A Current -Mode DTCNN Universal Chip

Hubert Harrer[1]
Institute for Network
Theory and Circuit Design
TUM, D-80290 Munich
Germany

Josef A. Nossek
Institute for Network
Theory and Circuit Design
TUM, D-80290 Munich
Germany

Tamás Roska
Computer and Automation
Research Institute
Hungarian Academy of Sciences
H-1518 Budapest, Hungaria

Leon O. Chua
Dept. of Electr. Eng.
and Comp. Sciences
UCB, Berkeley CA 94720
USA

ABSTRACT

The paper describes an analog current mode realization of Discrete-Time Cellular Neural Networks (DTCNNs) with high cell density, which have local analog and local logic memory. Hence, some important parts of the CNN Universal Machine concept are implemented. The computation speed can be adjusted simply to the application by changing the clock rate. The circuit components are described in detail and SPICE level 2 simulation results are given for the ORBIT 2.0 μm process. A layout has been designed for a chip with 12 by 12 cells on a square grid realizing a one-neighborhood with 9 feedback and 9 control coefficients. The cell size is 619 μm by 425 μm and the simulated speed is between 1MHz and 10MHz depending on the minimum value of the state current. For the latter this leads to a simulated performance of 25.9 10^9 XPS for a single chip operation with an effective area of 0.379 cm^2 and a worst case power consumption of 0.86 W. Another important feature of the chip is its capability for a spatial cascaded connection.

1. INTRODUCTION

The Cellular Neural Network (CNN), invented in [2], is a nonlinear dynamic array processor, where the elementary processor cells are connected within a final spatial neighborhood only. The CNN Paradigm is now a general framework in case of different cell-, grid-, and interaction types and different modes of operations. This array, combined with local logic, is the first stored program analogic array computer [7]. Discrete-Time Cellular Neural Networks have the same local connectivity structure with translational invariant weights, but are time-discrete and have a threshold function [8] as nonlinearity. The local cell connectivity leads to efficient VLSI realizations. In the present paper an analog current mode implementation of a DTCNN is given, where the important features of the chip layout are

- an efficient realization of the feedback coefficients by cascoded current mirrors
- the use of a simple four-quadrant multiplier for the control coefficients
- the use of an efficient current comparator
- additional local analog and logic memory
- spatial cascadability of several chips on a board supported by a fast data transfer of specific boundary cells for the binary outputs

In Section 2 the network architecture is discussed and the circuits of the single components are given in Section 3. Section 4 describes the layout and the simulation results.

[1] supported by a DFG stipend at the Department of Electrical Engineering and Computer Sciences, UCB in 1993

2. NETWORK ARCHITECTURE

Multiple-Layer Discrete-Time-Cellular Neural Networks are defined by the following recursive algorithm

$$x_l^c(k) = \sum_{d \in N_r(c)} a_l^{c,d}(k) y_l^d(k) + \sum_{d \in N_r(c)} b_l^{c,d}(k) u_l^d(k) + i_l(k) \quad (1)$$

$$y_l^c(k) = f(x_l^c(k-1)) = \begin{cases} 1 & \text{for} \quad x_l^c(k) \geq 0 \\ -1 & \text{for} \quad x_l^c(k) < 0, \end{cases} \quad (2)$$

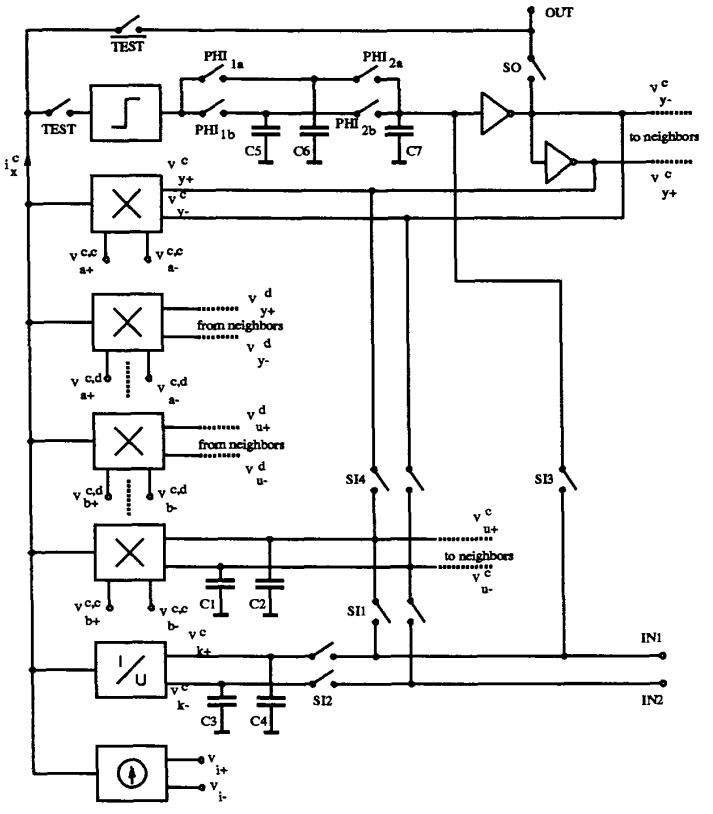

Figure 1: Block structure for a single cell.

$$k_l^c(k) = \sum_{d \in N_r(c)} b_l^{c,d}(k) u_l^d(k) + i_l(k) \qquad (3)$$

if time-variant templates and inputs are assumed [3]. The variables and coefficients denote:

$x_l^c(k)$:	cell state	$a_l^{c,d}(k)$:	feedback coeff.
$y_l^c(k)$:	cell output	$b_l^{c,d}(k)$:	control coeff.
$u_l^c(k)$:	cell input	$i_l^c(k)$:	threshold
$k_l^c(k)$:	cell bias	$N_r(c)$:	r-neighborhood
c:	cell index	d:	neighbor index
l:	layer index		

The block diagram is given in Fig. 1 for a single cell, which is similar to that in [4]. Since analog memories can be realized simply by capacitors, the feedback- and control coefficients as well as the cell input and cell output are implemented as voltages. The differential structure takes advantage of a higher accuracy and a particular compensation of disturbances such as the feedthrough effect from switching transistors or cross-coupling effects of signal lines. The core of each cell consists of 9 multipliers for the feedback coefficients and 9 multipliers for the control coefficients. They all have a common current output, which performs the summation in (1) and represents the cell state. In addition to the threshold, which is implemented by a current source controlled by the differential control voltage v_{i+} and v_{i-}, a local analog memory is included in each cell. It can be used for offset compensation or simple motion detection tasks.

The cell input $v_{u\pm}^c$ and the analog memory are read in by two global bus lines IN1 and IN2, when activating the signals SI1 or SI2. They are stored on the capacitors C_1 and C_2, or C_3 and C_4, respectively. Thus, a hole column is loaded in parallel, which accelerates the data transfer. The control signal SI3 is used to read in the initial value $v_y^c(0)$ from the same input bus IN2. The signal SI4 allows a parallel data transport of all cells to the cell input.

Each cell has an implemented test modus, in which the state current can be connected to the global output bus OUT for analog measurements. During normal operation the current is lead to a comparator, which decides the sign and extracts the binary outputs. The switching into the subsequent output state is performed by the signals φ_{1a}, φ_{2a}, φ_{1b} and φ_{2b}. The capacitors C_5 and C_6 realize two local logic memories. Depending on the switch configuration only one of them is connected to C_7 and determines the output state for the following iteration. This enables a sequential processing of multiple layers.

The complementary value is generated by an inverter and both signals are lead to the corresponding feedback multipliers of adjacent cells. The outputs are read out column by column by selecting SO.

At the boundaries of the regular grid specific border cells have been designed, which provide the cell input and output and decode the control signals SI1 to SI3, TEST and SO from a common bus. Besides, four specific link lines enable a sequential data transfer of the cell outputs from boundary cells and allow a fast processing, when several chips are connected on a board in a spatial cascade.

3. CIRCUIT COMPONENTS

A simple circuit for realizing the multiplication of the binary outputs with the feedback coefficients is given in Fig. 2.

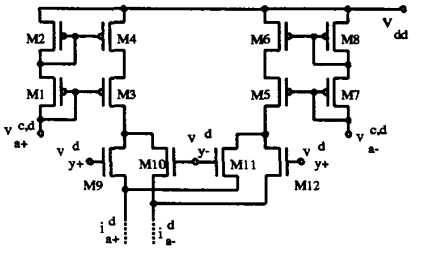

Figure 2: Circuit structure for the feedback multipliers.

It consists of only two cascaded current mirrors, whose outputs are connected to four switch transistors. They are controlled by the binary outputs from neighboring cells. The current is switched to two current lines i^d_{a+} and i^d_{a-}. Only one of the two current mirrors is active depending if the weight is positive or negative. A single-ended current is generated by an additional current mirror realized by n-transistors. It has to be implemented only once for a cell, since the two current lines sum up the outputs of all feedback multipliers. This realization of the feedback coefficients implies a nonlinear relationship between the output current and the global weight voltage. In first approximation it is described by

$$v^{c,d}_{a+} = V_{dd} - \sqrt{\frac{2i^d_{a+}}{c_4}}\left(1 + \sqrt{\frac{c_2}{c_1}}\right) - 2V_{Tp} \qquad (4)$$

with the simplified square-law model for the saturated transistors. Here, V_{Tp} denotes the p-channel threshold voltage for identical bulk and source voltages and $c_{1,2,4}$ is defined by

$$c_i = \frac{W_i}{L_i}\mu C_{ox}, \qquad (5)$$

where W_i and L_i give the transistor width and length, μ is the mobility of the charge carriers and C_{ox} is the oxide capacitance per unit area.

For higher accuracy, a look up table should be used instead of (4) for the mapping between weight voltage and weight current, which is obtained from measurement results. As the weights are applied from outside the chip, this nonlinear characteristic is not disturbant. For $V_{dd} = 8$ V, the dynamic range of the weight voltages is between 4.56 V and 6.22 V for a maximum current of 10 μA. Here, the dc power consumption amounts to 80 μW. The same circuit is used without the switch transistors for the realization of the constant threshold, too.

In contrast to the feedback coefficients (A-template), the realization of the B-template needs four-quadrant multipliers with a good linearity for the analog inputs. A circuit structure, which makes a good compromise in chip area, accuracy and power consumption, is introduced in [1].

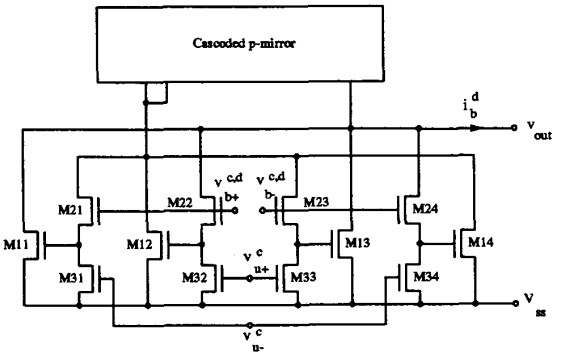

Figure 3: Circuit structure for the control multipliers.

Its schematic is shown in Fig. 3. If all transistors are operated in saturation, the output current is described by

$$i^d_b = c(v^{c,d}_{b+} - v^{c,d}_{b-})(v^c_{u+} - v^c_{u-}) \qquad (6)$$

for identical transistor geometries. Here, only one current mirror has been used to reduce the power consumption. Since the output voltage v_{out} is kept approximately constant for small currents, the error due to channel length modulation is very small. The circuit has an input range of 3.1 V $\leq v^{c,d}_{b+}, v^{c,d}_{b-} \leq$ 4.1 V, and 1 V $\leq v^c_{u+}, v^c_{u-} \leq$ 2 V for obtaining an output current $-9.68\ \mu\text{A} \leq i^d_b \leq 9.59\mu$ A. The simulated full scale error of $v^{c,d}_u$ amounts to 1.58 % for $\Delta v^{c,d}_b = 1$ V and 1.92 % for $\Delta v^{c,d}_u = 1$ V, respectively. The maximum dc power consumption is $P_{max} = 400\ \mu$W.

This circuit can also implement a linear transconductance, if only the left part of Fig. 3 is used with a bias voltage v_{ref} instead of $v^{c,d}_{b+}$. It provides a very high input resistance, which is requested by the local analog memory in Fig. 1.

The sign of the state current i_x^c, which represents the binary output v_y^c is extracted by a current comparator. An efficient circuit, which combines the advantages of a resistive input comparator with that of a capacitive input comparator, is given in [5], [6]. Its schematic is shown in Fig. 4.

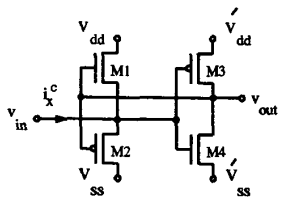

Figure 4: Circuit structure for the current comparator.

The circuit has a small chip area, high accuracy, fast transient behavior, independence of fabrication tolerances and renders an approximately constant input voltage for $i_x^c = 0$.

4. LAYOUT AND SIMULATION RESULTS

The layout has been designed for the ORBIT 2.0 μm process implementing 12 by 12 regular cells and 52 surrounding border cells. The cell geometry is 425 μm by 619 μm. It includes the whole connectivity structure and all global bus lines. A network is simply built up by placing the cells to a regular grid. The dc power consumption for a single cell was simulated to 1.89 mW for zero weights and reaches a maximum value of 4.4 mW for the worst case.

The maximum clock frequency for φ_1 and φ_2 depends on the minimum absolute state current i_x^c, which can appear for the worst case. This state current has to charge the parasitic capacitances of the summing current line, until the output of the current comparator can detect the change of the sign. Since the voltage is amplified by the inverter of the current comparator, it is very small for the critical range (about 0.1 V). Table 1 gives the time of the transient response for different state currents. For this simulation only a negative self-feedback coefficient $v_{a-}^{c,c}$ has been chosen to generate the state current. It shows that the maximum clock frequency can be chosen between 1 MHz and 10 MHz for an application.

The loading-time of the analog inputs has been simulated to about 30ns for an accuracy of 0.5 %. This includes the signal delay caused by the decoder circuits. The storage time amounts to 7.5ms for an accuracy of 0.5 %.

$\Delta i_a^{c,c}$	T	$\Delta i_a^{c,c}$	T	$\Delta i_a^{c,c}$	T
0.1 μA	985 ns	0.2 μA	680 ns	0.4 μA	456 ns
0.9 μA	302 ns	4.9 μA	119 ns	9.9 μA	75 ns

Table 1: Settling time of the output voltage of the comparator for different weight currents.

The outputs of the boundary cells can be transferred within 14 × 50ns = 700ns between different chips.

5. REFERENCES

[1] K. Bult and H. Wallinga. A cmos four-quadrant analog multiplier. *IEEE Journal of Solid State Circuits*, 21:430–435, 1986.

[2] L. O. Chua and L. Yang. Cellular neural networks: Theory. *IEEE Transactions on Circuits and Systems*, 35:1257–1272, 1988.

[3] H. Harrer. Multiple layer discrete-time cellular neural networks using time-variant templates. *IEEE Transactions on Circuits and Systems II*, 40:191 – 199, March 1993.

[4] H. Harrer, J. A. Nossek, and R. Stelzl. An analog implementation of discrete-time cellular neural networks. *IEEE Transaction on Neural Networks*, 3:466–477, 1992.

[5] A. Rodriguez-Vazques, R. Dominguez-Castro, F. Medeiro, J. L. Huertas, and M. Delgado-Restituto. High resolution cmos current comparators and piecewise-linear current-mode circuits. *Internal Report.*

[6] A. Rodriguez-Vazques, S. Espejo, R. Dominguez-Castro, J. L. Huertas, and E. Sanchez-Sinencio. Current-mode techniques for the implementation of continuous and discrete-time cellular neural networks. *IEEE Transactions on Circuits and Systems II*, 40:132–146, March 1993.

[7] T. Roska and L. O. Chua. The cnn universal machine: An analogic array computer. *IEEE Transactions on Circuits and Systems II*, 40:163–173, March 1993.

[8] C. L. Sheng. *Threshold Logic.* Academic Press, London, 1969.