Recurrent neural network as a KWTA selector: a synthesis procedure

Ruxandra L. Costea and Corneliu A. Marinov
Dept. of Electrical Engineering
Polytechnic University of Bucharest, Romania
Email: ruxandra.costea@upb.ro; cmarinov@rdslink.ro

Abstract—A recurrent continuous time Hopfield network is rigorously designed to find the K largest elements in a list of currents. Simple formulae link the given parameters (as list separation and the processing time) with computed constants (as gain of amplifiers and interconnection conductance). The results unveil ways of improving circuit performances.

I. INTRODUCTION

Sorting the K largest elements from a longer list is a fundamental operation in data processing. The analog solution relies on the huge parallelism of neural circuits. In this respect, the recurrent (or Hopfield type) neural networks, started in [1] and quickly found applications in KWTA sorting [2], [3], [4], [20]. Recently, in [5], [6], [21]-[24] the KWTA operation has been transformed into an optimization problem which, in turn, was solved by a continuous time recurrent network. Other KWTA algorithms of discrete time type can be found in [7]-[11]. Concrete hardware realizations can be found in [12], [13] and [18]. The K = 1 case has a larger bibliography, see [17] and [19].

Our work continues the preoccupations in [14]-[17] toward a rigorous theory and design of a KWTA machine processing efficiently an uninterrupted sequence of lists. The elements of each list are electric currents and they are admitted and treated simultaneously (or “in parallel”) by the circuit. After processing one list, the network needs resetting, [17], and then a new list is admitted. We search for analytical relations between circuit and list parameters such that a synthesis procedure can lead to a network with feasible parameters and sorting large lists in a short time. The main theoretical result is in Theorem 1 (Section III), which gives inequality type constraints to obtain a KWTA asymptotically stable steady state. Its binary KWTA splitting can be “read” at an imposed time when the transients surpass a ± threshold. Section IV builds a design scenario with given and computed parameters related by simple formulae. Section V discusses the results and give some numerical tests.

II. FUNDAMENTALS

A list \( D = \{d_1, d_2, \ldots, d_N\} \) of distinct elements is given and its KWTA splitting \( D = D_a \cup D_b \) should be revealed, where \( D_a \) gathers the K largest elements and \( D_b \) the rest of them. In other words, we have to signal the indices \( \sigma(1), \sigma(2), \ldots, \sigma(K) \) from \( D_a = \{d_{\sigma(1)}, \ldots, d_{\sigma(K)}\} \) and the rest of indices in \( D_b = \{d_{\sigma(K+1)}, \ldots, d_{\sigma(N)}\} \) where \( d_{\sigma(i)} > d_{\sigma(j)} \) when \( i \in \overline{1, K} \), \( j \in \overline{K+1, N} \). To fulfill this task, our network in Figure 1 with outputs \( V = \{v_1, \ldots, v_N\} \), will perform the binary splitting \( V = V_a \cup V_b \) where \( V_a = \{v_{\sigma(1)}, \ldots, v_{\sigma(K)}\} \), \( V_b = \{v_{\sigma(K+1)}, \ldots, v_{\sigma(N)}\} \) and \( v_{\sigma(i)} \approx m \) for any \( i \in \overline{1, K} \) and \( v_{\sigma(j)} \approx \Delta \) for any \( j \in \overline{K+1, N} \), where \( m \) is a given voltage level. The list \( D \) to be processed should be confined to the interval \([0, d_{\max}]\) and we call “separation” the shortest distance between elements: \( \Delta = \min\{d_i - d_j \mid i, j \in \overline{1, N}, i \neq j\} \). The “relative separation” \( z = \frac{\Delta}{d_{\max}/(N-1)} \) is \( z \in (0, 1] \) expresses how crowded the \( d_i \)-s are comparing to the uniform distribution distance.

![KWTA circuit](image)

Referring to Fig. 1, each of the \( N \) cells consists of an ideal nonlinear voltage amplifier described by \( v_i = mg(\lambda u_i) \).

Here \( m \) is the amplitude, \( g(x) = \frac{\exp x - \exp(-x)}{\exp x + \exp(-x)} \) is the hyperbolic tangent and \( G = m\lambda \) is the open-loop amplifier gain, i.e. the characteristic slope in \( u_i = 0 \). Any transient of \( v_i \) transposes equivalently into a transient of \( u_i = \frac{1}{2\lambda} \ln \frac{m - v_i}{m + v_i} \) and vice-versa.

Each cell sends a negative feedback to all other cell inputs. With \( p \) the interconnection conductance and \( l = (N - 1)p \), with \( C \) the total capacitance at each input and with \( M \) a bias...
current, the circuit is described by
\[ C \frac{d}{dt} u_i = -lu_i - p \sum_{j=1}^{N} v_{ij} + d_i + M \]  
(1)

By denoting \( u_{ij}(t) = u_i(t) - u_j(t) \) and similarly for \( v_{ij} \) and \( d_{ij} \), from (1) we get
\[ C \frac{d}{dt} u_{ij} = -lu_{ij} + pv_{ij} + d_{ij} \]  
(2)

The steady states of components \( u_i(t) \) and \( v_i(t) \) are denoted by \( \bar{u}_i \) and \( \bar{v}_i \), respectively, and obviously \( \bar{v}_i = mg(\lambda \bar{u}_i) \) and
\[ 0 = -lu_i - p \sum_{j=1}^{N} v_{ij} + d_i + M \]  
(3)

Also, we immediately get
\[ 0 = -\bar{v}_{ij} + p\bar{v}_i + d_{ij} \]  
(4)

where \( \bar{u}_{ij} = \bar{u}_i - \bar{u}_j \) and \( \bar{v}_{ij} \) alike.

III. THEORY

Lemma 1: If \( u_i(0) = u_j(0) \) and \( d_i > d_j \) then \( u_i(t) > u_j(t) \) for all \( t > 0 \) and \( \bar{u}_i > \bar{u}_j \).

Proof: Due to \( \frac{d}{dt} u_{ij}(0) = d_{ij} > 0 \), there is \( T > 0 \) such that \( u_{ij}(t) > 0 \) on \( (0,t) \). Then \( \frac{d}{dt} u_{ij} > -lu_{ij} + d_{ij} \) on \( (0,T) \), giving \( u_{ij}(t) > \left[ 1 - \exp \left( -\frac{t}{C} \right) \right] d_{ij} / l \). From here \( u_{ij}(T) > 0 \), such that \( u_{ij}(t) > 0 \) on \( (0,\infty) \). The steady state inequality \( \bar{u}_i \geq \bar{u}_j \) is obvious, while the hypothesis \( \bar{u}_i = \bar{u}_j \) would imply \( d_i = d_j \).

In what follows we suppose \( d_{\sigma(i)} > d_{\sigma(i+1)} \) all over, and, for simplicity, we write "\( \bar{v} \)" instead of "\( \sigma(i) \)" as list or voltage indices. Thus, the list is ordered as
\[ d_{\text{max}} \geq d_1 > d_2 > \cdots > d_N \geq 0 \]  
(5)

and, by virtue of Lemma 1 we search for conditions assuring
\[ \bar{v}_1 > \bar{v}_2 > \cdots > \bar{v}_K > \xi > 0 > -\xi > \bar{v}_{K+1} > \cdots > \bar{v}_N \]  
(6)

Due to \( \bar{v}_i = mg(\lambda \bar{u}_i) \), (6) is equivalent with
\[ \bar{v}_1 > \bar{v}_2 > \cdots > \bar{v}_K \geq \beta > 0 > -\beta \geq \bar{v}_{K+1} > \cdots > \bar{v}_N \]  
(7)

Here \( \xi \) is the KWTA threshold, related with \( \beta \) by \( \xi = mg(\lambda \beta) \). We are looking for \( \xi \geq 0.99m \) which, together with \( -m < \bar{u}_i < m \), assures the binary KWTA result \( \bar{v}_i \simeq v_i \simeq 0 \) and \( \bar{v}_{K+1} \simeq \cdots \simeq \bar{v}_N \simeq -m \).

We denote
\[ a = 1 - z \frac{N}{N-1} \]  
(8)

\[ M_1 = l\beta + (K - 1) pm - p\xi (N - K) - (N - K) \Delta \]  
(9)

\[ M_2 = -l\beta - (N - K - 1) pm + p\xi K - d_{\text{max}} + K\Delta \]  
(10)

\[ F = \frac{2l\beta + pm (N - 2) + ad_{\text{max}}}{pmN} \]  
(11)

We also put \( T_p \), the first moment when \( v_K(T_p) \geq \xi \) and \( v_{K+1}(T_p) \leq -\xi \). Due to Lemma 1, this means that at \( T_p \) we have exactly the KWTA splitting of steady state (6) or (7). At this moment the transient process can be halt and the result can be read.

**Theorem 1:** If the following restriction are met
\[ z < \frac{N - 1}{N} \]  
(12)

\[ 2pm - ad_{\text{max}} > 0 \]  
(13)

\[ \beta < \min \left\{ \beta_1 = \frac{\Delta}{2l} \left( 1 - \exp \left( -\frac{t}{C} \right) \right); \beta_2 = \frac{2pm - ad_{\text{max}}}{2l} \right\} \]  
(14)

\[ \lambda > \max \left\{ \lambda_1 = \frac{1}{2\beta} \ln \frac{1 + F}{1 + F} + \lambda_2 = \frac{1}{\beta} \ln \frac{4m}{(N - 1)\beta} \right\} \]  
(15)

\[ M \in [M_1, M_2] \]  
(16)

then our network has an asymptotically stable KWTA steady state. This splitting can be read at \( T_p \), moment when the \( u_i(t) \) variables overcome \( \pm \beta \) values.

Proof: (sketch) First, we can easily see that \( \lambda > \lambda_1 \) imposes \( M_1 < M_2 \). Also (13) implies \( \beta_2 > 0 \). Note also that \( F < 1 \) due to \( \beta < \beta_2 \). Now suppose \( \bar{v}_K \leq \xi \) or, equivalently \( \bar{v}_K < \beta \). From (4) written for the \((K, K+j)\) pair, we get \( \bar{v}_{K+j} < -j - \bar{v}_K < -\beta \) for \( j \in [1, K-1] \), \( K > 1 \), where \( \beta < \beta_1 < \frac{\Delta}{2l} \) has been used. Thus \( \bar{v}_{K+j} < -\xi \) for

\[ j \in \{1, N-K\} \]

Now we take the \( K \)-th equation in (3), split the sum of \( v_i \)-s in \( \sum_{j=1}^{K-1} \bar{v}_{K-j} \) and \( \sum_{j=1}^{K} \bar{v}_{K+j} \), and derive
\[ 0 < l\beta + p(K - 1)m + p(N - K)(-\xi) - (N - K)\Delta - M \]

This contradicts \( M > M_1 \) and shows that \( \bar{v}_K > \xi \).

Further on, we consider the \( K+1 \)-th equation from (3) and use the above proven \( \bar{v}_1 > \bar{v}_2 > \cdots > \bar{v}_K > \xi \) and \( \bar{v}_j > -m \) for \( j > K \), as well. We obtain
\[ |\bar{v}_{K+1} < -pK\xi + pm(N - K - 1) + d_{\text{max}} - K\Delta + M \]

i.e.
\[ |\bar{v}_{K+1} < -l\beta - M_2 + M < -l\beta \]

which is exactly \( \bar{v}_{K+1} < -\xi \). Next we deal with the stability issue and observe that the Jacobian of the system (1) computed in \( \bar{u} \), \( i \in [1, N] \) can be written as \( J = -AQ \) where \( Q = \text{diag} \{ pm\lambda(\bar{u}_i) \}; i \in [1, N] \). Each non-diagonal term of \( A \)
equals 1 while the diagonal terms are \( 1 + \frac{1 - mp\lambda(\bar{u}_i)}{pmN\lambda(\bar{u}_i)} \). We can show that \( A \) is positive definite if \( \lambda > \lambda_2 \). This proves the negativity of real parts of eigenvalues of \( J \) [25].
Finally, from (2) written for \( i = K \) and \( j = K + 1 \) we derive
\[
u_{K,K+1}(t) > \frac{\Delta}{T} \left[ 1 - \exp\left( -\frac{l}{C}t \right) \right]
\]
for any \( t \in (0, \infty) \).

If \( T_p \) is the stopping time, then we have to impose \( u_{K,K+1}(T_p) > 2\beta \). This leads to the restriction \( \beta < \beta_1 \).

**IV. Synthesis**

In this section we set up a design procedure starting from the following given parameters:

- \( N \) - number of list elements, \( N \geq 3 \);
- \( K \in \mathbb{N}, N - 1 \) - the number of ranks of largest elements of \( D \);
- \( z \) - the relative separation of \( D \), \( z \in \left( 0, \frac{N-1}{N+1} \right) \);
- \( T_p \) - the stopping time for the transient process;
- \( m \) - the amplifier output saturation amplitude;
- \( C \) - total ground capacity per cell.

The network admits the list \( D \) inside an interval \([0, d_{\text{max}}]\) and adds to each current a bias \( M \). With a certain gain \( G \) of amplifiers and a certain interconnection conductance \( p \), the output voltages are split in \( T_p \) seconds towards \( z \approx m \) and \(-\xi \approx -m \) in a KWT.A manner near a stable steady state.

The mentioned parameters \( d_{\text{max}}, p, \xi, G \) and \( M \) are computed as follows:

\[
d_{\text{max}} \geq \frac{2mC}{z(N-1)} \ln \frac{1}{z}
\]

\[
p > p_{\text{min}} = \frac{C}{z(N-1)} \ln \frac{1}{z}
\]

\[
\beta \leq \beta_{\text{max}} = \frac{mz}{(N-1)^2}
\]

\[
G > G_{\text{min}} = \frac{(N-1)^2}{z} \ln \frac{4(N-1)}{z}
\]

\[
\xi = m \tanh \left( \frac{G}{m} \right)
\]

\[
\xi \geq 0.99m \text{ for } G \beta \geq G_{\text{min}} \beta_{\text{max}}
\]

\[
M \in [M_1, M_2]
\]

where \( M_1 \) and \( M_2 \) have the expressions in (9) and (10).

These formulae are derived from the restrictions in Theorem 1 in a manner described briefly below.

Due to (13) we take \( p = S m d_{\text{max}} \frac{2m}{(N-1)^2} \) where \( S > 1 \) will be precised later. With this, \( \beta_2 = \frac{(S-1)m}{S(N-1)} \) while \( \beta_1 = \frac{\Delta}{2l} = \frac{mz}{S a(N-1)^2} \). In order to have \( \beta_1 \leq \beta_2 \) we have to take \( S \geq \frac{1 - \frac{a}{l}}{\frac{1}{C}T_p} \). To meet (14) we put \( \beta = s\beta_1 \) with
\[
s \leq \frac{1 - \frac{a}{l}}{\frac{1}{C}T_p}
\]

with this value of \( \beta \), and with \( p \) above, we compute \( F, \lambda_1 \) and \( \lambda_2 \) in (15). It can be proven that for \( z < \frac{N-1}{N+1} \) and \( s \leq 1 - z \) we have \( \lambda_2 > \lambda_1 \). Thus \( \lambda > \lambda_2 \) and choosing \( S = \frac{1 - z}{s} \) and \( s = 1 - z \) we get
\[
G = m\lambda > m\lambda_2 = G_{\text{min}} \text{ i.e. (20}).
\]

Also, from \( s = 1 - z < 1 - \exp\left( -\frac{1}{C}T_p \right) \) we get (17).

Now \( g(\lambda_{\text{min}}, \lambda_{\text{max}}) = g \left[ \ln \frac{4(N-1)}{z_{\text{min}}} \right] \approx 1 - \frac{2}{x^3 + 1} \)

where \( x = \frac{4(N-1)}{z} \). Due to \( x^2 \) \( \gg \) \( 1 \) we have
\[
g(\lambda_{\text{min}}, \lambda_{\text{max}}) \approx 1 - \frac{1}{8} \frac{z}{(N-1)^2}. \text{ If } N 
\]

we immediately show that \( \xi = m \left[ \frac{G_{\text{min}} \beta_{\text{max}}}{m} \right] \geq 0.99m \).

**V. Comments and Numerical Tests**

The formulae (17)-(23) show:

- the minimum gain of amplifiers depends on \( N \) and \( z \) only, increasing with \( N^2 \) and decreasing with list separation. The gain is influenced neither by \( T_p \) - the processing time, nor by \( K, m \) and \( C \);
- the interconnection conductance \( p_{\text{min}} \) increases for shorter times \( T_p \) and smaller separation \( z \). It does not depend on \( K \) or \( m \);
- let us take \( M = \frac{M_1 + M_2}{2} \) in (23). With \( z \approx m \) and
\[
p \approx p_{\text{min}} = \frac{d_{\text{max}}}{2m} (1 - z) \text{ this gives }
\]
\[
M = -\frac{d_{\text{max}}}{2} \left[ (N-2K) \left( 1 - \frac{z}{N-2} \right) + 1 \right]
\]

The bias current is the only parameter depending on the number of winners \( K \): \( K < \frac{N}{2} \) means \( M < 0 \) and \( K > \frac{N}{2} \) gives \( M > 0 \). The size \( |M| \) has a minimum value of \( \frac{d_{\text{max}}}{2m} \) for \( K = N/2 \) and increases both sides for \( K \neq N/2 \), especially for small separations \( z \);
- it is clear that the (17)-(23) recipe can lead to optimal performances if proper combinations of parameters are chosen.

Let’s consider a numerical example with the given parameters are in Table I-left side. The incoming list we use is given by
\[
d_i = \begin{cases} (2i - 1) \Delta, & i \in 1,50 \; \text{\( \Delta \) timesteps} \\ 2(i - 1) \Delta, & i \in 51,100 \end{cases}
\]

With \( K = 40 \), the largest elements are \( \{d_{31}, d_{32}, \ldots, d_{50}, d_{61}, d_{62}, \ldots, d_{100} \} \). By using the formulae (17)-(23) we compute the parameters in Table I-right side and with \( u_i(0) = 0 \), \( i \in 1,100 \) we solve (MATLAB ODE113) the dynamic system (1). Cutting the dynamic evolution at
TABLE I
PARAMETERS SET UP

<table>
<thead>
<tr>
<th>Given parameters</th>
<th>Computed parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>100</td>
</tr>
<tr>
<td>$d_{max}$</td>
<td>3.4746 × 10^{-7} Amp</td>
</tr>
<tr>
<td>$K$</td>
<td>40</td>
</tr>
<tr>
<td>$p$</td>
<td>1.22 × 10^{-8} /Ohm</td>
</tr>
<tr>
<td>$z$</td>
<td>0.3</td>
</tr>
<tr>
<td>$T_p$</td>
<td>10^{-6} s</td>
</tr>
<tr>
<td>$m$</td>
<td>10 V</td>
</tr>
<tr>
<td>$C$</td>
<td>$10^{-12}$ F</td>
</tr>
</tbody>
</table>

TABLE II
DIVERTING FROM COMPUTED PARAMETERS

<table>
<thead>
<tr>
<th>Chosen</th>
<th>$d_{max}$</th>
<th>$p$</th>
<th>$G$</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$3 \times 10^{-7}$</td>
<td>$10^{-8}$</td>
<td>$10^4$</td>
<td>$-2 \times 10^{-6}$</td>
</tr>
<tr>
<td>Number of winners</td>
<td>39</td>
<td>37</td>
<td>39</td>
<td>42</td>
</tr>
</tbody>
</table>

$T_p = 1 \mu$s we read the $v_1(t)$ voltages above $+\xi$ and those below $-\xi$. We clearly obtain the winners $v_{11}(T_p) = \cdots = v_{50}(T_p) = v_{81}(T_p) = \cdots = v_{100}(T_p) \simeq m = 10 \ V$ and the losers $\simeq -m$. This shows the correct $40 \ - WTA$ separation.

Now we disregard the rules (17)-(23) by modifying slightly the values of certain parameters - Table II. Thus we take $d_{max} = 3 \times 10^{-7}$ instead of at least $d_{max} = 3.4746 \times 10^{-7}$, leaving all other parameters unchanged, i.e. as in Table I. The result of processing (read at $T_p$) gives 39 winners instead of 40. Thus the $KWTA$ machine gives a wrong result. Similarly, we modify in turn $p$, $M$ and $G$ (keeping all others) and each time the selection is wrong - Table II. This shows that our recipe (17)-(23) gives good bounds for the correct computed parameters.

VI. CONCLUSION
Simple formulae to synthesize a $KWTA$ network with a recurrent structure were given and checked by numerical tests. They allow a flexible design for good performances.

ACKNOWLEDGMENT
The work has been co-funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557 and by CNCSIS-UEFISCUS, project number PN II-RU code PD_32/2010.

REFERENCES