Improving the Robustness of Winner-Take-All Cellular Neural Networks

Lachlan L. H. Andrew

***Abstract*— This paper describes two improvements on a recently proposed winner-take-all (WTA) architecture with linear circuit complexity based on the cellular neural network paradigm. The general design technique originally used to select parameter values is extended to allow values to be optimized for robustness against relative parameter variations as well as absolute variations. In addition, a modified architecture, called clipped total feedback winner-take-all (CTF-WTA) is proposed. This architecture is shown to share most properties of standard cellular neural networks, but is shown to be better suited to the WTA application. It is shown to be less sensitive to parameter variations and under some conditions to converge faster than the standard cellular version. In addition, the effect of asymmetry between the neurons on the reliability of the circuit is examined, and CTF-WTA is found to be superior.**

I. Introduction

A GENERAL inputless cellular neural network (CNN) [1], [2] is governed by equations of the form

$$\tau \frac{dx^c(t)}{dt} = -x^c(t) + \sum_{d\in\mathcal{CG}} a_d^c y^d(t) + k^c \qquad (1)$$
$$y^c(t) := f(x^c(t))$$

where f is the output function of each cell, x^c and $y^c, c = 1,\cdots,n$, are the state and output variables, respectively, and $\mathcal{CG}$ is the cell grid defined in [1], which in this case consists of all of the neurons. The time constant, τ, only affects the time scale of the solution and it will be assumed that $\tau = 1$, except in the section on convergence speed. Typically a bipolar clipping function is used [1]

$$f(x) := \tfrac{1}{2}(|x+1| - |x-1|) = \begin{cases} -1 & \text{if } x \le -1 \\ x & \text{if } |x| < 1 \\ 1 & \text{if } x \ge 1 \end{cases}.$$

It was shown in [3] that a network built up of such neurons can perform the winner-take-all (WTA) function. The symmetry of the WTA problem allows a very regular network to be used. Its parameters are given by

$$a_d^c = \begin{cases} \alpha & \text{if } c \ne d \\ \alpha+\delta+1 & \text{if } c = d \end{cases}$$
$$k^c = \kappa$$

so the dynamic equation (1) becomes

$$\tau \frac{dx^c(t)}{dt} = -x^c(t) + \alpha\sigma + (\delta+1)y^c(t) + \kappa$$
$$\sigma := \sum_{d\in\mathcal{CG}} y^d(t). \qquad (2)$$

It is stated in [3] that that such networks are limited to sizes of less than 10 neurons if reasonable fabrication tolerances are assumed. The major contributions of this work are to show two ways of overcoming this limit. First, an extension of the design algorithm used in [3] is proposed and shown to provide a nominal design whose tolerance decreases linearly with the number of nodes, rather than quadratically, as does that of [3]. Second, it will be shown that using a unipolar activation function and clipping the feedback σ produces a network whose sensitivity to component variation is independent of the network size, thus removing entirely the principal factor limiting the size of the circuit in [3]. This new circuit will be called clipped total feedback winner-take-all (CTF-WTA). Furthermore, it is argued that the CTF-WTA architecture will converge significantly faster than the system (2).

The rest of this paper is organized as follows. In Section II, improved parameters for the original architecture in [3] are derived. Section III describes the CTF-WTA architecture, and optimal parameters for CTF-WTA and their allowable tolerances are derived in Section IV. Section V shows that CTF-WTA shares most of the important convergence properties of standard CNN's, and in Section VI expressions are derived for the amount of asymmetry that is allowable in each architecture for WTA functionality to be maintained.

II. Improved Parameter Selection

Seiler and Nossek [3] applied a robust design technique to the task of choosing α, δ and κ for the WTA system described by (2). This design procedure, which is described in detail in [4], is summarized below. For notational convenience, let $\boldsymbol{p} = (\delta, \alpha, \kappa)$ denote the vector consisting of the parameters of the CNN.

1) Determine a set of N linear inequalities of the form $w_j + \boldsymbol{w}_j \cdot \boldsymbol{p} > 0, j = 1,\cdots,N$, between the parameters which is sufficient for the circuit to be functional.

2) Re-express these as

$$w_j + \boldsymbol{w}_j \cdot \boldsymbol{p} > \|\boldsymbol{w}_j\|_* r \qquad (3)$$

where $\|\cdot\|_*$ denotes the dual of the PDF norm of the manufacturing process [5], and r denotes the permissible parameter variation.

Manuscript received February 28, 1994; revised August 18, 1995. This work was supported in part by a scholarship from the Australian Telecommunications and Electronics Research Board (ATERB). This paper was recommended by Associate Editor J. L. Meador.

The author is with the Department of Electrical and Electronic Engineering at the University of Melbourne, Parkville, Victoria 3052, Australia.

Publisher Item Identifier S 1057-7130(96)02326-9.

3) Find nominal parameters $\hat{\boldsymbol{p}}$ which maximize r subject to the constraints (3).

One difficulty with this technique is that if the manufacturing process allows components to be fabricated with relative tolerances, $\|\boldsymbol{w}_j\|_*$ is dependent on $\boldsymbol{p}$, so the constraints cease to be linear. In [3] it was stated that the use of relative tolerance in design required an iterative algorithm, and the results were only "a bit" better, and so the design was completed using absolute tolerances. In what follows, a simple extension to the algorithm will be given which is shown to yield markedly better results.

After step 3 above, add the following stages:

4) Observe the *asymptotic* behavior of the parameters with network size, n, and define a new *weighted* norm, $\|\cdot\|_+$. Appropriate weights [5] are such that the variance of each weighted component of the PDF norm is equal, so the weights of the dual norm, $\|\cdot\|_+$ have the same asymptotic form as the components themselves.

5) Recalculate the final $\hat{\boldsymbol{p}}$ to maximize r subject to the new constraints where $\|\boldsymbol{w}_j\|_+$ replaces $\|\boldsymbol{w}_j\|_*$ in (3).

For the WTA circuit, optimal parameters, $\hat{\alpha}, \hat{\delta}$ and $\hat{\kappa}$, and their tolerance, $\hat{r}$, are related by [3, (4.23)–(4.25)]

$$\hat{\kappa} = n\hat{\alpha} + \hat{\delta} + \|[1, n, 1]\|_* \hat{r} \tag{4a}$$

$$\hat{\kappa} = (n-4)\hat{\alpha} - \hat{\delta} - \|[1, n-4, 1]\|_* \hat{r} \tag{4b}$$

$$\hat{\alpha} + \hat{\delta} = \|[1, 1, 0]\|_* \hat{r} \tag{4c}$$

which eventually gives asymptotic solutions $\hat{\kappa} = O(n), \hat{\alpha} = O(1)$, and $\hat{\delta} = O(1)$.

Thus the weighted dual norm becomes

$$\|(\delta, \alpha, \kappa)\|_+ = |\delta| + |\alpha| + n|\kappa|. \tag{5}$$

Substituting this into (4) gives

$$\hat{r} = -\frac{2}{|n-4| + 3n + 6}\underline{\alpha} \tag{6a}$$

$$\hat{\delta} = -\frac{|n-4| + 3n + 10}{|n-4| + 3n + 6}\underline{\alpha} \tag{6b}$$

$$\hat{\kappa} = \frac{(n-1)|n-4| + 3n^2 - n - 12}{|n-4| + 3n + 6}\underline{\alpha} \tag{6c}$$

and the permissible ranges become

$$\delta \in (\hat{\delta} - \hat{r}, \hat{\delta} + \hat{r}) \tag{7a}$$

$$\alpha \in (\hat{\alpha} - \hat{r}, \hat{\alpha} + \hat{r}) \tag{7b}$$

$$\kappa \in (\hat{\kappa} - n\hat{r}, \hat{\kappa} + n\hat{r}). \tag{7c}$$

(Note that the absolute tolerance of κ has been increased by a factor of n, so its relative tolerance is comparable with those of α and δ.)

This produces a useful improvement of 5 fold to 48 neurons at 1% tolerance. However the real gains come when the net sizes increase, since sensitivity is now *linear* in n, so at 0.1%, a 498 neuron net can be implemented, compared to 31 for absolute tolerances. These gains are achieved with no modification to the original cellular architecture, merely by choosing appropriate parameters.

Note that the above are the sizes which are *guaranteed* to be attainable with the stated tolerances. In real implementations, it is likely that many values will be well within tolerance and thus larger nets can often be made in practice. (Hardware implementations of a similar architecture with 170 neurons have already been reported [6], with no indication that larger networks cannot be made.) Appropriate nominal values, however, will always improve yield beyond that obtainable with suboptimal nominal values.

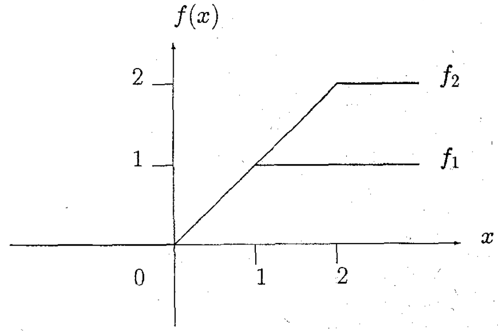

Fig. 1. Unipolar clipping functions for CTF-WTA.

III. Clipped Total Feedback

In this section, a novel WTA architecture based on the cellular circuit of [3] will be presented. The new architecture, clipped total feedback WTA (CTF-WTA), differs from that proposed in [3] in two key aspects. First, neurons are unipolar. This means that neurons which have ceased to compete make no contribution to the total feedback, σ. Second, the total feedback, σ, is clipped to the range [0, 2] (hence the name). This means that the control signals used to cause neurons to change are bounded by limits independent of the number of neurons, n. As will be seen in Section IV, these simple changes produce an architecture such that, when optimal nominal parameters are used, the precision required to realize the parameters is independent of n.

Define the following unipolar clipping functions (Fig. 1):

$$f_2(x) := 1 + f(x-1)$$
$$f_1(x) := f_2(2x)/2.$$

The new dynamical equation will be of the form

$$\tau\frac{dx^c(t)}{dt} = -x^c(t) + \alpha f_2(\sigma) + (\delta+1)y^c(t) + \kappa \tag{8}$$
$$\sigma := \sum_{d\in\mathcal{CG}} y^d(t)$$
$$y^c(t) := f_1(x^c(t)).$$

IV. Parameter Selection for CTF-WTA

In this section, the algorithm described in section II will be used to select optimal parameters for the new architecture. For ease of description, let a *WTA state* be a state such that there is one neuron, c, such that $y^c = 1$ and $y^d = 0$ for all $d \neq c$. For a circuit to perform the WTA function, a state must be stable if and only if it is a WTA state.

Stage 1 of the algorithm requires that inequalities be determined which are sufficient for correct circuit operation. The *first* such inequality in this case is the binary output condition (section V), which guarantees that all final outputs are either 0 or 1:

$$\delta + \alpha \geq 0. \tag{9}$$

For the state in which all outputs are zero not to be an equilibrium state, it is necessary that the derivative in (8) be positive so that one of the outputs can become positive. Now in this state $y^c = 0, f_2(\sigma) = 0$ and $x^c \leq 0$, so this condition reduces to

$$\kappa \geq 0. \tag{10}$$

In any state in which two or more neurons output a 1, at least one of these must decrease so that eventually only one neuron outputs a 1. This neuron will satisfy $x^c \geq 1$ and $y^c = 1$, and in this state $\sigma \geq 2$, so $f_2(\sigma) = 2$. Thus from (8)

$$\frac{dx^c(t)}{dt} \leq -1 + 2\alpha + (\delta + 1) + \kappa \leq 0$$

whence

$$\kappa + \delta \leq -2\alpha. \tag{11}$$

Finally, the condition that WTA states must be stable follows automatically. The binary output condition guarantees that the network will end up in a binary output state. Since the only such states which are not forced to be unstable by the above inequalities are WTA states, and since all WTA states are equivalent by symmetry, all WTA states must be stable.

Notice that all of these parameter constraints are independent of the number of neurons, n. Following [3], the robustness of the system to errors in these parameters is found by fitting a maximal norm-body (of radius $\hat{r}$) into the polytope formed by the above inequalities and the further artificial requirement that

$$\alpha \geq \underline{\alpha}.$$

Thus in stage 3, $\hat{r}$ must be maximized subject to

$$\begin{aligned} \hat{\kappa} &\geq \|[0,0,1]\|_* \hat{r} \\ -\hat{\delta} - 2\hat{\alpha} - \hat{\kappa} &\geq \|[1,2,1]\|_* \hat{r} \\ \hat{\alpha} + \hat{\delta} &\geq \|[1,1,0]\|_* \hat{r} \\ \hat{\alpha} &\geq \underline{\alpha}. \end{aligned}$$

These are satisfied with equality if

$$\hat{r} = -\underline{\alpha}/7 \tag{12a}$$
$$\hat{\kappa} = -\underline{\alpha}/7 \tag{12b}$$
$$\hat{\delta} = -9\underline{\alpha}/7. \tag{12c}$$

Since these are all independent of n, the weighting of $\|\cdot\|_+$ in stage 4 will be uniform, and stage 5 will give the same results as stage 3, so (12) is the final set of optimal nominal parameters for the CTF-WTA. Here $\hat{r}$ is the tolerance on each α, δ and κ, and $\underline{\alpha}$ is a free parameter.

V. Convergence

Since the dynamics of the CTF-WTA are not the same as the dynamics of standard CNN's, standard convergence results must be verified. The results in this section establish that the most important properties of CNN's also apply to CTF-WTA networks.

A. Guarantee of Convergence

Following [1], convergence will be proved by showing that the state variables x^c are bounded and thence that the outputs y^c converge. These proofs will follow those presented in [1] closely, and so only differences from those proofs will be noted.

Boundedness

In the notation of this paper, (4a) in [1] can be replaced for the CTF-WTA by

$$\frac{dx^c}{dt} = -x^c(t) + \underline{f}^c(t) + \kappa$$

where

$$\underline{f}^c(t) = \alpha f_2(\sigma) + (\delta + 1) y^c.$$

This has solution

$$x^c(t) = x^c(0)e^{-t} + \int_0^t e^{-(t-\tau)}(\underline{f}^c(\tau) + \kappa)\, d\tau.$$

From this it can be shown that, if $0 \leq x^c(0) \leq 1$,

$$|x^c| \leq |x^c(0)| + \max_t(|\underline{f}^c(t) + \kappa|)$$

giving the bounds

$$|x^c| \leq r^c = 2 + 2|\alpha| + |\delta| + |\kappa|.$$

Convergence of Outputs

Convergence of the outputs can then be proved by a proof similar to that for the convergence of a standard CNN.

Define the function

$$F(x) := \begin{cases} 0 & \text{if } x \leq 0 \\ x^2/2 & \text{if } 0 < x < 2 \\ 2x - 2 & \text{if } x \geq 2 \end{cases} \tag{13}$$

such that $F'(x) = f_2(x)$. Then the function

$$E(\boldsymbol{y}) = \frac{\delta}{2}\sum_c (y^c)^2 + \kappa \sum_c y^c + \alpha F\left(\sum_c y^c\right) \tag{14}$$

is bounded with $\max_t |E(\boldsymbol{y}(t))| \leq E_{\max}$ where

$$E_{\max} = n(|\delta|/2 + |\kappa|) + |\alpha| F(n).$$

In addition

$$\begin{aligned} \frac{dE(\boldsymbol{y}(t))}{dt} &= \sum_{0 < x^c < 1} \left[\frac{dx^c(t)}{dt}\right]^2 \\ &= \sum_c \left[\frac{dy^c(t)}{dt}\right]^2 \\ &\geq 0. \end{aligned}$$

Thus $E(\boldsymbol{y}(t))$ is a bounded monotonic function of time, and hence converges to a limit as $t \to \infty$. The corollary that $\boldsymbol{y}$ converges to a constant is also still valid.

If the piecewise linear output function is replaced with a continuous, strictly monotonic function, f, such as is likely to be produced by a VLSI implementation, then the function may be replaced by

$$E(\boldsymbol{y}) = -\sum_c \int_0^{y^c} f^{-1}(v)\, dv + \frac{1+\delta}{2}\sum_c (y^c)^2 + \kappa \sum_c y^c + \alpha F\left(\sum_c y^c\right). \tag{15}$$

B. Convergence to Binary Outputs

An essential element of a WTA network is that the final output should consist of binary values (± 1 or 0, 1); it should be clear which node has won and which have not. The standard CNN has this property when the magnitude of the self feedback of each node is greater than 1 [2, Theorem 2]. It is easily shown that the CTF-WTA has the same property under the same conditions. This can be proved by replacing $e(t)$ in Theorem 2 of [2] with the function

$$e(t) = \alpha\left(f_2\left(\sum_d y^d\right) - y^c\right) + \kappa$$

and the intervals $(-1, +1)$ and $[-r^c, -1]$ with $(0, +1)$ and $[-r^c, 0]$, respectively. In the proposed architecture, the self feedback is $1+\alpha+\delta$, so the binary output condition is $\alpha+\delta > 0$, as stated in Section IV.

C. Speed of Convergence

It has been pointed out [7] that, when the cross-coupling between neurons is fixed, the rate of convergence drops dramatically as the number of neurons whose activations are above the minimum threshold decreases (as it must in a functional WTA circuit). The change of convergence rate is important since the *peak* rate of change is often limited by supply voltage or current limits in hardware realizations. Thus a fast initial response followed by a slower tail would require the time constants of the entire circuit to be increased. Yen and Chang [7] proposed altering the weights of the cross-coupling to ensure a constant convergence rate. Introducing a clipping nonlinearity at the output of the global adder, σ, in the CTF-WTA performs this task in an eminently realisable manner. By limiting the initial response it allows a small time constant, τ, to be used leading to faster overall convergence.

Robust design dictates that the system must operate reliably for any combination of inputs. Denote the smallest τ permissible in the CTF-WTA by τ_C and that in the nonclipped WTA by τ_N. For the original architecture, the worst case, in which the fastest rate of change occurs, is when all but one of the inputs are $+1$ and the other $x^c = -1$, giving $\tau_N\, dx^c/dt = 1 + (n-2)\alpha - (\delta + 1) + \kappa = (2n-3)\alpha$ using the values of δ and κ from [3]. In comparison, the worst case in the proposed architecture is again for all inputs but one $+1$, now with the other $x^c = 0$, giving $\tau_C\, dx^c/dt = 2\alpha + \kappa = 13\alpha/7$ for $n > 2$.

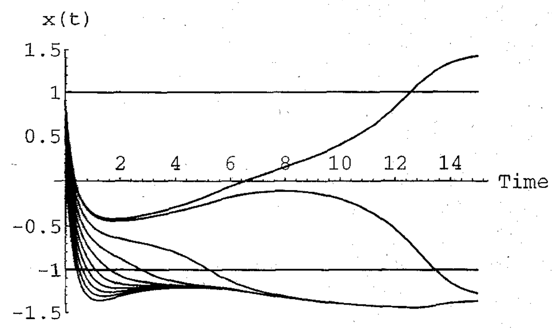

Fig. 2. Convergence of a standard WTA network, $\tau = 1$. Convergence is attained when all state variables, x, are out of the range [-1, 1].

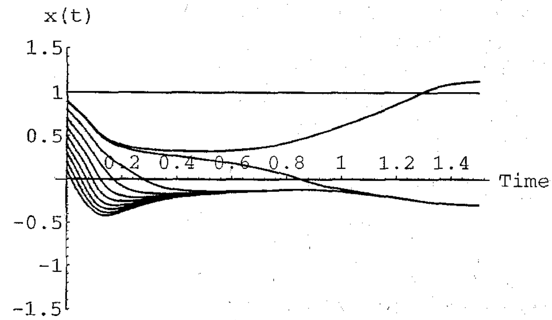

Fig. 3. Convergence of a CTF-WTA network, $\tau = 0.1$. Convergence is attained when all state variables, x, are out of the range [0, 1].

Equating rates gives $|\tau_N| = 7(2n-3)|\tau_C|/13 \approx n|\tau_C|$. It is shown in the appendix that the worst case total convergence time for the CTF-WTA is bounded above by

$$t_C \le \frac{7\tau_C}{9|\alpha|}\log\left(\frac{1}{x_1(0) - x_2(0)}\right) + \frac{4\tau_C}{|\alpha|}$$

which is independent of the network size, while that of the nonclipped WTA is bounded below by

$$t_N \ge \tau_N/|\alpha| \approx n\tau_C/|\alpha|.$$

Since τ_N must grow linearly with n for a constant initial rate of change, the convergence time of the conventional network is effectively linear in network size, meaning that the total convergence time is much smaller for the proposed architecture for substantial networks.

Figs. 2 and 3 show the convergence behavior of a standard WTA network, and a CTF-WTA network, respectively. Each has 10 inputs (0.9, 0.89, 0.8, 0.7, $\cdots$, 0.2, 0.1) and is designed with optimal parameters (those for the standard network being those of [3] rather than Section II from this paper) with $\alpha = -0.35$. Time constants have been chosen to correspond to an equal worst-case initial rate of change ($\tau_N = 1, \tau_C = 0.1$). Note the difference in scale for the time axes. From this it can be seen that the CTF-WTA network converges in approximately 1.3 arbitrary units of time, while the standard WTA network takes over 13.

VI. Asymmetry

In [3] it was shown that the symmetry of the WTA circuit preserves the order of the activations of the neurons. However, in a VLSI implementation the symmetry will not be exact due to component mismatch. Conditions must therefore be found under which the standard WTA and CTF-WTA will reliably select the largest input as the winner. The smaller the initial difference between inputs which can be detected reliably, the more robust the network is to asymmetry.

Consider the case when each parameter has a variation of $\pm m/2$, where m is assumed to be much smaller than any of the nominal values. Assume without loss of generality that x_1 corresponds to the largest input and x_2 to the second largest. Define the variable $z = x_1 - x_2$. Winner take all functionality is maintained as long as z increases monotonically ($dz/dt \geq 0$). Let α_1, δ_1 and κ_1 denote the exact parameters for neuron 1, and α_2, δ_2 and κ_2 denote those for neuron 2.

In the case of the standard circuit it is required that

$$\begin{aligned}\frac{dz}{dt} &= (\alpha_1 - \alpha_2)\sigma + \delta_1 x_1 - \delta_2 x_2 + \kappa_1 - \kappa_2 \\ &\geq -|m\sigma| + (\delta - m/2)z - mx_2 - m \\ &\geq 0.\end{aligned}$$

For this to be true at all times for all initial conditions, this requires that $-mn + (\delta - m/2)z - 2m \geq 0$. Thus it is required that $z \geq 2m(n+2)/(2\delta - m) \approx m(n+2)/\delta$, so sensitivity increases linearly with network size.

For the CTF-WTA, it is required that

$$\begin{aligned}\frac{dz}{dt} &= (\alpha_1 - \alpha_2)f_2(\sigma) + \delta_1 x_1 - \delta_2 x_2 + \kappa_1 - \kappa_2 \\ &\geq -2m + (\delta - m/2)z - mx_2 - m \\ &\geq 0\end{aligned}$$

which is satisfied by $z \geq 8m/(2\delta - m) \approx 4m/\delta$, so sensitivity is independent of network size.

This shows that as well as improving the robustness to errors in the mean value of κ, δ and α, CTF-WTA also provides improved robustness to mismatch between neurons.

VII. Conclusion

A new WTA architecture called Clipped Total Feedback Winner-Take-All (CTF-WTA) has been presented. Its convergence has been proved and verified by simulation. It has been shown using a method due to Seiler and Nossek [3] to be totally scalable, in that parameter tolerances and variable ranges are independent of the number of neurons, and has been shown to converge faster than a conventional WTA layer and be less sensitive to component mismatch.

Appendix

This appendix proves the results used in Section V-C. All initial activations will be assumed to be within the linear region of the activation function, [−1, 1] for the standard architecture or [0, 1] for CTF-WTA. To simplify notation, much of this section will use a normalized time variable, $T = t/\tau$. To analyse the time taken for the network to converge, this time is broken into two parts: the time, T_1, during which both of the two largest neuron activations are in the linear region, and the time $T_2 - T_1$ for the remaining neuron to leave the linear region.

It will be useful to note that while x_i and x_j are both in the linear region, their difference obeys the equation

$$\frac{d}{dT}(x_i - x_j) = \delta(x_i - x_j)$$

which, if $x_i > x_j$, has solution

$$x_i(T) - x_j(T) = (x_i(0) - x_j(0))e^{\delta T} \geq x_i(0) - x_j(0). \quad (16)$$

In the remainder of this section it will be assumed, without loss of generality, that the neurons are numbered in decreasing order, so that x_1 is the largest and x_n the smallest.

A. Convergence Time of CTF-WTA

In this section an upper bound will be derived for the convergence time of the CTF-WTA network, using the nominal parameters of section IV. Since $|x_1 - x_2| < 1$ while x_1 and x_2 are in the linear region, [0, 1], an upper bound on T_1 can be found from (16) to be

$$\overline{T_1} = \frac{1}{\delta}\log\left(\frac{1}{x_1(0) - x_2(0)}\right). \quad (17)$$

There are now two possibilities: either x_1 has reached 1, or x_2 has reached 0. If $x_2(T_1) = 0$, then x_1 continues according to

$$\frac{dx_1}{dT} = (\alpha + \delta)x_1 + \kappa$$

which has solution

$$x_1 = \frac{-\kappa}{\alpha + \delta} + \left(x_1(T_1) + \frac{\kappa}{\alpha + \delta}\right)\exp((\alpha + \delta)(T - T_1))$$

thus reaching 1 at time

$$\begin{aligned}T_2 &= T_1 + \frac{1}{\alpha + \delta}\log\left(\frac{\alpha + \delta + \kappa}{x_1(T_1)(\alpha + \delta) + \kappa}\right) \\ &\leq T_1 + 4/|\alpha| \end{aligned} \quad (18)$$

by (12) and since $x_1(T_1) \geq 0$. Alternatively, if $x_1(T_1) = 1$, then x_2 continues according to

$$\frac{dx_2}{dT} \leq (x_2 + 1)\alpha + \delta x_2 + \kappa. \quad (19)$$

This is satisfied with equality if $x_3(T_1) = 0$. An upper bound on convergence time is obtained by solving (19) with equality, giving

$$x_2 = \frac{-\alpha - \kappa}{\alpha + \delta} + \left(x_2(T_1) + \frac{\alpha + \kappa}{\alpha + \delta}\right)\exp((\alpha + \delta)(T - T_1))$$

so, again by (12) and since $x_2(T_1) \leq 1, x_2 = 0$ by time

$$\begin{aligned}T_2 &= T_1 + \frac{1}{\alpha + \delta}\log\left(\frac{(\alpha + \kappa)}{(\alpha + \kappa) + x_2(T_1)(\alpha + \delta)}\right) \\ &\leq T_1 + 3/(2|\alpha|). \end{aligned} \quad (20)$$

By (17), (18), and (20) the total convergence time for the CTF-WTA circuit is

$$t_2 \leq \frac{7\tau_C}{9|\alpha|} \log\left(\frac{1}{x_1(0) - x_2(0)}\right) + \frac{4\tau_C}{|\alpha|} \tag{21}$$

where τ_C is the time constant for the CTF-WTA circuit.

B. Convergence Time of Conventional WTA

This section presents a lower bound on the worst case convergence time of a bipolar lateral inhibition WTA network with no feedback clipping. The actual parameters used will be those specified in [3], but similar results would be obtained with all such networks.

Let x_3 reach -1 at time T_3. The least upper bound on $x_2(T_3)$ is 1 since the greatest lower bound on T_3 is 0. After T_3, there are $n-2$ neurons contributing -1 to σ, and y_1 contributes at most 1, so $\sigma \leq 3 - n + x_2$, which gives

$$\frac{dx_2}{dT} \geq (3 - n + x_2)\alpha + \delta x_2 + \kappa \tag{22}$$

$$= \left(\frac{-2}{n+2}\alpha\right)x_2 + \left(\frac{n}{n+2}\alpha\right) \tag{23}$$

using $\delta = -(n+4)\alpha/(n+2)$ and $\kappa = (n^2-6)\alpha/(n+2)$ from [3]. At any time, T, x_2 must be greater than another variable, z, satisfying (22) with equality and which has the same initial condition. Such a z satisfies

$$z(T) = \frac{n}{2} + \left(x_2(T_3) - \frac{n}{2}\right)\exp\left(\frac{-2\alpha(T - T_3)}{n+2}\right)$$

and reaches -1 at

$$\underline{T}_2 = T_3 + \frac{n+2}{|2\alpha|}\log\left(\frac{n+2}{n - 2x_2(T_3)}\right).$$

Thus $\underline{T}_2$ is a lower bound on the worst case time to convergence of x_2. But in the case of $x_2(T_3) = 1$,

$$\underline{T}_2 = T_3 + \frac{n+2}{|2\alpha|}\log\left(1 + \frac{4}{n-2}\right)$$
$$> T_3 + 1/|\alpha|$$

since $(n+2)\log(1+4/(n-2)) > 2$ for $n > 2$. Thus the worst case convergence time is bounded below by

$$\underline{t}_2 > \tau_N/|\alpha| \tag{24}$$

where τ_N is the time constant for the nonclipped WTA circuit.

Acknowledgment

The author wishes to thank the reviewers for comments which have greatly improved the presentation of this paper.

References

[1] L. O. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 1257–1272, Oct. 1988.

[2] J. A. Nossek, G. Seiler, T. Roska, and L. O. Chua, "Cellular neural networks: Theory and circuit design," *Int. J. Circuit Theory Applicat.*, vol. 20, pp. 533–553, Sep. 1992.

[3] G. Seiler and J. A. Nossek, "Winner-take-all cellular neural networks," *IEEE Trans. Circuits Syst.- II*, vol. 40, pp. 184–190, Mar. 1993.

[4] G. Seiler, A. J. Schuler, and J. A. Nossek, "Design of robust cellular neural networks," *IEEE Trans. Circuits Syst.- I*, vol. 40, pp. 358–364, May 1993.

[5] R. K. Brayton, G. D. Hachtel, and S. W. Director, "Arbitrary norms for statistical design via linear programming," in *Proc. IEEE ISCAS-78*, New York, NY, 1978, pp. 161–164.

[6] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of $O(N)$ complexity," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, 1989, pp. 703–711.

[7] J.-C. Yen and S. Chang, "Improved winner-take-all neural network," *Electron. Lett.*, vol. 28, pp. 662–664, Mar. 1992.

Lachlan L. H. Andrew received the B.E. degree in electrical engineering and the B.Sc. degree in computer science from the University of Melbourne, Australia, in 1992 and 1993, respectively.

He was awarded an IEE undergraduate prize in 1992. He is currently studying toward the Ph.D degree at the University of Melbourne. His research interests include neural networks, image coding, and vector quantization.