sum_{j:\, y_j>0} \left(x_j - \sum_{k:\, y_k>0} w y_k \right) \\ &= x_m + \sum_{j:\, y_j>0, j\neq m} x_j - Jw \sum_{k:\, y_k>0} y_k.\end{aligned}$$

Rearranging the terms,

$$(1+Jw) \sum_{j:\, y_j>0} y_j = x_m + \sum_{j\neq m} x_j = x_m + (J-1)x_m - \sum_{j\neq m} \beta_j,$$

where J is the total number of units such that $y_j > 0$. Then we can express z as

$$\begin{aligned}z &= (w+1) \sum_{j:\, y_j>0} y_j \\ &= \frac{Jw+J}{Jw+1} x_m - \frac{w+1}{Jw+1} \sum_{j\neq m} \beta_j \\ &= x_m + \frac{w+1}{Jw+1} \sum_{j\neq m, y_j>0} \left(\frac{x_m}{w+1} - \beta_j \right),\end{aligned}$$

We know $\beta_n < x_m/(w+1)$, so z is maximized if β_j is minimized toward 0 and if J is maximized toward N (see Figure 3c). It is clear that using uniform inputs with decreasing β_n is a valid way of examining the network response as the inputs approach the worst-case scenario: $z = x_m(1+(N-1)/(Nw+1))$, where the error factor is inversely proportional to w (see Figure 4c) and relatively independent of N (see Figure 5c). For identical and nonzero $\beta_n = \beta$, and large w,

$$z = x_m + \frac{(w+1)(N-1)}{Nw+1}\left(\frac{x_m}{w+1} - \beta\right) \approx x_m,$$

where the error term is roughly independent of N. We also have the equilibrium condition,

$$(w+1)y_m \approx \frac{x_m}{N} + \frac{(N-1)}{N}w\beta,$$

indicating that y_m has an inverse dependence on N initially, but then becomes relatively independent of N as N becomes large.

A.4 Spiking Network. Because of the deterministic dynamics of equation 2.4, given identical initial conditions, y_m always fires before the other units can reach the threshold, and therefore y_m is the only unit that is active. In the degenerate case that y_n is identical for all inputs, all the units fire synchronously, and the overall firing rate, z, is N times higher (see Figure 3d). These behaviors are independent of magnitude of w (see Figure 4d), as long as it is large enough to reset all the membrane potential, m_n, of all the nonfiring cells to 0 after each spike. If this is not the case, then the unit receiving the second largest input x_{m_2} may have greater membrane potential than m_m after unit m fires: $m_{m_2} > m_m$, and may subsequently be the first to reach the firing threshold, as is the case in Figure 6d. In this way, the firing may alternate between two or more units, and the overall firing rate z increases as w decreases.

7 Acknowledgments

We thank Peter Dayan, Richard Hahnloser, and Maximilian Riesenhuber for helpful discussions. This work was supported by the Office of Naval Research under contract N00014-93-1-3085, Office of Naval Research under contract N00014-95-1-0600, National Science Foundation under contract IIS-9800032, and the National Science Foundation under contract DMS-9872936. A.Y. was also supported by Massachusetts Institute of Technology under the UROP program, National Science Foundation under Graduate Research Fellowship Program, and the Gatsby Charitable Foundation under a grant

to the Gatsby Computational Neuroscience Unit. M.G. was also supported by the Deutsche Forschungsgemeinschaft, Honda R&D Americas, and the Deutsche Volkswagenstiftung.

References

Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., & Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *Journal of Neuroscience, 10*(4), 1176–1196.

Anderson, C. H., & Van Essen, D. C. (1987). Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences (USA), 84*(17), 6297–6301.

Anderson, J. S., Carandini, M., & Ferster, D. (2000). Orientation Tuning of Input Conductance, Excitation, and Inhibition in Cat Primary Visual Cortex. *Journal of Neurophysiology, 84*, 909–926.

Ben-Yishai, R., Lev Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences (USA), 92*, 3844–3848.

Borg-Graham, L., Monier, C., & Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature, 393*, 367–373.

Carandini, M., & Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *Journal of Neurophysiology, 20*(1), 470–484.

Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science, 264*, 1333–1336.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17*(21), 8621–8644.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience, 17*(9), 3201–3214.

Doiron, B., Longtin, A., Berman, N., & Maler, L. (2001). Subtractive and divisive inhibition: Effect of voltage-dependent inhibitory conductances and noise. *Neural Computation, 13*(1), 227–248.

Ferster, D., & Jagadesh, B. (1992). EPSP-IPSP interactions in cat visual cortex studied with in vivo whole-cell patch recording. *Journal of Neuroscience, 12*(4),1262–1274.

Fukai, T., & Tanaka, S. (1997). A simple neural network exhibiting selective activation of neural ensembles: From winner-takes-all to winner-shares-all. *Neural Computation, 9*, 77–97.

Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36*(4), 193–202.

Giese, M. A. (2000). Neural model for the recognition of biological motion. In G. Baratoff & H. Neumann (Eds.), *Dynamische Perzeption* (pp. 105–110). Berlin: Infix Verlag.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron, 9*(1), 1–13.

Grossberg, S. (1973). Contour enhancement: Short term memory, and constancies in reverbarating neural networks. *Studies in Applied Mathematics, 52*, 213–257.

Grzywacz, N. M., & Yuille, A. L. (1990). A Model for the estimate of local image velocity by cells in the visual cortex. *Proc. R. Soc. Lond. B, 239*, 129–161.

Hahnloser, R. (1998). On the piecewise analysis of networks of linear threshold neurons. *Neural Networks, 11*, 691–697.

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

Hartline, H. K., & Ratliff, F. (1957). Spatial summation of inhibitory influences in the eye of limulus, and the mutual interaction of receptor units. *Journal of General Physiology, 41*(5), 1049–1066.

Hatsopoulos, N., Gabbiani, F., & Laurent, G. (1995). Hysteresis reduction in proprioception using presynaptic shunting inhibition. *Journal of Neurophysiology, 73*(3), 1031–1042.

Holt, G. R., & Koch, C. (1997). Shunting inhibition does not have a divisive effect on firing rates. *Neural Computation, 9*, 1001–1013.

Hubel, D. H., & Wiesel, T. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Koch, C., & Poggio, T. (1987). Biophysics of computational systems: Neurons, synapses, and membranes. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Synaptic functions* (pp. 637–697). New York: Wiley.

Koch, C., Poggio, T., & Torre, V. (1983). Nonlinear interactions in the dendritic tree: Localization timing and the role in information processing. *Proceedings of the National Academy of Sciences (USA), 80*, 2799–2802.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Kohonen, T. (1995). *Self Organizing Maps*. Berlin: Springer-Verlag.

Lampl, I., Riesenhuber, M., Poggio, T., & Ferster, D. (2001). The MAX operation in cells in the cat visual cortex. *Society of Neuroscience Abstracts, 619*, 30.

Lazzaro, J., Ryckenbusch, S., Mahowald, M. A., & Mead, C. A. (1989). Winner-take-all network of $O(n)$ complexity. In D. Touretzky (Ed.), *Advances in neural information processing systems, 1* (pp. 703–711). San Mateo, CA: Morgan Kaufmann.

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-takes-all competition among visual filters. *Nature Neuroscience, 2*, 375–381.

Li, Z. (2000). Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. *Neural Computation, 15*(8), 1749–1780.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5*, 552–563.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience, 14*(4), 2178–2189.

Naka, K. I., & Rushton, W. A. H. (1966). S-potential from luminosity units in the retina of fish (cyprinidae). *Journal of Physiology, 185*, 587–599.

Nowlan, S. J. & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *Journal of Neuroscience, 15*(2), 1195–1214.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology, 82*(5), 2490–2502.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86*(1), 159–173.

Reichardt, W., Poggio, T., & Hausen, K. (1983). Figure-ground discrimination by relative movement in the visual system of the fly. Part II. Towards the neural circuitry. *Biological Cybernetics, 46*, 1–30.

Riesenhuber, M. K., & Poggio, T. (1999a). Are cortical models really bound by the "binding problem"? *Neuron, 24*, 87–99.

Riesenhuber, M. K., & Poggio, T. (1999b). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Riesenhuber, M. K., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3* (Supp.), 1199–1204.

Rockland, K. S., & Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology, 216*, 303–318.

Sakai, K., & Tanaka, S. (1997). A spatial pooling mechanism underlies the second-order structure of V1 complex cell receptive fields: A computational analysis. *Soc. Neurosci. Abs., 23*, 453.

Salinas, E., & Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences (USA), 93*,11956–11961.

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferiortemporal neurons in awake monkeys. *Experimental Brain Research, 77*, 23–30.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neural responses in visual area MT. *Vision Research, 38*, 743–761.

Starzyk, J. A., & Fang, X. (1993). CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback. *Electronics Letters, 29*(10), 908–910.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109–139.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Torre, V., & Poggio, T. (1978). A synaptic mechanism possibly underlying directional selectivity motion. *Proceedings of the Royal Society London B, 202*, 409–416.

Treue, S., & Maunsell, J. H. L. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature, 382*, 539–541.

Van Opstal, A. J., Hepp, K., Suzuki, Y., & Henn, V. (1995). Influence of eye position on activity in monkey superior colliculus. *Journal of Neurophysiology, 74*(4), 1593–1610.

White, E. L. (1989). *Cortical circuits*. Boston: Birkhauser.

Wilson, H. R., & Humanski, R. (1993). Spatial frequency adaptation and contrast gain control. *Vision Research, 33*, 1133–1149.

Yu, A. J., Giese, M. A., & Poggio, T. (2000a). Neural circuits for the realization of the maximum operation. *Society of Neuroscience Abstracts, 26*, 1202.

Yu, A. J., Giese, M. A., & Poggio, T. (2000b). Neural mechanisms for the realization of maximum operations. *European Journal of Neuroscience, 12*, 74.
Yuille, A. L., & Grzywacz, N. M. (1989). A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation, 1*, 334–347.

Received September 18, 2001; accepted May 10, 2002.