Pergamon

Neural Networks, Vol. 9, No. 7, pp. 1141–1154, 1996

Printed in Great Britain
0893–6080/96 $15.00+.00

PII: S0893-6080(96)00019-6

CONTRIBUTED ARTICLE

Dynamics of a Winner-Take-All Neural Network

YUGUANG FANG,[1] MICHAEL A. COHEN[2] AND THOMAS G. KINCAID[1]

[1] Department of Electrical, Computer and Systems Engineering, [2] Department of Cognitive and Neural Systems, Boston University

(*Received* 28 *March* 1995; *revised and accepted* 17 *January* 1996)

Abstract—*This paper describes a neural network with lateral inhibition, which exhibits dynamic winner-take-all (WTA) behavior. The equations of this network model a current input MOSFET WTA circuit, which motivates the discussion. A very general sufficient condition for the network to have a WTA equilibrium point is obtained and sufficient conditions for the network to converge to the WTA point are presented. This gives explicit expressions for the resolution and lower bound of the input currents. We also show that whenever the network gets into the WTA region, it will stay in that region and settle down exponentially fast to the WTA point. This provides a speed up procedure for the decision making: as soon as it gets into the region, the winner can be picked up. Finally, we show that this WTA neural network has a self-resetting property. Copyright © 1996 Elsevier Science Ltd*

Keywords—Winner-take-all, Shunting inhibition, Lateral inhibition, Competition, VLSI neural networks, MOSFET, nonlinear dynamics.

1. INTRODUCTION

Neural networks which pick the maximum from a collection of inputs are known as winner-take-all (WTA) networks (Feldman & Ballard, 1982). The operation of these networks is a mode of extreme contrast enhancement where only the maximally stimulated neuron responds and all other neurons in the network are inhibited. Such networks have been used extensively in decision making, pattern recognition, and competitive-learning networks, and general-purpose self-organizing neural networks (Nabet & Pinter, 1991; Sheu et al., 1992; Choi & Sheu, 1993; Kohonen, 1993; Haykin, 1994; Kaski & Kohonen, 1994 and references therein). The WTA networks have also been used in various signal processing applications, including image feature extraction (Mahowald & Delbruck, 1989), nonlinear inhibition (Lazzaro et al., 1989), Hamming network (Robinson et al., 1992), subthreshold-region signal processing (Andreaou et al., 1991), silicon binaural hearing (Mead et al., 1991), image processing applications (Lee & Sheu, 1991), and video compression (Fang et al., 1992).

The current literature frequently describes WTA networks that are constructed using lateral inhibition among the neurons so that the system is a competitive neural network. Intuitively, if the competitive system is initiated from a fair start with sufficiently strong inhibition, the players in this competition will go to two extremes: win or lose, so that WTA behavior can be expected (Grossberg, 1973). However, if the lateral inhibition is weak or lateral excitation is involved in the competition, the dynamics of the system can be very complex as illustrated by Ermentrout (1992) and Lemmon and Kumar (1989). Thus, conditions for WTA behavior are desirable for the practical design of such neural networks. There are many strategies for WTA network designs. MAXNET (Lippmann, 1987) is an architecture of mutual inhibition to select a maximum, though its efficiency and implementation are a problem. Motivated by self-organizing algorithms, some iterative neural networks are designed to pick the largest number in a data set (Yen et al., 1994, 1995 and references therein). Although these networks are convenient for computer computation, they may be very hard to implement in analog hardware. A number of investigations on analog implementation have been undertaken and electronic networks have been designed to implement the WTA function. Lazzaro et al. (1989) developed a current–input voltage–output WTA circuit using MOSFET transistors. This circuit has the advantage that it has $O(N)$ interconnections. However, the paper only analyzed the circuit at steady-state and did not provide the dynamic analysis, which is obviously a

Requests for reprints should be sent to Professor Thomas G. Kincaid, ECS Department, College of Engineering, Boston University, 44 Cummington Street, Boston, MA 02215, USA.

very important issue for WTA network design. Majani et al. (1989) generalized the WTA analog networks to K-winners-take-all networks. Starzyk and Fang (1993) modified the Lazzaro circuit to improve the resolution and speed performance by introducing excitatory feedback. Sheu and his coworkers (Lee & Sheu, 1991; Fang et al., 1992; Sheu et al., 1992; Choi & Sheu, 1993) systematically investigated the WTA networks, their VLSI implementations and various applications. A voltage-input circuit was developed by Choi and Sheu (1993) using a cascade configuration to significantly increase the competition resolution and the high-speed performance. More recent circuits described by Serrano and Linares-Barranco (1995) and by Smedley et al. (1995) offer solutions to the transistor matching problem.

However, there has been no complete theoretical dynamic analysis of a circuit to show under what conditions on the parameters and the inputs the circuit does have a WTA equilibrium point and converge to this point. In the examples cited above, simulations or experiments have been used to demonstrate that there is an input resolution and a minimum input required to guarantee WTA operation. Some cited MOSFET implementations built the WTA circuits with transistors biased in the subthreshold regions in order to achieve the low-power operation. However, as Choi and Sheu (1993) noted, these circuits may have some significant limitations such as low operation speed, small dynamic range, and limited noise immunity for engineering applications.

This paper gives the detailed dynamic analysis for a WTA analog circuit design which is suitable for MOSFET implementation. The circuit is a fully connected $O(N^2)$ implementation. Although this may be only suitable for small scale applications, the dynamic analysis present in this paper provides for the first time explicit expressions for resolution and minimum inputs in terms of the circuit parameters, which serves as the design guide for MOSFET fabrication and network implementation. Moreover, we do not assume that all transistors in the circuit operate in the specific region during the dynamic transient. The conditions under which the WTA network exhibits the desired behavior is divided into two parts. First, necessary and sufficient conditions on the network parameters are derived for the existence of a WTA equilibrium point. Second, sufficient conditions are derived under which the network goes to the WTA point from a fair start at the origin. It is also shown that a fair start can be obtained by zeroing the inputs. To the authors' knowledge, this is the first systematic dynamic analysis for the WTA MOSFET circuits.

The organization of the paper is as follows. In Section 2, we present the MOSFET implementation of the WTA network, serving as the motivation for the subsequent development. This description of this network is formalized in a set of differential equations which use a MOSFET device model. A new class of dynamical neural networks is proposed which covers a few known neural networks such as shunting and additive neural networks. In Section 3, the first of our two part derivation is presented. A set of necessary and sufficient conditions for the existence of a WTA point is derived for the general networks. In Section 4, the second part of the derivation is given. Using a novel analysis, sufficient conditions are obtained for the MOSFET network to go to the WTA point from a fair start. The resulting conditions are either resolution and upper bound conditions on the inputs or resolution and gain conditions. We also show that if the neural network gets into the WTA region, then it will stay there and settle down at the WTA point. This suggests a quick decision-making procedure: it is not necessary to wait for the network to reach the equilibrium point; as long as it gets into the WTA region, a decision can be made immediately. At the end of this section we show that the system can be reset automatically to the fair starting point by switching off the inputs for a while. Simulation results are presented in the fifth section to illustrate the effectiveness of the designed WTA neural network. Finally, we conclude with some comments about this research and future research directions.

2. MOFSET IMPLEMENTATION OF WTA NETWORK

The winner-take-all neural networks we present in this paper use the metal oxide semiconductor field effect transistor (MOSFET), and the configuration for this network is shown in Figure 1.

In such an N neuron network, each neuron consists of $N-1$ MOSFETs, one resistor and a capacitor, which can be the stray capacitance. Therefore, we need $N(N-1)$ MOSFETs and N resistors in this network. All MOSFETs, resistors, and capacitors are identical; therefore, all neurons in the network are identical. In this circuit, we use the n-channel enhancement mode MOSFET with threshold voltage V_t and physical parameter K. Let $i \in \{1, 2, \ldots, N\}$. The inputs to the neurons are currents, which can be generated by photodiodes. Let I_i denote the ith input current, and let $v_i(t)$ denote the voltage on the capacitor of the ith neuron. The ith neuron is shown in Figure 2.

Consider the ith neuron. For $j \neq i$, let $I_{\mathrm{ds}}^{(ij)}$, $V_{\mathrm{gs}}^{(ij)}$, $V_{\mathrm{ds}}^{(ij)}$ denote the current through the drain, the voltage between the gate and source, and the voltage between the drain and the source, respectively, for the jth MOSFET whose gate connects to the jth neuron (i.e., the inhibition from the jth neuron). Let $I_{\mathrm{R}}^{(i)}$ and $I_{\mathrm{C}}^{(i)}$

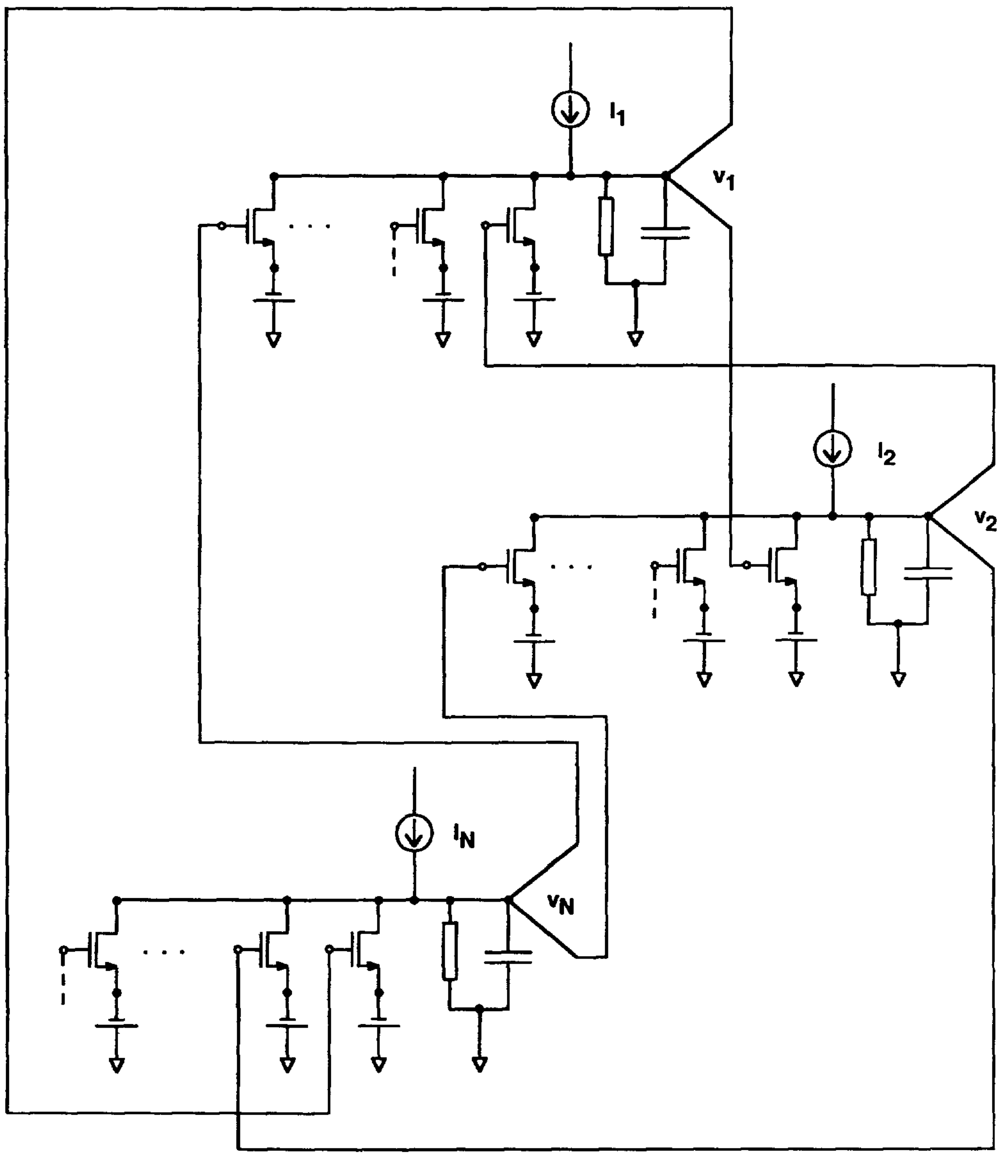

FIGURE 1. The MOSFET WTA neural network.

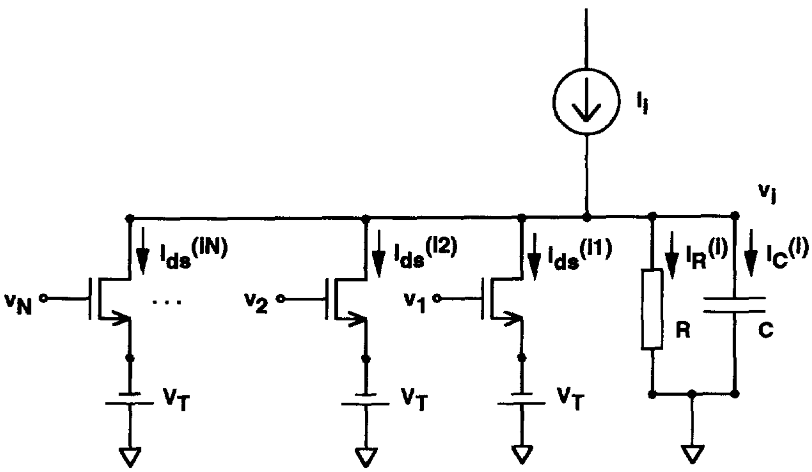

FIGURE 2. The *i*th neuron.

denote the currents through the resistor and the capacitor in the ith neuron. From Kirchhoff's current law, we have

$$I_i = I_{\rm R}^{(i)} + I_{\rm C}^{(i)} + \sum_{j\neq i} I_{\rm ds}^{(ij)} = \frac{v_i}{R} + C\frac{dv_i}{dt} + \sum_{j\neq i} I_{\rm ds}^{(ij)},$$

hence we obtain

$$C\frac{dv_i}{dt} = -\frac{1}{R}v_i + I_i - \sum_{j\neq i} I_{\rm ds}^{(ij)}. \tag{2.1}$$

From the characteristics of a MOSFET (cf. Horenstein, 1990; Sedra & Smith, 1991), we have

$$I_{\rm ds}^{(ij)} = \begin{cases} K(2(V_{\rm gs}^{(ij)} - V_{\rm T})V_{\rm ds}^{(ij)} - [V_{\rm ds}^{(ij)}]^2), \\ \qquad\qquad V_{\rm T} \leqslant V_{\rm gs}^{(ij)}, 0 \leqslant V_{\rm ds}^{(ij)} \leqslant V_{\rm gs}^{(ij)} - V_{\rm T} \\ K(V_{\rm gs}^{(ij)} - V_{\rm T})^2, \quad V_{\rm T} \leqslant V_{\rm gs}^{(ij)}, V_{\rm ds}^{(ij)} > V_{\rm gs}^{(ij)} - V_{\rm T} \\ 0, \qquad \text{otherwise} \end{cases}$$

and $V_{\rm ds}^{(ij)} = v_i + V_T$, $V_{\rm gs}^{(ij)} = v_j + V_T$. Taking these last two equations into the first equation and letting $I_{\rm ds}^{(ij)} = h(v_i, v_j)$, we have

$$I_{\rm ds}^{(ij)} = h(v_i, v_j) = \begin{cases} K[2(v_i + V_{\rm T})v_j - (v_i + V_{\rm T})^2], \\ \qquad\qquad v_j \geqslant 0, -V_{\rm T} \leqslant v_i \leq v_j - V_{\rm T} \\ Kv_j^2, \qquad v_j \geqslant 0, v_i > v_j - V_{\rm T} \\ 0, \qquad \text{otherwise} \end{cases}$$

where the physical parameter K is determined by the material and the physical shape of the MOSFET. Finally, taking this into eqn (2.1), we obtain the neural network differential equation:

$$C\frac{dv_i}{dt} = -Gv_i + I_i - \sum_{j\neq i} h(v_i, v_j), i = 1, 2, \ldots, N, \tag{2.2}$$

where $G = 1/R$ is the conductance of the resistor. For convenience, we call the function $h(x, y)$ the MOSFET function, and the network (2.2) with the MOSFET function is called the MOSFET neural network. For clarity, we rewrite the function $h(x, y)$ in the following form:

$$h(x,y) = \begin{cases} K[2(x + V_{\rm T})y - (x + V_{\rm T})^2], \\ \qquad\qquad y \geqslant 0, -V_{\rm T} \leqslant x \leqslant x - V_{\rm T} \\ Ky^2, \qquad y \geqslant 0, x > y - V_{\rm T} \\ 0, \qquad \text{otherwise,} \end{cases} \tag{2.3}$$

which is just the characteristics of a MOSFET with a battery of voltage $V_{\rm T}$ connected to the source.

In the region $y \geqslant 0$, $-V_{\rm T} \leqslant x \leqslant y - V_{\rm T}$, the function $h(x, y)$ has a multiplicative term, which corresponds to the shunting part of the network, this region is the triode region of a MOSFET. In the region $y \geqslant 0$, $x > y - V_{\rm T}$, $h(x, y)$ only depends on the y, which corresponds to the additive part of the network; this region is the constant current region of a MOSFET.

It is easily observed that the MOSFET function $h(x, y)$ satisfies the following property:

PROPERTY A. *$h(x, y)$ is a nonnegative continuous function which is monotonically nondecreasing in both variables x and y. Moreover, $h(x, y) = 0$ for any $y \leqslant 0$.*

Obviously, the neural network (2.2) is a competitive system. In certain regions, it is shunting, while in other regions it is additive. Therefore, comparing with previous implementations such as the shunting networks and additive networks (Nabet & Pinter, 1991), this is a new implementation of a neural network.

3. CONDITIONS FOR THE WTA POINT TO EXIST

The MOSFET implementation above leads us to the study of the neural network (2.2) for a general continuous function $h(x, y)$. Most importantly, we are interested in the WTA properties of the network. Let $I_1, I_2, \ldots, I_N$ be the external inputs, $I_{\rm max}$ and $I_{\rm submax}$ be the largest and second largest external inputs. We assume throughout this paper that $I_i > 0$ and $I_i \neq I_j (i \neq j, i, j = 1, 2, \ldots, N)$. We will use dx/dt and $\dot{x}$ interchangeably to denote the derivative of x whenever convenient. A point $v = (v_1, v_2, \ldots, v_N)^{\rm T}$ (where the superscript T denotes the transpose of a vector or matrix) is called a WTA point if it is an equilibrium point of (2.2) and only the component corresponding to the largest input is positive while other components are nonpositive. We let $v_{\rm max}$ denote the state variable which corresponds to the largest input $I_{\rm max}$. Then v is the WTA point if it is such an equilibrium point of (2.2) satisfying: $v_{\rm max} > 0$ and $v_j \leqslant 0 (j \neq \max)$. For convenience, let the set $C^+ = \{v = (v_1, \ldots, v_N)^{\rm T} | v_{\rm max} > 0, v_j \leqslant 0, j \neq \max\}$ be called the WTA region.

Based on this definition, in order to make the system (2.2) be a winner-take-all (WTA) network, we must guarantee that the WTA point exists so that a decision can be made where the system settles down. In this section, we present some necessary and sufficient conditions for the existence of the WTA point for the system (2.2).

We first present a result for the system (2.2) with a more general function $h(x, y)$.

THEOREM 1. *Suppose that the function $h(x, y)$ has the Property A. Then the system (2.2) has a WTA point if and only if*

$$I_j \leqslant h(0, I_{\max} R), j \neq \max, j = 1, 2, \ldots, N. \quad (3.1)$$

Proof. Without loss of generality in the proof, we assume that $I_1 > I_2 > \cdots > I_N > 0$. In this case, $I_{\max} = I_1$.

Sufficiency. Suppose that (3.1) is true; we want to show that the system (2.2) has an equilibrium point v such that $v_1 > 0$ and $v_j \leqslant 0 (j \neq 1)$. We only need to show that there is a WTA point in the WTA region $C^+ = \{v | v_1 > 0, v_j \leqslant 0, j \neq 1\}$. In fact, in this region, noticing that $h(x, y) = 0$ for $y \leqslant 0$, we have $-Gv_1 + I_1 = 0$, i.e., $v_1 = I_1 R > 0$ and $-Gv_j + I_j - h(v_j, v_1) = 0$, we only need to show that for any $j \neq 1$, the equation $-Gv_j + I_j - h(v_j, I_1 R)$ has a nonpositive solution. Let $F(x) = -Gx + I_j - h(x, I_1 R)$, then $F(x)$ is a continuous function, and $F(0) = I_j - h(0, I_1 R) \leqslant 0$ and for $T > 0$, we have $F(-T) = GT + I_j - h(-T, I_1 R) \geqslant GT - h(0, I_1 R) > 0$ for sufficiently large T. From the intermediate theorem of continuous functions, we conclude that there exists a point $v_j \in [-T, 0]$ such that $F(v_j) = 0$. This means the existence of a WTA point.

Necessity. We need to show that if the system (2.2) has a WTA point, then (3.1) must be true. In fact, suppose that (3.1) is not true, then there is $j \neq 1$ such that $I_i > h(0, I_1 R)$. Let v be the WTA point, i.e., $v_1 > 0$ and $v_j \leqslant 0 (j \neq 1)$. We know that $v_1 = I_1 R > 0$ and $-Gv_j + I_j - h(v_j, I_1 R) = 0$. In particular, $-Gv_i + I_i - h(v_i, I_1 R) = 0$. From this and the monotone nondecrease of $h(x, y)$ in x, we have $Gv_i = I_i - h(v_i, I_1 R) \geqslant I_i - h(0, I_1, R) > 0$, which contradicts the fact that $v_i \leqslant 0$. Therefore, (3.1) must be true. □

The proof is illustrated graphically in Figure 3. The graph shows that the theorem is simply a statement that $h(x, y)$ and the "load line" for R and source I_j have an intersection where $v_j \leqslant 0 (j \neq 1)$.

For the MOSFET neural network, we have the following result.

COROLLARY 1. *The MOSFET neural network (2.2) with the MOSFET function $h(x, y)$ has a WTA point if and only if either*

$$R \geqslant \max\left\{\frac{V_T}{I_{\max}}, \frac{1}{2KV_T}\frac{I_j}{I_{\max}} + \frac{V_T}{2I_{\max}}\right\}, \quad (3.2)$$

or

$$\frac{1}{\sqrt{K}}\frac{\sqrt{I_j}}{I_{\max}} \leqslant R < \frac{V_T}{I_{\max}}, \quad (3.3)$$

for all $j \neq$ *max.*

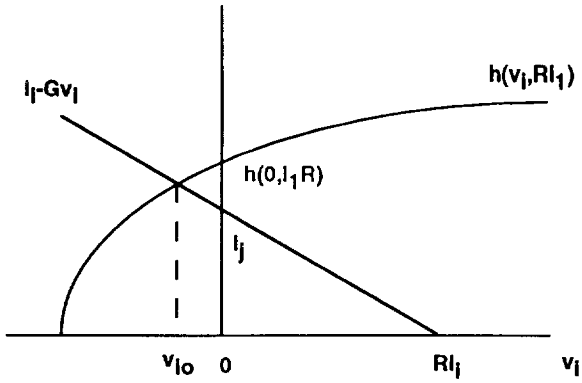

FIGURE 3. The graphic proof of Theorem 1.

Proof. For the MOSFET neural network, the function $h(x, y)$ can be specified, and we evaluate $h(0, I_{\max} R)$. Since $I_{\max} R > 0$, the corresponding MOSFET is either in the triode region or the constant current region. In the triode region, we have $-V_T \leqslant 0 \leqslant I_{\max} R - V_T$, i.e., $R \geqslant V_T / I_{\max}$. The condition (3.1) becomes

$$I_j \leqslant h(0, I_{\max} R) = 2RKV_T I_{\max} - KV_T^2,$$

i.e.,

$$R \geqslant \frac{1}{2KV_T}\frac{I_j}{I_{\max}} + \frac{V_T}{2I_{\max}}.$$

This reduces to the first condition in the corollary. Similarly, in the constant current region, we have $0 > I_{\max} R - V_T$, i.e., $R < V_T / I_{\max}$. The condition (3.1) in this case becomes

$$I_j \leqslant h(0, I_{\max} R) = K I_{\max}^2 R^2,$$

i.e.,

$$R \geqslant \frac{1}{\sqrt{K}}\frac{\sqrt{I_j}}{I_{\max}}.$$

This reduces to the second condition of the corollary. This completes the proof. □

For practical network design, the following corollary gives a useful set of sufficient conditions for the existence of a WTA point.

COROLLARY 2. *The MOSFET neural network (2.2) with the MOSFET function $h(x, y)$ has a WTA point if*

$$KV_T R \geqslant 1, \quad (3.4)$$

and

$$I_{\max} > \frac{1}{KR^2}. \quad (3.5)$$

Proof. As indicated in the proof of Corollary 1, when $I_{\max} \geqslant V_T/R$, the MOSFET is in the triode region, and condition (3.1) becomes $I_j \leqslant 2KV_TRI_{\max} - KV_T^2$, which can be guaranteed by $I_{\max} \leqslant 2KV_TRI_{\max} - KV_T^2$. Solving this for $I_{\max}$ gives

$$I_{\max} \leqslant \frac{V_T}{R}\left(\frac{KV_TR}{2KV_TR-1}\right).$$

Since $I_{\max} \geqslant V_T/R$, this condition is met if the term in the parentheses is no more than one, which is the case if $KV_TR \geqslant 1$. From Corollary 1, if $I_{\max} \geqslant V_T/R$ and (3.4) is true, there exists a WTA point.

If $I_{\max} < V_T/R$, the MOSFET is in the constant current region, and the condition (3.1) becomes $I_j \leqslant KI_{\max}^2R$, which can be assured by $I_{\max} \leqslant KI_{\max}^2R$. This is met if (3.5) is true. This completes the proof. □

The two conditions of this corollary represent a gain condition and a threshold condition, respectively. The dimensionless expression KV_TR is the incremental gain of the transistor–resistor combination when the input gate voltage is $V_T/2$. Inequality (3.4) says this gain exceeds unity to ensure the existence of a WTA point. The second condition, the inequality (3.5), says that if the largest input current $I_{\max}$ is more than a threshold $1/KR^2$, then the existence of a WTA point is guaranteed.

4. WTA BEHAVIOR AND CONVERGENCE ANALYSIS

In the last section, we provided a set of conditions for the system (2.2) to have a WTA equilibrium point. However, in order to guarantee that the neural network (2.2) is so designed as a WTA network, we have to show that the system (2.2) will settle down to the WTA point. It seems impossible to design a network (2.2) that will always converge to the WTA point for any external inputs no matter where it starts, because of the nonlinearity involved in the function $h(x,y)$. In order to get the natural winner from a closed competitive system, a fair start should be expected. A natural fair starting point is the origin, i.e., each competitor has nothing at the very beginning. In this paper, we will always start the system (2.2) from the origin, and a resetting procedure will be used whenever necessary.

There has been intensive research on the convergence of dynamical neural networks (Grossberg, 1988; Hirsch, 1989 and references therein). The Cohen–Grossberg theorem (Cohen & Grossberg, 1983) provided the most general convergence theorem for neural networks, and a general Lyapunov function construction guideline was proposed. It has been shown that the additive and shunting neural networks with symmetric interconnections between neurons use global Lyapunov functions so that convergence of the networks can be concluded.

It is not known if the system (2.2) can be converted into the Cohen–Grossberg model. Even for the MOSFET neural network (2.2), the MOSFET function $h(x,y)$ cannot be written as a sum or product of two independent functions $h_1(x)$ and $h_2(y)$; hence we cannot apply the Cohen–Grossberg theorem.

In this section, we will utilize the special structure of the MOSFET function $h(x,y)$ for the principal result, Theorem 5. However, results for a general $h(x,y)$ with Property A are obtained in the theorems leading to this result.

We first want to show the boundedness of the trajectory of the neural network (2.2).

THEOREM 2. *If the function $h(x,y)$ has the Property A, the trajectory of the neural network (2.2) is bounded.*

Proof. We only need to show that the trajectory of the system (2.2) will eventually stay in a bounded set. If $v_i > I_iR$, we have

$$\frac{dv_i}{dt} = -\frac{1}{RC}v_i + \frac{I_i}{C} - \frac{1}{C}\sum_{j\neq i} h(v_i, v_j) \leqslant -\frac{1}{RC}v_i + \frac{I_i}{C} < 0,$$

hence the $v_i(t)$ will decrease and $v_i(t) \leqslant I_iR$, i.e., $v_i(t)$ is bounded from above. If

$$v_i < I_iR - \sum_{j\neq i} Rh(0, I_jR) < 0,$$

then, from $h(v_i, v_j) \leqslant h(0, v_j) \leqslant h(0, I_jR)$, we have for sufficiently large t

$$\frac{dv_i}{dt} \geqslant -\frac{1}{RC}v_i + \frac{I_i}{C} - \frac{1}{C}\sum_{j\neq i} h(0, I_jR) > 0,$$

where we use the fact that $h(x,y)$ is nondecreasing in both x and y; hence $v_i(t)$ will increase, i.e., $v_i(t)$ is also bounded from below. This completes the proof. □

The trajectory of (2.2) also has the so-called order-preserving property which we formalize as follows.

THEOREM 3. *For any continuous function $h(x,y)$, the neural network (2.2) is order-preserving: if $I_i > I_j$ and $v_i(t_0) > v_j(t_0)$, then for any $t \geqslant t_0$, $v_i(t) > v_j(t)$.*

Proof. $\Delta(t) = v_i(t) - v_j(t)$, then $\Delta(t)$ is a continuous function and $\Delta(t_0) > 0$. Suppose that the claim in the theorem is not true, then there exists a $t^* > t_0$ such that $\Delta(t^*) = 0$ and $\Delta(t) > 0$ for $t_0 < t < t^*$. At t^*, we have $v_i(t^*) = v_j(t^*)$; thus, subtracting the jth equation from the ith equation in (2.2), we obtain

$$C\frac{d\Delta(t^*)}{dt} = -\frac{1}{R}\Delta(t^*) + (I_i - I_j) - [h(v_i(t^*), v_j(t^*)) - h(v_j(t^*), v_i(t^*))] = I_i - I_j > 0.$$

Since $d\Delta(t)/dt$ is also a continuous function, from the above inequality, there exists a small $\delta > 0 (\delta < t^* - t_0)$ such that $d\Delta(t)/dt > 0$ in the interval $[t^* - \delta, t^* + \delta]$, i.e., $\Delta(t)$ is strictly increasing in this interval, hence $\Delta(t^*) > \Delta(t^* - \delta) > 0$; this is contradictory to the choice of t^*. □

From the system equations of (2.2), it is easy to see that when the system starts from the rest state, i.e., the fair starting condition, all output voltages will increase and get into the positive orthant. As we expect, the voltage corresponding to the largest input current will initially be the largest. From the order-preserving property and the system equations, this voltage will always stay positive, since the trajectory cannot get into the negative orthant. Therefore, it suffices to study the conditions under which the system must enter and stay in the WTA region. Since the system equations have a solution when all but one voltage are negative, the following result, Theorem 2, guarantees that whenever the network enters the WTA region, the network will stay in that region and converge to the WTA point. Hence the system has at least local WTA behavior. Before we give the result, we need the following lemma.

LEMMA. (a) *(Comparison principle). Let $g(t,x)$ be a scalar continuous function, and let $m(t)$ be the solution of the scalar differential equation: $\dot m(t) = g(t, m(t))$ with $m(t_0) = m_0$. Then for any continuous function $x(t)$ satisfying the differential inequality: $\dot x(t) \leqslant g(t, x(t))$ with $x(t_0) = m_0$, we have $x(t) \leqslant m(t)$ for any $t \geqslant t_0$. Similarly, for any continuous function $x(t)$ satisfying the differential inequality: $\dot x(t) \geqslant g(t, x(t))$ with $x(t_0) = m_0$, we also have $x(t) \geqslant m(t)$ for any $t \geqslant t_0$.*

(b) *If the continuous function $y(t) \geqslant 0$ satisfying $\dot y(t) \leqslant -\alpha y(t) + g(t)$ where $\alpha > 0$ and $g(t)$ exponentially converges to zero, i.e., there exist positive numbers $M > 0$ and $\beta > 0$ such that $|g(t)| \leqslant Me^{-\beta t}$, then $y(t)$ also exponentially converges to zero. In fact, for any $\gamma < \min\{\alpha, \beta\}$, there exists positive number M_1 such that $|y(t)| \leqslant M_1 e^{-\gamma t}$.*

Proof. The rigorous version of (a) and its proof can be found in Lakshmikantham and Leela (1969). We only need to prove (b). From $\gamma < \min\{\alpha, \beta\}$, there exists a small positive number ϵ such that $\gamma + \epsilon < \min\{\alpha, \beta\}$. Then from the comparison principle, we have

$$0 \leqslant y(t) \leqslant e^{-\alpha t} y(0) + \int_0^t e^{-\alpha(t-\tau)} g(\tau) d\tau \leqslant e^{-\alpha t} y(0) + \int_0^t e^{-\alpha(t-\tau)} e^{-\beta\tau} d\tau \leqslant e^{-\gamma t} + \int_0^t e^{-(\gamma+\epsilon)(t-\tau)} e^{-(\gamma+\epsilon)\tau} \times d\tau = e^{-\gamma t}[y(0) + Mte^{-\epsilon t}].$$

However, $y(0) + Mte^{-\epsilon t}$ is bounded; therefore, there exists a positive number $M_1 > 0$ such that $y(0) + Mte^{-\epsilon t} \leqslant M_1$. Taking this into the above inequality, we can complete the proof of (b). □

THEOREM 4. *Let C^+ denote the WTA region. Suppose that the function $h(x,y)$ has the Property A; then whenever the trajectory of (2.2) starting from the origin enters the WTA region C^+, it will stay there forever. Moreover, if the function $h(x,y)$ is a Lipschitzian function, it will converge exponentially to the WTA point with convergence rate at least γ for any $\gamma < 1/RC$.*

Proof. Since $h(0,0) = 0$ and $I_i > 0$, initially, the system (2.2) will get into the positive orthant. Without loss of generality in the proof, we assume that $I_1 > I_2 > \cdots > I_N > 0$ and $C = 1$. Then, $C^+ = \{v = (v_1, v_2, \ldots, v_N)^{\mathrm T} | v_1 > 0 \text{ and } v_j \leqslant 0, j \neq 1\}$. From (2.2) and the nonnegativeness of $h(x,y)$, we have [using $\dot x(t)$ to denote the differentiation $dx(t)/dt$ for convenience] $\dot v_i(t) \leqslant -Gv_i(t) + I_i$, and

$$\dot v_i(t) \leqslant e^{-Gt} v_i(0) + \int_0^t e^{-G(t-\tau)} I_i d\tau = \frac{I_i}{G}(1 - e^{-Gt}) \leqslant \frac{I_i}{G} = I_i R.$$

Suppose the first claim in Theorem 4 is not true, i.e., the network (2.2) enters the WTA region and then gets out of the region later on. Then there exist $t_1 > 0$ and $t_2 > 0$ such that $v_2(t_1) = v_2(t_2) = 0$ [notice that in considering only v_2 we have used the order-preserving property of the trajectory of (2.2)], i.e., $v_2(t)$ gets into the region C^+ at t_1 and gets out of the region C^+ at t_2. In this case, we must have $\dot v_2(t_1) \leqslant 0$ and $\dot v_2(t_2) > 0$ [otherwise, if $\dot v_2(t_1) > 0$, then $v_2(t) > v_2(t_1) = 0$ for the time t sufficiently close to t_1 from the right, hence the network does not enter the WTA region at t_1, which contradicts the definition of t_1. A similar argument applies to the other case]. On the other hand, in $[t_1, t_2]$, we have $v_i(t) \leqslant 0 (i \neq 1)$; therefore, $\dot v_1(t) = -Gv_1(t) + I_1 \geqslant -G(I_1 R) + I_1 = 0$, so we have $v_1(t_1) \leqslant v_1(t_2)$. Moreover, from (2.2) we have

$$\dot v_2(t_2) = -Gv_2(t_2) + I_2 - \sum_{j\neq 2} h(v_2(t_2), v_j(t_2)) \leqslant I_2 - h(0, v_1(t_2)) \leqslant I_2 - h(0, v_1(t_1)) = \dot v_2(t_1) \leqslant 0,$$

where we have used the fact that $h(x,y)$ is monotonically nondecreasing in both x and y. This contradicts the fact that $\dot{v}_2(t_2) > 0$, and completes the proof of the first part.

Next, we want to show the exponential convergence. Let t_0 be the time instant which the trajectory enters the WTA region C^+. Then from the first part of the theorem, for any $t \geqslant t_0$, we have

$$\begin{aligned} C\dot{v}_1(t) &= -Gv_1(t) + I_1 \\ C\dot{v}_j(t) &= -Gv_j(t) + I_j - h(v_j(t), v_1(t)), j \neq 1. \end{aligned} \tag{4.1}$$

Solving the first equation, we obtain

$$v_1(t) = I_1R + Me^{-\alpha(t-t_0)},$$

where

$$M = v_1(t_0) + I_1R, \alpha = 1/RC.$$

Since the function $h(x,y)$ is a Lipschitzian function, there is a positive constant $L > 0$ such that $|h(x_1,y) - h(x_2,y)| \leqslant L|x_1 - x_2|$ and $|h(x,y_1) - h(x,y_2)| \leqslant L|y_1 - y_2|$. For any $j \neq 1$, we have $\dot{v}_j = -\alpha v_j + I_j/C - h(v_j, v_1)/C$. Let v^* be the WTA point, then $0 = -\alpha v_j^* + I_j/C - h(v_j^*, I_1R)$. If $e(t) = v_j(t) - v_j^*$, from the above two equations, we obtain

$$\dot{e}(t) = -\alpha e(t) - C^{-1}(h(v_j(t), v_1(t)) - h(v_j^*, I_1R)). \tag{4.2}$$

For simplicity, we let

$$g_1(t) = \frac{h(v_j(t), v_1(t)) - h\left(v_j^*, v_1(t)\right)}{v_j(t) - v_j^*},$$

$$g_2(t) = \frac{h\left(v_j^*, v_1(t)\right) - h\left(v_j^*, I_1R\right)}{v_1(t) - I_1R}.$$

From (4.2), we obtain [noticing that $v_1(t) \leqslant I_1R$]

$$\begin{aligned} \dot{e}(t) &\geqslant -\alpha e(t) - C^{-1}\left[h(v_j(t), v_1(t)) - h\left(v_j^*, v_1(t)\right)\right] \\ &\geqslant -\left[\alpha + C^{-1}g_1(t)\right]e(t). \end{aligned}$$

From the comparison principle and noticing that $g_1(t) \geqslant 0$, we can easily obtain that

$$e(t) \geqslant \exp\left(-\int_{t_0}^{t} (\alpha + C^{-1}g_1(\tau))d\tau\right) \geqslant -|e(0)|e^{-\alpha(t-t_0)}. \tag{4.3}$$

On the other hand, from (4.2), we have

$$\begin{aligned} \dot{e}(t) &= -\alpha e(t) - C^{-1}g_1(t)e(t) - C^{-1}g_2(t)(v_1(t) - I_1R) \\ &\leqslant -\left(\alpha + C^{-1}g_1(t)\right)e(t) + MC^{-1}Ke^{-\alpha(t-t_0)}. \end{aligned}$$

Following the same procedure as in the proof of Lemma (b) and the fact that $g_1(t) \geqslant 0$, we can easily obtain that there exists a positive number $M_1 > 0$ such that $e(t) \leqslant M_1 e^{-\gamma(t-t_0)}$ where $\gamma < \alpha$. Combining this and (4.3), we can obtain that for any $\gamma < \alpha$ there exists a positive number $M_2 > 0$ such that $|e(t)| \leqslant M_2 e^{-\gamma(t-t_0)}$. This completes the proof. □

This theorem also suggests a speeding procedure for decision making. If the network (2.2) has a fair start, i.e., initiates near the origin, then whenever the network enters the WTA region, we can stop. The winner is found, because the corresponding WTA region has the same property as the WTA point for the purpose of maximum selection. This may reduce the processing time significantly.

Now, in order to prove the global convergence, we only need to show that the trajectory of (2.2) starting from the origin cannot stay in the region where at least two components are positive. However, for the general function $h(x,y)$, even under the conditions in Property A, the convergence of the network (2.2) is unknown. The general Cohen–Grossberg theorem cannot be applied to the model (2.2) because of the inseparability of the function $h(x,y)$. We will only give a global convergence analysis for the MOSFET neural networks. Sufficient conditions are obtained so that the trajectory of the MOSFET neural network (2.2) will not stay forever in the region where at least two components are positive; thus the network (2.2) initiating from the origin will converge to the WTA point.

THEOREM 5. *Suppose that (3.2) or (3.3) holds. Let I_{max} and I_{submax} be the largest and the second largest inputs, respectively. Under (a), the resolution condition:*

$$I_{max} - I_{submax} > \frac{1}{4KR^2};$$

(b) either the upper boundedness condition:

$$\begin{aligned} I_{sub\,max} < KV_T[\max\{V_T, 2(I_{max} - I_{sub\,max})R \\ + 2KRV_T^2 - V_T\}], \end{aligned} \tag{4.4}$$

or the gain condition:

$$KV_TR > \frac{1}{2};$$

the trajectory of the MOSFET neural network (2.2) starting from the origin does not have an invariant set in the region where at least two components are *positive, i.e., the trajectory will not stay in such a region forever and it will enter the WTA region. Therefore, the trajectory of the MOSFET neural network initiating from the fair start will always converge to the WTA point. This implies that the network so fabricated is a WTA network.*

Proof. Without loss of generality, we assume that $C = 1$ and $I_1 > I_2 > \cdots > I_N > 0$. From the order-preserving property, we have $v_1(t) > v_2(t) \cdots > v_N(t)$. Suppose that the claim in the theorem is not true, then for any $t > 0$, we must have $v_1(t) > v_2(t) > 0$. The first two network equations in (2.2) become

$$\dot{v}_1 = -Gv_1 + I_1 - \sum_{j\neq 1} h(v_1, v_j) \tag{4.5}$$

$$\dot{v}_2 = -Gv_2 + I_2 - \sum_{j\neq 2} h(v_2, v_j). \tag{4.6}$$

Let $e(t) = v_1(t) - v_2(t)$; then we have

$$\dot{e}(t) = -Ge(t) + (I_1 - I_2) + [h(v_2, v_1) - h(v_1, v_2)] + \sum_{j\neq 1,2} [h(v_2, v_j) - h(v_1, v_j)]. \tag{4.7}$$

Since for any $x \geqslant y \geqslant 0$, we have $x \geqslant y - V_T$, hence $h(x,y) = Ky^2$. Moreover, for $y \leqslant 0$, $h(x,y) = 0$. Taking these into (4.7), we obtain

$$\dot{e} = -Ge + (I_1 - I_2) + h(v_2, v_1) - Kv_2^2. \tag{4.8}$$

If $v_2 \leqslant v_1 - V_T$, i.e., $v_2 + V_T \leqslant v_1$, then

$$h(v_2, v_1) - Kv_2^2 = K[2(v_2 + V_T)v_1 - (v_2 + V_T)^2] - Kv_2^2 \geqslant K(v_2 + V_T)^2 - Kv_2^2 > 0.$$

If $v_2 > v_1 - V_T$, then

$$h(v_2, v_1) - Kv_2^2 = Kv_1^2 - Kv_2^2 > 0.$$

Since the system (2.2) is initiated from the origin, the system starts in the region $v_2 > v_1 - V_T$, i.e., $e(t) < V_T$. In this region, we have (noticing that $v_2 > 0$)

$$\begin{aligned}\dot{e} &= -Ge + I_1 - I_2 + Kv_1^2 - Kv_2^2 \\ &= -Ge + I_1 - I_2 + K(e + 2v_2)e \\ &> -Ge + I_1 - I_2 + Ke^2 = K\left(e - \frac{G}{2K}\right)^2 \\ &\quad + \left(I_1 - I_2 - \frac{1}{4KR^2}\right) \\ &\geqslant I_1 - I_2 - \frac{1}{4KR^2} > 0,\end{aligned}$$

where we have used the resolution condition (a). Thus, for $e(t) < V_T$, $e(t)$ will eventually exceed the threshold V_T. Let T_0 be such time instant that $e(t) \geqslant V_T$ for $t \geqslant T_0$. This means that the system will stay in the region $v_2 < v_1 - V_T$; therefore, we have

$$\begin{aligned}\dot{e} &\geqslant -Ge + (I_1 - I_2) + K(v_2 + V_T)^2 - Kv_2^2 \\ &> -Ge + (I_1 - I_2) + KV_T^2.\end{aligned}$$

From the comparison principle, we obtain for $t \geqslant T_0$

$$e(t) > e^{-G(t-T_0)}e(T_0) + [(I_1 - I_2)R + KRV_T^2](1 - e^{-G(t-T_0)}),$$

i.e.,

$$v_1(t) > v_2(t) + e^{-G(t-T_0)}e(T_0) + [(I_1 - I_2)R + KRV_T^2](1 - e^{-G(t-T_0)}).$$

Then, for $t > T_0$, we have $v_2(t) \leqslant v_1(t) - V_T$, and substituting into (4.6)

$$\begin{aligned}\dot{v}_2(t) &\leqslant -Gv_2(t) + I_2 - h(v_2, v_1) \\ &= -Gv_2(t) + I_2 - K[2(v_2 + V_T)v_1 - (v_2(t) + V_T)^2].\end{aligned} \tag{4.9}$$

Suppose that the boundedness condition holds. Let

$$a(\delta) = \max\{V_T, [2(I_{\max} - I_{\text{sub max}})R + 2KRV_T^2](1 - \delta) - V_T\}.$$

If (4.4) is true, then there is a sufficiently small $\delta > 0$ such that $I_2 < KV_T[a(\delta)]^2$. Also, there exists a $T > T_0 > 0$ such that

$$e^{-G(t-T_0)}e(T_0) + [(I_1 - I_2)R + KRV_T^2](1 - e^{-G(t-T_0)}) > [(I_1 - I_2)R + KRV_T^2](1 - \delta).$$

Thus, for $t > T$, we have $v_1(t) > v_2(t) + [(I_1 - I_2)R + KRV_T^2](1 - \delta)$ and $v_1(t) \geqslant v_2(t) + V_T$. Taking this into (4.9), we have for $t > T$

$$\begin{aligned}\dot{v}_2(t) &= -Gv_2(t) + I_2 - K(v_2(t) + V_T)[2v_1(t) - (v_2 + V_T)] \\ &\leqslant -Gv_2(t) + I_2 - K(v_2(t) + V_T)(v_2(t) + a(\delta)).\end{aligned} \tag{4.10}$$

Let $m(t)$ be the solution of the equation:

$$\dot{m}(t) = -Gm(t) + I_2 - K(m(t) + V_T)(m(t) + a(\delta)), m(T) = v_2(T).$$

If $I_2 < KV_T[a(\delta)]^2$, this system has a negative stable equilibrium point (similar procedure as in the proof of Theorem 1) and when $m(T) = v_2(T) > 0$, $m(t)$ will become negative for sufficiently large t. However, using the comparison principle in (4.10), we have $v_2(t) \leqslant m(t)$ for $t \geqslant T$. Hence for sufficiently large t, $v_2(t)$ will become negative. This contradicts the assumption that $v_2(t) > 0$. Therefore, the first part of the theorem is proved if the boundedness condition holds.

Suppose that the gain condition holds, following the same procedure, we know that for sufficiently large t, we have $e(t) \geqslant V_T$, so, in a similar manner, from (4.8) we have

$$\begin{aligned} \dot{e} &\geqslant -Ge + I_1 - I_2 + h(v_2, v_1) - Kv_2^2 \\ &= -Ge + I_1 - I_2 + K[2(v_2 + V_T)v_1 - (v_2 + V_T)^2] - Kv_2^2 \\ &= -Ge + I_1 - I_2 + K(v_2 + V_T)(2v_1 - v_2 - V_T) - Kv_2^2 \\ &= -Ge + I_1 - I_2 + K(v_2 + V_T)(2e - V_T + v_2) - Kv_2^2 \\ &= (2KV_T - G)e + I_1 - I_2 - KV_T^2 + 2Kv_2 e \\ &\geqslant (2KV_T - G)e + I_1 - I_2 - KV_T^2 = ae + b, \end{aligned} \tag{4.11}$$

where $a = 2KV_T - G$ and $b = I_1 - I_2 - KV_T^2$. From the gain condition, we know that $a > 0$. Let t_0 be the time instant that $e(t_0) \geqslant V_T$, then from the comparison principle and (4.11), we have for all $t \geqslant t_0$

$$\begin{aligned} e(t) &\geqslant e^{a(t-t_0)} e(t_0) + b \int_{t_0}^{t} e^{a(t-\tau)} d\tau \\ &= -\frac{b}{a} + \left(e(t_0) + \frac{b}{a} \right) e^{a(t-t_0)}. \end{aligned} \tag{4.12}$$

Since $e(t_0) \geqslant V_T$, we have

$$\begin{aligned} e(t_0) + \frac{b}{a} &\geqslant V_T + \frac{(I_1 - I_2) - KV_T^2}{2KV_T - G} \\ &= V_T + \frac{(I_1 - I_2)R - KV_T^2 R}{2KV_T R - 1} \\ &= \frac{2KV_T^2 R - V_T + (I_1 - I_2)R - KV_T^2 R}{2KV_T R - 1} \\ &= \frac{KV_T^2 R - V_T + (I_1 - I_2)R}{2KV_T R - 1} \\ &\geqslant \frac{KV_T^2 R - V_T + \left(\frac{1}{4KR^2}\right)R}{2KV_T R - 1} \\ &= \frac{\left(KV_T R - \frac{1}{2}\right)^2}{4KR(2KV_T R - 1)} > 0, \end{aligned}$$

where we have used the resolution condition. From this and (4.12), we obtain $\lim_{t\to\infty} e(t) = +\infty$; hence $v_1(t)$ is unbounded. This contradicts the fact that the trajectory is bounded. This completes the proof for the case when the gain condition holds. The rest of the theorem can be obtained from the previous theorems. □

Notice that the resolution condition involves the difference between the largest input and the second largest input. This is reasonable, because we can easily observe that when these two are equal, the system will not go to the WTA region; the voltages corresponding to these two inputs will be identical. We also notice that the gain condition is independent of the inputs, which is a desirable property for practical design.

As we have mentioned before, in order to effectively use this kind of WTA network (2.2), we have to start from the origin, i.e., the fair starting point. One way is to use switches to reset the network to the origin whenever a new input vector is about to be tested, i.e., discharge the capacitors before the network is used. However this needs some additional circuits to implement. Fortunately, the network (2.2) has an intrinsic resetting procedure, which will be discussed next. The following result shows that we only need to switch off the input currents for some time and the network (2.2) will finish the resetting task.

THEOREM 6. *(Self-Resetting Theorem) If the function $h(x, y)$ has the Property A and also satisfies Lipschitz condition, then when the external inputs are switched off, i.e., $I_i = 0 (i = 1, 2, \ldots, N)$, the neural network (2.2) will globally exponentially converge to the origin, the fair starting point, with the convergence rate γ for any $\gamma < 1/RC$; hence the network has an intrinsic resetting property.*

Proof. When the external inputs are switched off, the network reduces to the system: (without loss of generality, we assume that $C = 1$)

$$\dot{v}_i = -Gv_i - \sum_{j \neq i} h(v_i, v_j), \quad i = 1, 2, \ldots, N. \tag{4.13}$$

First, we want to show that (4.13) has a unique equilibrium point. Obviously, the origin is an equilibrium point of (4.13). Let $v = (v_1, \ldots, v_N)^T$ be the equilibrium point. Then

$$-Gv_i - \sum_{j \neq i} h(v_i, v_j) = 0,$$

so $v_i \leqslant 0$ for any $i = 1, 2, \ldots, N$. According to the Property A, we have $h(x, y) = 0$ for $y \leqslant 0$. Therefore, we have

$$0 = -Gv_i - \sum_{j \neq i} h(v_i, v_j) = -Gv_i,$$

so $v_i = 0$, i.e., the origin is the unique equilibrium point of (4.13).

Next, we want to show that if there exists a t_1 such that $v_i(t) \leqslant 0$, then for all $t \geqslant t_1$, $v_i(t) \leqslant 0$. This is almost obvious, because if there exists another t_2 such that $v_2(t_2) \geqslant 0$, we have

$$\dot{v}_i(t_2) = -Gv_i(t_2) - \sum_{j \neq i} h(v_i(t_2), v_j(t_2)) \leqslant -Gv_i(t_2) \leqslant 0,$$

so $v_i(t)$ cannot cross the boundary.

From this argument, we obtain that there exists a $T > 0$ and m such that for all $t \geqslant T$, we have (without loss of generality, we can still use such indexing)

$$v_i(t) \geqslant 0,\ i = 1, 2, \ldots, m;\ v_j(t) \leqslant 0,\ j = m+1, \ldots, N.$$

Then the system (4.13) reduces to the following system:

$$\dot{v}_i = -Gv_i - \sum_{k \neq i}^{m} h(v_i, v_k), i = 1, 2, \ldots, m. \qquad (4.14)$$

$$\dot{v}_j = -Gv_j - \sum_{k=1}^{m} h(v_j, v_k), j = m+1, \ldots, N. \qquad (4.15)$$

From (4.14), we have $\dot{v}_i \leqslant Gv_i$, and from the comparison principle, we have $0 \leqslant v_i(t) \leqslant e^{-G(t-T)} v_i(T)$; thus, there exists an $M > 0$ such that $|v_i(t)| \leqslant Me^{-G(t-t_0)}$, i.e., (4.14) globally exponentially converges to its origin.

Because $h(x, y)$ satisfies the Lipschitz condition, there exists a constant $L > 0$ such that $h(0, y) \leqslant L|y|$. Also, since $v_j(t) \leqslant 0$ for $t \geqslant T (j = m+1, \ldots, N)$, we have $h(v_j(t), y) \leqslant |y|$. Thus, from (4.15), for $t \geqslant T$ and $j \in \{m+1, \ldots, N\}$, we have

$$\dot{v}_j(t) \geqslant -Gv_j(t) - \sum_{k=1}^{m} h(0, v_k(t)) \geqslant -Gv_j(t) - L \sum_{j=1}^{m} v_k(t).$$

Applying the same procedure as in the proof of the lemma (b) and noticing that $0 \leqslant v_k(t) \leqslant Me^{-G(t-t_0)} (k = 1, 2, \ldots, m)$, we can conclude that $v_j(t)$ will globally exponentially converge to zero with convergence rate $\gamma < 1/RC$. This completes the proof. □

It is useful to extract from the previous results a set of conditions which is easier to use for network design. The following corollary gives simplified conditions for the MOSFET neural network to always converge to a WTA point.

COROLLARY 3. *For the MOSFET neural network (2.2) with the MOSFET function defined in (2.3), under (i) the gain condition*

$$KV_T R > 1;$$

(ii) the resolution condition and lower bound condition:

$$I_{\max} - I_{\text{submax}} \geqslant \frac{1}{4KR^2}, I_{\max} > \frac{1}{KR^2};$$

the trajectory of the MOSFET network (2.2) starting from the fair condition, i.e., the origin, will exponentially converge to the WTA point, hence the network has WTA behavior. Moreover, the network also has the self-resetting property: when the network external inputs are switched off, then the network will globally exponentially converge to the fair condition; hence in order to effectively use this WTA network, switch off the inputs before applying the new inputs. The time constant for the overall network is approximately equal to the time constant of the RC circuit, i.e., RC; hence the convergence rate for the WTA network is approximately equal to 1/RC.

Proof. From Corollary 2, the gain condition and the lower bound condition guarantee the existence of a WTA point. From Theorem 5, the gain condition and resolution condition assure the convergence property. □

Remarks.

(1) The conditions for the existence of a WTA point and the convergence do not depend on the capacitance; this may be helpful in practical design. However, the capacitance does affect the convergence speed.
(2) Since if $I_{\max} - I_{\text{submax}} > 1/KR^2$, then $I_{\max} > 1/KR^2$ can be guaranteed, so simpler conditions for the MOSFET network (2.2) to have WTA behavior are (a) the gain condition $KV_T R > 1$ and (b) the resolution condition $I_{\max} - I_{\text{submax}} > 1/KR^2$. The basic design guideline is to choose R and K to be appropriately large. However, R should not be too large because that will slow down the convergence speed. A compromise has to be made.

Examples show that when the resolution, gain, and boundedness conditions are not satisfied, the network still often exponentially converges to the WTA point if it exists. We conjecture that without the resolution, gain and boundedness conditions, as long as the WTA point exists, the MOSFET network (2.2) starting from the fair starting point will always converge to the WTA point. However, we have not been able to prove this yet. This forms a fruitful research direction for the future.

5. ILLUSTRATIVE SIMULATIONS

In this section, we present a simulation to demonstrate the use of the theoretical results. A circuit

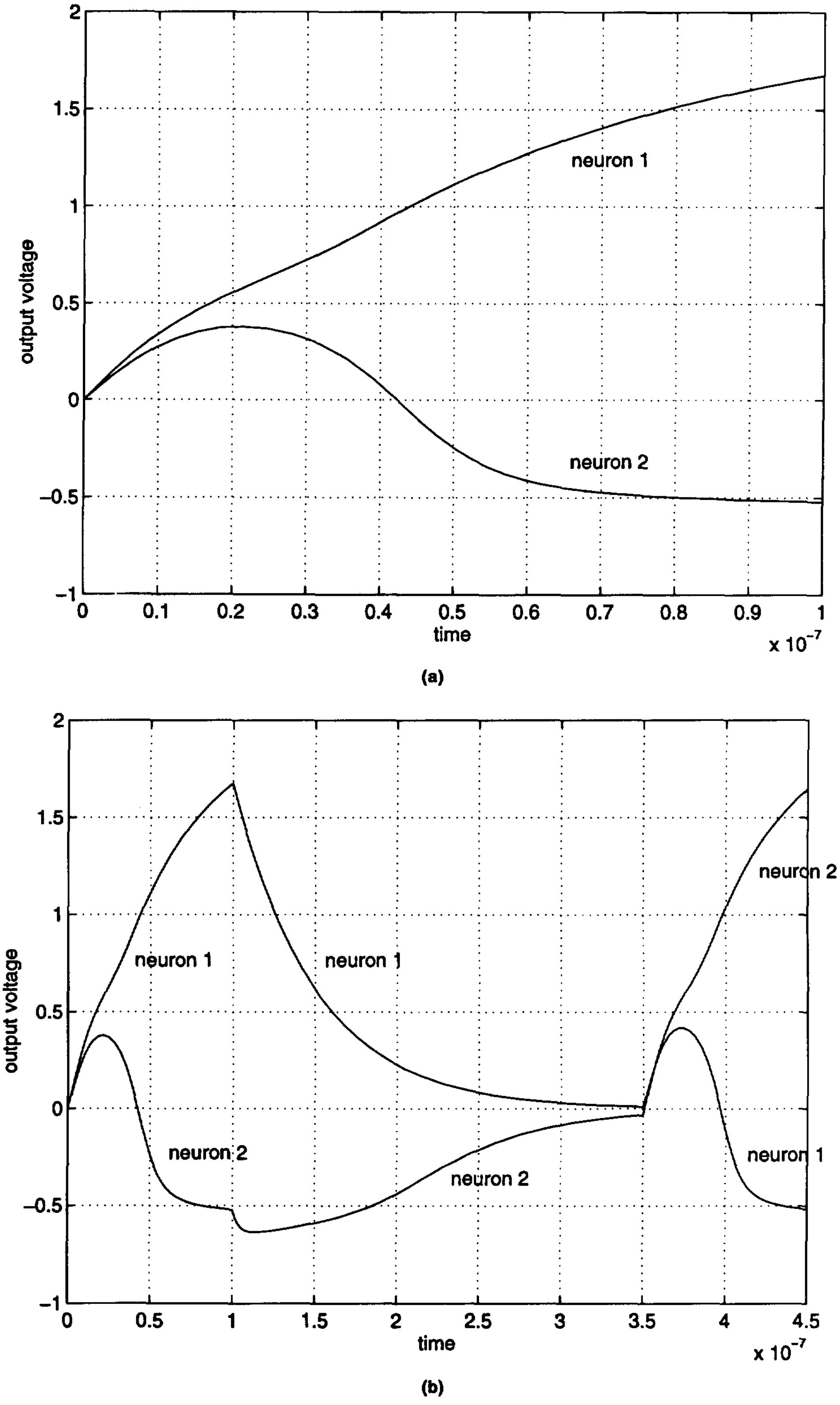

FIGURE 4(a). WTA behavior of a two-neuron MOSFET neural network; (b) WTA behavior of a two-neuron network with self-resetting.

implementation of this network has been reported by Kane and Kincaid (1995); however, the parameters were not optimized and the circuit was rather slow. In our present example, let $N = 2$, assume that $I_1 > I_2 > 0$ and use a capacitor of 1 pF. In practice, this may actually have to be an integrated capacitor to avoid using the variable stray capacitances. Choose a resistor of 100 kΩ, and a MOSFET with physical parameters $K = 40\ \mu A/V^2$ and a threshold voltage $V_T = 0.7V$. From the simplified conditions of

Corollary 3, the resolution is $I_1 - I_2 > 1/(4KR^2) = 0.625\ \mu A$, and the current lower bound is $I_1 > 2.5\ \mu A$. Since $KV_T R = 2.8 > 1$, the gain condition is satisfied. Figure 4 shows the results of the differential equation (2.2) associated with this network. The current inputs are $I_1 = 20\ \mu A$, $I_2 = 17\ \mu A$. Notice that the first condition in Corollary 1 is also easily established. Therefore, the network with these inputs will converge to the WTA point. Figure 4a is the time history of the voltage outputs. Although the network is not settling down in 100 ns, from our convergence analysis we know that a decision can be made at 43 ns, because at that instant, the network gets in the WTA region ($v_1 > 0$ and $v_2 < 0$), so the neuron 1 corresponds to the winner.

Figure 4b shows that the WTA network is self-resetting. We first test the WTA behaviour with above input currents, switch off the input at 100 ns for about 250 ns, then swap the input values between I_1 and I_2 at 350 ns. The simulation shows that at the beginning, the network picks up the winner I_1. After we switch off the currents, the network will settle down automatically to the fair starting condition. When the values of I_1 and I_2 are swapped, the network can then pick up the new winner I_2. This simulation confirms our theoretical results for the MOSFET WTA networks.

6. DISCUSSION AND CONCLUSIONS

In this paper, we have studied the modelling, analysis, and MOSFET implementation of dynamical WTA neural networks. Motivated from this, a new class of WTA networks is formalized. We have obtained a set of necessary and sufficient conditions for the existence of a WTA point for this class of neural networks; then a rigorous proof for the convergence of the MOSFET network is given. The computer simulations and experimental results show that the network so designed worked according to the theory. However, the convergence analysis of this general class of neural networks has not been done; it is still under investigation.

REFERENCES

Andreaou, A. G. et al. (1991). Current-mode subthreshold MOS circuits for analog VLSI neural networks. *IEEE Transactions on Neural Networks*, **2**(2), 205–213.

Choi, J., & Sheu, B. J. (1993). A high-precision VLSI winner-take-all circuit for self-organizing neural networks. *IEEE Journal of Solid-State Circuits*, **28**(5), 576–594.

Cohen, M. A., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel pattern formation by competitive neural networks. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-13**, 815–826.

Ermentrout, B. (1992). Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Networks*, **5**, 415–431.

Fang, W. C., Sheu, B. J., Chen, O. T. C., & Choi, J. (1992). A VLSI neural processor for image data compression using self-organizing neural network. *IEEE Transactions on Neural Networks*, **3**(3), 506–518.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, **6**, 205–254.

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

Grossberg, S. (1978). Competition, decision, and consensus. *Journal of Mathematical Analysis and Applications*, **66**, 470–493.

Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, **1**, 17–61.

Haykin, S. (1994). *Neural networks; a comprehensive foundation.* New York: Macmillan College Publishing Company.

Hirsch, M. W. (1982). Systems of differential equations which are competitive or cooperative, I: Limit sets. *SIAM Journal of Mathematical Analysis*, **13**, 167–179.

Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks*, **2**, 331–349.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences USA*, **81**, 3088–3092.

Horenstein, M. N. (1990). *Microelectronic circuits and devices.* Englewood Cliffs, NJ: Prentice-Hall.

Kane, J. S., & Kincaid, T. G. (1995). Optoelectric winner-take-all VLSI shunting neural networks. *IEEE Transactions on Circuits and Systems II*, to appear.

Kaski, S., & Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, **7**, 973–984.

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

Lakshmikantham, V., & Leela, S. (1969). *Differential and integral inequalities: theory and applications, Vol. 1: ordinary differential equations.* New York: Academic Press.

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

Lee, J. C., & Sheu, B. J. (1991). Analog neuroprocessors for early vision processing. In H. S. Moscovitz, K. Yao & R. Jain (Eds) (pp. 319–328). *VLSI signal processing IV.* New York: The IEEE Press.

Lemmon, M., & Kumar, B. V. K. V. (1989). Emulating the dynamics for a class of laterally inhibited neural networks. *Neural Networks*, **2**, 193–214.

Lippman, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, **4**, 4–22.

Mahowald, M. A., & Delbruck, T. (1989). Cooperative stereo matching using static and dynamic image features. In C. Mead and M. Ismail, *Analog VLSI implementation of neural systems.* (Eds), Norwell, MA: Kluwer Academic.

Majani, E., Erlanson, R., & Abu-Mostafa, Y. (1989). On the K-winners-take-all network. In D. S. Touretzky (Ed.), *Advances in neural information procesing systems I* (pp. 634–642). San Mateo, CA: Morgan Kaufmann.

Mead, C. A., Arreguit, X., & Lazzaro, J. (1991). Analog VLSI model of binaural hearing. *IEEE Transactions on Neural Networks*, **2**(2), 230–236.

Nabet, B., Darling, R. B., & Pinter, R. B. (1989). Analog implementation of shunting neural networks. In D. S. Touretzky (Ed.), *Advances in neural information processing systems I* (pp. 695–702). San Mateo, CA: Morgan Kaufmann.

Nabet, B., & Pinter, R. B. (1991). *Sensory neural networks: lateral inhibition.* Boston, MA: CRC Press.

Robinson, M. E., Yoneda, H., & Sanchez-Sinencio, E. (1992). A modular CMOS design of a Hamming network. *IEEE Transactions on Neural Networks*, **3**(3), 444–456.

Sedra, A. S., & Smith, K. C. (1991). *Micro-electronic circuits*, 3rd edn. New York: HRW.

Serrano, T., & Linares-Barranco, B. (1995). A modular current-mode high-precision winner-take-all circuit. *IEEE Transactions on Circuits and Systems—II: Analog and Digital Processing*, **42**(2), 132–134.

Sheu, B. J., Choi, J., & Chang, C. F. (1992). An analog neural network processor for self-organizing mapping. In *ISSCC digital technology papers* (pp. 136–137). San Francisco. CA.

Simpson, P. K. (1990). *Artificial neural networks: foundations, paradigms, applications, and implementations*. New York: Pergamon Press.

Smedley, S., Taylor, J., & Wilby, M. (1995). A scalable high-speed current-mode winner-take-all network for VLSI neural applications. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, **42**(5), 289–291.

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.

Yang, J. F., Chen, C. M., Wang, W. C., & Lee, J. Y. (1995). A general mean-based iterative winner-take-all neural network. *IEEE Transactions on Neural Networks*, **6**, 14–24.

Yen, J. C., Chang, F. J., & Chang, S. (1994). A new winner-take-all architecture in artificial neural networks. *IEEE Transactions on Neural Networks*, **5**, 838–843.

Yuille, A. L., & Geiger, D. (1995). Winner-take-all mechanism. In: M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 1056–1060). Cambridge, MA: The MIT Press.