BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
Vol. 54, No. 4, 2006

Transresistance CMOS neuron for adaptive neural networks implemented in hardware

R. WOJTYNA* and T. TALAŚKA

Institute of Telecommunication, University of Technology and Agriculture, 7 Kaliskiego St., 85-796 Bydgoszcz, Poland

Abstract. A simple analog circuit is presented which can play a neuron role in static-model-based neural networks implemented in the form of an integrated circuit. Operating in a transresistance mode it is suited to cooperate with transconductance synapses. As a result, its input signal is a current which is a sum of currents coming from the synapses. Summation of the currents is realized in a node at the neuron input. The circuit has two outputs and provides a step function signal at one output and a linear function one at the other. Activation threshold of the step output can be conveniently controlled by means of a voltage. Having two outputs, the neuron is attractive to be used in networks taking advantage of fuzzy logic. It is built of only five MOS transistors, can operate with very low supply voltages, consumes a very low power when processing the input signals, and no power in the absence of input signals. Simulation as well as experimental results are shown to be in a good agreement with theoretical predictions. The presented results concern a 0.35 μm CMOS process and a prototype fabricated in the framework of Europractice.

Key words: neural networks, learning on silicon, hardware intelligence, CMOS analog circuits, low-power electronics.

1. Introduction

Artificial Neural Networks (ANN's) have been a subject of interests since a first mathematical model of a biological neuron appeared more than 60 years ago. Recently, most of works concerning ANN's lies in the area of mathematical considerations and they are mainly implemented in software [1,2]. However, hardware implementations of ANN's are becoming more and more popular recently and one can indicate many examples of successful applications of such ANN's in practice. One reason for the interests in hardware implemented ANN's is because they can be faster compared to the software ones. Another is that they can consume less power and exhibit a higher level of intelligence.

So far, the hardware implemented ANN's are in-advance-trained specific application circuits without a possibility to adaptive self-learning on silicon during operation. Thinking about implementing in a chip form a densely connected neural network capable of learning on silicon in the recall phase has become realistic only recently as a result of advances in CMOS processes [3–17].

The hardware implementation of intelligent self-learning ANN's is still a big challenge. A lot of conditions must be fulfilled and problems solved yet to achieve this goal. First, proper low power electronics must be worked out to realize basic operations required in the ANN's. Second, signals transmitted between neurons must be voltages, in order to avoid power losses in conductive paths. Since summation of currents is much easier to implement than summation of voltages, synapses should operate in a transconductance and neurons in a transresistance mode. Moreover, information about the synapse weight should be storied within a chip, and analog memories seem to be best for this purpose [7].

Various models of hardware neural networks have been developed and published [3–5]. We deal with a static model, where each synapse is accompanied by a local, short-term analog memory. This memory is needed to hold on silicon information about the synapse weight during the learning process. As the network training procedures we propose approaches based on Kohonen's [2] or Hebbian's methods [3], belonging to the group of unsupervised learning.

Recently, taking advantage of modern CMOS processes, several electronic circuits have been proposed to realize local analog memories [7,17], electronically controlled transconducatnce synapses [8], a transresistance neuron [9,10], Euclidean distance calculations [12,13], a conscience mechanism [14,15], a winning neuron detection [16]. The circuit of [9] can function as a power saving neuron with a step or signum activation function. In this paper, an improved version of the neuron circuit is proposed. The improvement relies on adding a second output at which the voltage is linearly dependent on the neuron input current. In this way, we obtain a neuron with two outputs, i.e. a step-function output and a linear one. The additional linear output is useful, among others, when the network performs classification tasks. Then, apart from classifying an input object to a given group (using the step-function output) we can asses a level of its belonging to this group (fuzzy logic approach). The linear output can also be used to detect the winning neuron in a learning process based on a WTA method (Winner Takes All). From our studies it results, however, that the WTA neuron detection can be carried out with a better effect when evaluating a similarity between a learning vector and a weight vector associated with a given output neuron. An Euclidean distanced metric can be applied for this purpose [12–16].

In this paper, simulations (SPICE) concerning the whole

*e-mail: woj@atr.bydgoszcz.pl

neuron and measurement results concerning its step output for a prototype made in a 0.35 μm CMOS process have been presented.

2. Proposed neuron circuit

Electrical scheme of the proposed neuron is shown in Fig. 1. The currents $I_1, I_2, ..., I_k$, are signals coming from synapses. They are summed at the neuron input node, and the resulting current I_{IN} is provided to a double-output transresistance activation circuit built of the transistors M01-M05. The pair M01/M02 creates a current mirror associated with one output, called "step output", while the other output, called "linear output" is based on the M01/M04 current mirror. M03 functions as a current source controlled by the voltage V_{th} and M05 as a quasi-linear resistor loading the M01/M04 mirror.

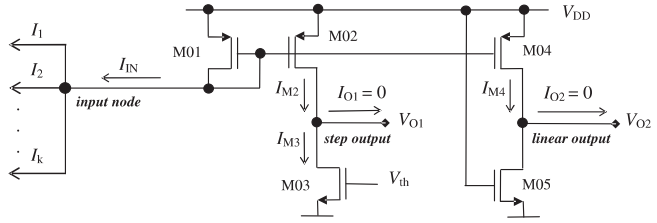

Fig. 1. Simple transresistance CMOS neuron with step and linear activation function outputs

At the neuron input, the current I_{IN} is a sum of currents delivered by synapses and can be expressed as:

$$I_{IN} = \sum_{i=1}^{k} I_i. \tag{1}$$

Notice, that I_{IN} can take only positive values, despite the fact that the summed input currents can be positive as well as negative. This is due to the transistor M01, whose gate-source voltage, V_{GS1}, can not be positive for input voltages being less than the supply voltage V_{DD}. The V_{GS1} voltage biases the transistor M02 and M04 forcing I_{IN} to be conveyed (current mirroring) to the output nodes. Thus, $I_{M2} \cong I_{IN}$ and $I_{M4} \cong I_{IN}$, provided that M02 and M04 operate in saturation.

Assuming that the currents I_{O1} and I_{O2} in Fig.1 can be neglected (neuron outputs loaded by a MOS transistor gate), we can write:

$$I_{M2} = I_{M3}. \tag{2}$$

Each of the transistors M02 and M03 can operate either in saturation or in triode region, depending on the control voltage V_{th} and the input current I_{IN}. Denote by I_{th} drain current of M03 operating in saturation. For the M03 operation in strong inversion, I_{th} is approximately described by:

$$I_{th} \cong K\left(V_{th} - V_p\right)^2, \tag{3}$$

where V_{th} is drain current, V_{th} is gate to source voltage, V_p is pinch-off voltage and K is a real-valued coefficient. If the input current I_{IN} is high and satisfies the inequality:

$$I_{IN} > I_{th}, \tag{4}$$

the transistor M03 operates in saturation, M02 in the triode region (current mirroring of the pair M01/M02 does not function) and the following relation is true:

$$I_{IN} > I_{M2} = I_{M3} = I_{th}. \tag{5}$$

This is an active state of the neuron. Its output voltage V_{O1} is then approximately equal to V_{DD}.

If the input current, I_{IN}, is less than the threshold value, I_{th}, i.e. when:

$$I_{IN} < I_{th}, \tag{6}$$

M02 is in saturation (the M01-M02 current mirror functions properly) and M03 is forced to operate in the triode region, which leads to:

$$I_{IN} = I_{M2} = I_{M3} < I_{th}. \tag{7}$$

Output voltage V_{O1} is then close to zero and this is an inactive state of the neuron.

From (6) and (7) it results that the neuron consumes no supply current (no supply power) if I_{IN} equals zero (important advantage). This takes place, for instance, when all synapses are inactive and provide no current to the neuron summation node, which can be expressed as:

$$I_1 = I_2 = \ldots = I_k = 0. \tag{8}$$

In addition, the presented neuron is well suited to low supply voltages and is able to carry out its tasks for the supply voltage V_{DD} being only slightly higher than the M01 transistor pinch-off voltage. This is desirable from the point of view of reducing power consumption associated with processing the input current I_{IN} for I_{IN} being different from zero (neuron in operation).

At the linear output in Fig. 1, the transistor M05 works in the triode region (non-saturated channel) and, as previously mentioned, plays a resistor role that loads the M04 transistor. An operation in this region takes place when drain-source voltage, V_{DS}, gate-source voltage, V_{GS}, and pinch-off voltage, V_p, of M05 fulfill the following inequality:

$$V_{DS} < V_{GS} - V_p. \tag{9}$$

The higher is the value on the right hand side of (9), compared to V_{DS}, the more linear is the transistor channel resistance. For this reason, gate of M05 is connected to the supply voltage V_{DD}. As a result, the V_{O2} voltage at the linear output in Fig. 1 is approximately proportional to the I_{IN} input current. A sufficiently large value of the channel resistance, required to operating with low currents (low power consumption) is obtained for a long and narrow channel of M05.

3. The neuron cooperation with synapses and local memories

Since signals transmitted between neurons should be voltages and output signals of the synapses should be currents, the neuron must operate in a transresistance mode and the synapses in a transconductance one. A common feature of ANN's implemented in both software and hardware is a partition into a

learning phase and recall phase, where the last starts after the first is finished. Such a partition results from the fact that learning procedure required by artificial networks lasts a very long time. If we want the ANN to be able to learn adaptively in operation during the recall phase, the learning time must be considerably shortened. To achieve this goal, we propose to use analog medium-term memories and locate them close to synapses as shown in Fig. 2. In these memories, information about the synapse weight is stored. The medium-term means that information should be held at least as long as it is required for one iteration of the learning procedure. This enables a quick, realized within a chip, variations of the synapse weights. After the learning process is finished, the weight information should be additionally recorded in an external digital memory. This memory is also needed to periodically refresh analog memories in the recall phase (by means of multiplexing techniques) until a next procedure of an adaptive self-learning stars.

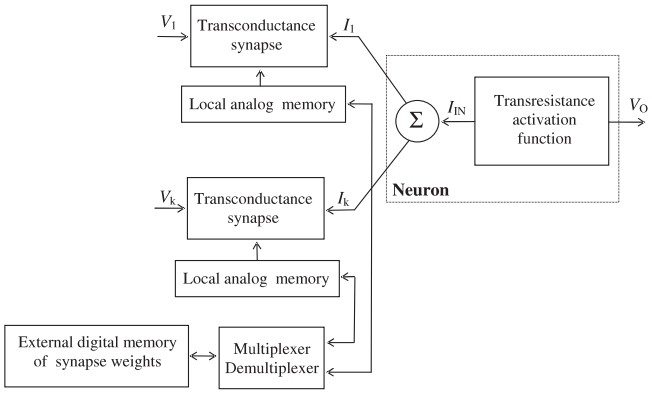

Fig. 2. Static neuron model including synapses coupled with local memories suitable for adaptive ANN's implemented in a chip form and trained on silicon

Scheme of a transconductance synapse [8] which is well suited to cooperate with the proposed neuron is presented in the next section. Like in the neuron case, its advantage is a zero power consumption when being inactive, i.e. when no voltage is delivered to its input, and a power economic operation when processing a different from zero voltage.

An analog medium-term memory of a capacitive type, suitable to be applied in our network is presented in [7,17]. Its advantage is an increased holding time, for a short acquisitions time, achieved due to applying a switched feedback around the holding capacitor. This allows us to obtain a relatively long holding time even for small capacitances of the holding capacitor. ANN's based on the scheme of Fig. 2 are attractive for networks with unsupervised learning on silicon using a "Winner takes all" mechanism.

4. Spice simulation studies

Results presented in this section concern a 0.35 μm CMOS process, for which the circuit was designed. The neuron properties as well as its cooperation with synapses were tested. Fig. 3 presents the tested circuit with only one synapse. Weights of synaptic connections are controlled by means of the voltages V_{C1} and V_{C2}, delivered from analog memories, in a differential way. The differential control is superior over a single voltage control in respect of damping common mode effects and improves the control precision. Other properties of the synapse have been described in [8]. Layout of the tested circuits was made using Cadence and simulations performed by means of HSPICE and PSPICE. Parasitic elements resulting from the layout have been taken into account in electrical schemes of the circuits examined. Pinch-off voltages of NMOS and PMOS transistors were equal to 0,4655 V and – 0,617 V, respectively. Transistor dimensions are shown in Tables 1 and 2.

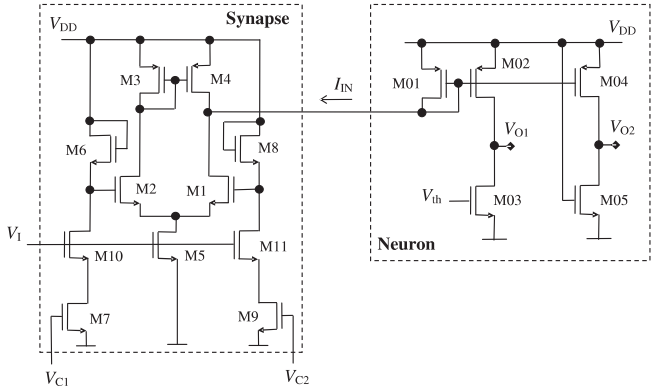

Fig. 3. Implemented in 0.35 μm CMOS process neuron with one synapse tested in the way of simulations

Table 1
Transistor dimensions of the simulated and experimentally tested neuron

Tran.	M01	M02	M03	M04	M05
W[μ]	2	2	2	2	1
L[μ]	2	2	2	2	140

In Fig. 4, principle of the neuron operation from the step output point of view is illustrated. The upper plot includes four curves presenting the neuron input current I_{IN} as a function of V_I voltage at the synapse input, for different values of the synapse weight, in the case with only one synapse like shown in Fig. 3.

Table 2
Transistor dimensions of the synapses used in simulations

Tran.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
W[μ]	18	18	18	18	2	7	2	7	2	30	30
L[μ]	20	20	0.8	0.8	200	4	200	4	200	0.8	0.8

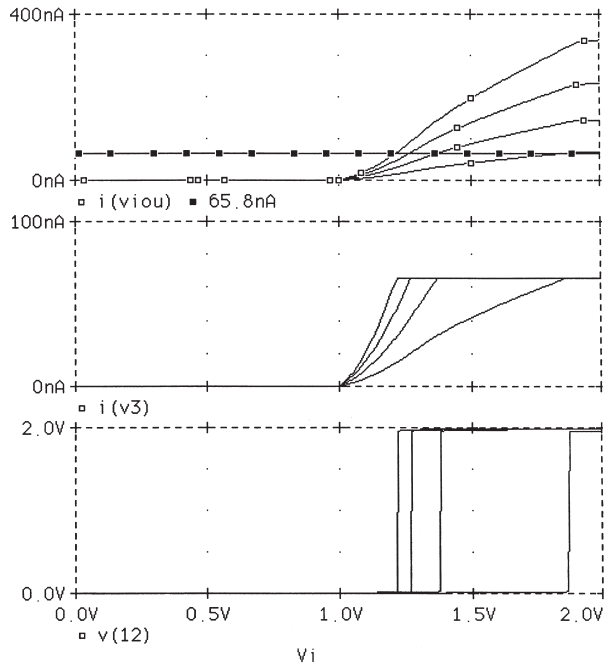

Fig. 4. The neuron basic signals for $V_{th} = 0.49$ V when driven by one synapse: a) input current I_{IN} for four synapse weights versus input voltage V_I of the synapse, b) drain current I_{M3} of the transistor M03 versus V_I, c) V_{O1} voltage at the step output versus V_I

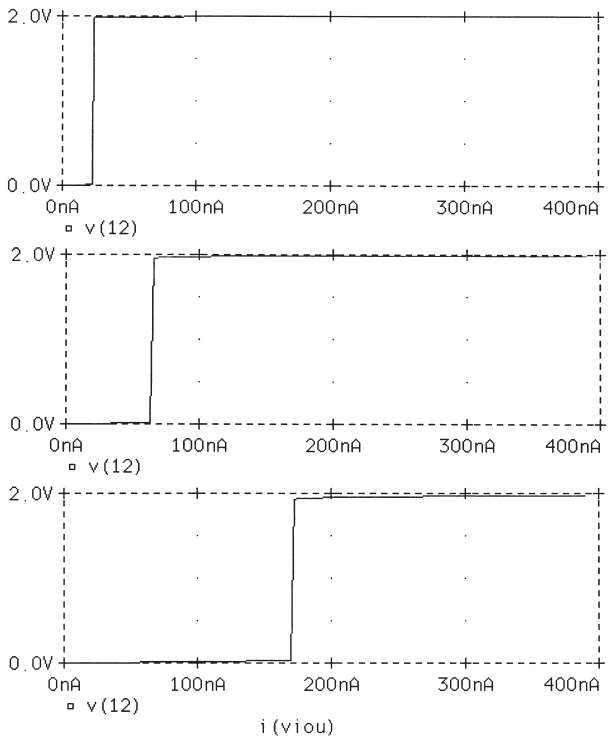

Fig. 5. Transfer characteristics concerning the step output of the transimpedance neuron showing the possibility of controlling the threshold current Ith by means of V_{th}: a) V_{O1} versus I_{IN} for $V_{th} = 0.48$ V (upper plot), b) V_{O1} versus I_{IN} for $V_{th} = 0.49$ V (middle plot), c) V_{O1} versus I_{IN} for $V_{th} = 0.505$ V (bottom plot)

For clarity reasons, a constant value of the neuron activation threshold $I_{th} = 65.8$ nA, corresponding to V_{th} = 0.49 V, is also marked on this plot (horizontal line). The middle plot presents the I_{M3} drain current of MO3 and the bottom output voltage V_{O1} of the neuron as functions of V_I. If I_{IN} crosses the I_{th} level (upper plot), I_{M3} gets into saturation (middle plot) and output voltage takes approximately the value $V_{O1} = 2$ V (bottom plot).

In Fig. 5, a possibility of controlling an activation threshold of the step output is demonstrated. The shown three curves correspond to a different value of V_{th}. For the upper trace we have $V_{th} = 0.48$ V, for the middle $V_{th} = 0.49$ V and for the bottom $V_{th} = 0.505$ V. Simulated properties of the neuron linear output are shown in Fig. 6.

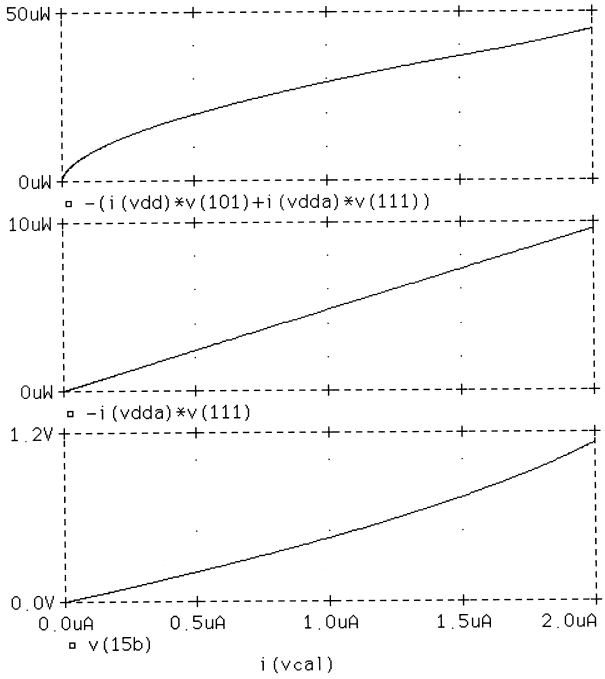

Fig. 6. Simulated DC properties at the linear output of the neuron driven by three synapses of the type shown in Fig. 3. a) power consumed by the whole circuit (neuron plus three synapses), b) power consumed by the neuron only, c) the linear output voltage, V_{O2}, of the neuron versus its input current I_{IN}

The upper plot presents total power consumed by the neuron and three synapses with which it cooperates as a function of the neuron input current I_{IN}. Voltages controlling the synapse weights are as follows: $V_{C1} = 2$ V and $V_{C2} = 1$ V for the first synapse, $V_{C1} = 1.9$ V and $V_{C2} = 1.7$ V for the second synapse and $V_{C1} = 1.9$ V and $V_{C2} = 1.7$ V for the last one. Input voltage, V_I, of each synapse is varied from zero to the supply voltage V_{DD}. As can be seen, the consumed power increases when V_I rises. In Fig. 6, the highest value of this power is less than 50 μW and coincides with $V_{in} = V_{DD} = 2$ V. The middle plot presents power consumed by the neuron only. This power does not exceed the 10 μW level and is rather low compared to that of the synapses. At the bottom plot we have the neuron transfer characteristic at the linear output (output voltage V_{O2} as a function of the input current I_{IN}). The obtained curve is almost linear, especially for low values of the input current I_{IN}.

Linearity of the transresistance transfer function for the V_{O2} output depends on the M01/M04 current-mirror properties as well as on a V_{DS}/I_D characteristic of the M05 transistor, which is used as a resistance loading the current mirror. As mentioned in Section 2, linearity of the M05 resistance is the higher, the lower is the V_{DS} voltage compared to the V_{GS}–V_p voltage difference. Lower values of V_{DS} correspond to lower values of the current flowing through the current mirror. As a consequence, linearity of the neuron transfer function improves when the neuron input current, I_{IN}, decreases. The achieved quasilinear transfer function shown in Fig. 6 (bottom plot) is in a good agreement with theoretical predictions. Concrete values of I_{IN}, for which the linearity is sufficiently good, depend on the M01, M02 and M03 transistor sizes. Higher values of W/L aspect ratios (width to length ratios of transistor channels) result in higher I_{IN} values for given V_{GS} and V_{DS} . This, of course, is at the cost of increasing the consumed power. Designing the transistor sizes, to ensure the neuron linear operation we have to take into account the number of synapses connected to its input and values of currents delivered by the synapses. Allowing higher values of the consumed power, one can obtain the neuron linear operation even for a large number of synapses.

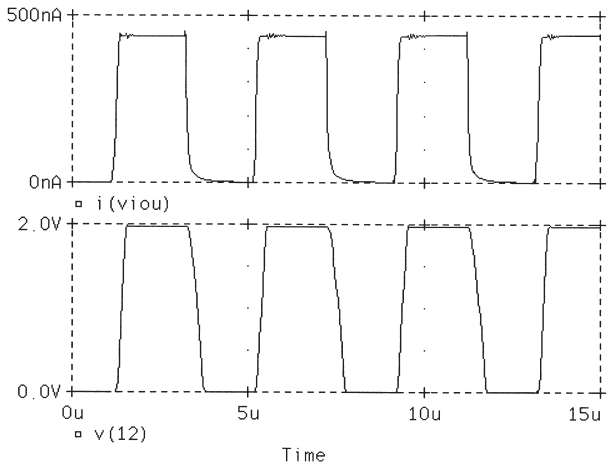

Fig. 7. Time response at the neuron step output: a) input current I_{IN} versus time (upper plot), b) output voltage V_{O1} versus time (bottom plot)

Speed and stability of the neuron operation, for the case with $V_{th} = 0.505$ V, is illustrated in Fig. 7. The top trace presents a current supplied to the neuron input and the bottom the neuron voltage response to this current. Some delay of the response can be observed when going from the higher to the lower voltage levels. Frequency of the signals is 250 kHz. Notice that no parasitic oscillations appear in the response voltage. This means a stable operation of the circuit. Its speed, however, is rather low. Fortunately, the low speed is not a great problem here because the synapses are not very fast as well. A slow operation of the synapses results, in general, from reducing their power consumption which is associated with reducing currents flowing through all transistors. In case of ANN's considered in this paper, a stable and power economic operation is more important that the speed.

5. Prototyping and experimental results

A first version of the neuron which included only one output (the step one) have already by prototyped and experimentally tested, while the present version with the additional linear output was, till now, only examined by means of simulations. Preparations for prototyping it are in progress. For this reason, we present measurement results concerning only the step output of the neuron.

As it is known, current measurements are in general more complex and less accurate than voltage ones. This is particularly true when the currents are very low, like it takes place in our case. That is way in the performed experiments, instead of the neuron input current I_{IN}, the synapse input voltage, V_I, was measured. As mentioned in the previous section, the prototyped circuit was implemented in a 0.35 CMOS technology by the firm NORDIC associated with AMS on the basis of our full custom design. We measured the fabricated chips, where among other circuits the neuron and synapses were included, for the supply voltage equal to $V_{DD} = 2.4$ V. Transistor dimensions of the fabricated neuron were exactly as shown in Table 1 and a little different in case of synapses (Table 3).

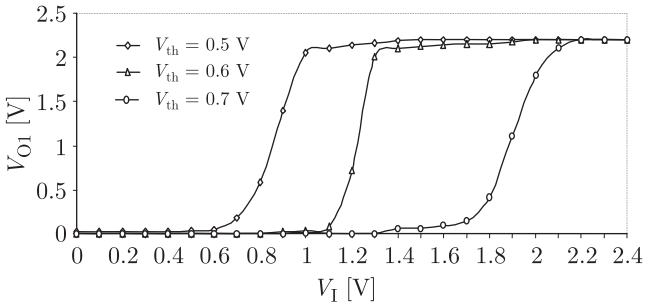

Fig. 8. Measured transfer characteristic (output voltage V_{O1} versus V_I voltage at the synapse input) of the circuit shown in Fig. 3, for V_{th} equal to 0.5 V, 0.6 V and 0.7 V

Table 3
Transistor dimensions of the synapses used in the prototype measurements

Tran.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
W[μ]	18	18	16	16	2	7	2	7	2	30	30
L[μ]	20	20	0.8	0.8	50	4	400	4	400	0.8	0.8

We have purposely presented the simulation results for different dimensions of some transistors (mainly M5, M9 and M10) to show that the most important transfer characteristic, i.e. the characteristic concerning the step output, has proved to be very similar in case of the performed simulations and measurements, despite these differences. This means a low sensitivity of the neuron characteristic to variations in some transistors dimensions. Effects of these variations are observed mainly in the area of power consumption and operation speed.

Experimentally determined transfer properties of the circuit of Fig. 3, i.e. relations between V_I and V_{O1}, are presented in Fig. 8. As can be seen, V_{O1} drops to zero for sufficiently low values of V_I. This, however, is true if V_{th} is not less than the M03 pinch-off voltage V_p (equal approximately to 0.47 V). If not, i.e. for $V_{th} < V_p$, the M03 channel resistance becomes extremely large causing an incorrect operation of the M01/M02 current mirror. Comparing Figs. 8 and 5, a big similarity can be noticed. In particular, a step character of the transfer characteristic and a possibility of controlling the neuron activation threshold is clearly seen.

Till now, the neuron with linear output have not been investigated experimentally because the authors were unable, for financial reasons, to realize another chip including the circuit. At present, such a possibility appeared in collaboration with University of Alberta in Edmonton, Canada. A layout prepared for a CMOS TSCM 0.18 μm process have already been made and sent for fabrication. Prototypes should be ready at the and of 2006. Experimental results of the full neuron will be presented in future publications of the authors.

6. Conclusions

A simple CMOS circuit has been proposed which can be used as a neuron in hardware implemented ANN's. Its usefulness to build a huge CMOS neural network results from the fact that it operates in a transconductance mode and is power economic. The transconductance operation means low losses in conductive path associated with using voltages as signals transmitted between neurons. It also means that a great number of synapses can be connected with one neuron because currents provided by the synapses can be easily summed in a single node at the neuron input. The power economic operation manifests itself in two ways. For different from zero input signals, a small amount of energy is consumed by the circuit. In the absence of input signals, no power is taken from supply sources. In contrast to the first version of the neuron published in [9], the presented version has two outputs. At one of them, voltage response to the input current is of a step-function type (step output) while at the other, voltage is linearly related to the input current. The step output enables an operation with binary output signals and the linear can be utilized, as an alternative to the Euclidean distance method [11–16], to point out the winner neuron in a WTA-based unsupervised learning on silicon. A scheme of cooperation between the neuron, synapses and analog memories, as a way of realizing the learning within a chip during the recall phase, has also been outlined. Properties of the neuron and its good cooperation with synapses have been investigated in details and positively verified by means of SPICE simulations. Experimental studies were also carried out but on a restricted scale because only the first version of the neuron have already been prototyped. The measurements were performed using a chip fabricated in a 0.35 μm CMOS process within the Europractice framework. Preparations for further prototypes are in progress.

References

[1] J. Żurada, *Introduction to Artificial Neural Systems*, West Publishing Company, USA, 1992.

[2] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlin, 1995.

[3] G. Cauwenberghs and M. Bayoumi, *Learning on Silicon, Adaptive VLSI Neural Systems*, Kluwer Academic Publishers, 1999.

[4] T. Roska and A. Rodrigues-Vazquez, *Towards the Visual Microprocessor*, John Willey & Sons, 2001.

[5] W. Maass and C. Bishop, *Pulsed Neural Networks*, Massachusetts Institute of Technology, MIT Press, 1999.

[6] K. Wawryn and B. Strzeszewski, "Low power VLSI neuron cells for artificial neural networks", *Proc. ISCAS* 3, 372–375 (1996).

[7] P. Grad, "Switched-feddback analog memories for CMOS neuroprocessing", Ph.D. Dissertation, University of Technology and Agriculture, Bydgoszcz, 2003, (in Polish).

[8] R. Wojtyna and T. Talaśka, "Improved power-saving synapse for hardware implemented ANN's", *Int. Conf. on Signals and Electronic Systems ICSES*, 27–30 (2004).

[9] R. Wojtyna and T. Talaśka "Simple low-power CMOS neuron to be used with transconductance synapses", *IEEE Workshop Signal Signal Processing*, 21–26 (2004).

[10] R. Wojtyna and T. Talaśka, "CMOS neuron with step and linear activation functions", *State Electronics Conference KKE*, 79–84 (2005).

[11] T. Talaśka, R. Wojtyna, and R. Długosz, "Hardware implemented neural network model with unsupervised learning on silicon", *Int. Workshop MIXDES'2005*, 133–136 (2005).

[12] R. Wojtyna, "Simple CMOS transconductance-mode differential squarer", *IEEE Workshop Signal Processing'2005*, Poznań, 171–177 (2005).

[13] R. Wojtyna, "Current-mode analog square rooter for hardware neuroprocessing", *IEEE Workshop Signal Processing'2006*, Poznań, 61–64 (2006).

[14] T. Talaśka, R. Wojtyna, R. Długosz, and K. Iniewski, "Implementation of the conscience mechanism for Kohonen's Neural Network in CMOS 0.18 μm technology", *International Conference Mixed Design of Integrated Circuits and Systems MIXDES'2006*, Gdynia, 319–315 (2006).

[15] T. Talaśka, R. Wojtyna, R. Długosz, K. Iniewski, and W. Pedrycz, "Analog-counter-based conscience mechanism in Kohonen's neural network implemented in CMOS 0.18 μm technology", *IEEE Workshop on Signal Processing Systems*, Banff, Canada, 420–425 (2006).

[16] R. Długosz, T. Talaśka, and R. Wojtyna, "New binary-tree-based winner-takes-all circuit for learning on silicon Kohonen's networks", *International Conference on Signals and Electronic Systems, ICSES 2006*, Łódź , 441–444 (2006).

[17] R. Wojtyna, "CMOS analog memory with increased storage time", *ICSES 2006*, Łódź , 437–440 (2006).