

# Analog Programmable Distance Calculation Circuit for Winner Takes All Neural Network Realized in the CMOS Technology

Tomasz Talaška, Marta Kolasa, Rafał Długosz, and Witold Pedrycz, *Fellow, IEEE*

**Abstract**—The paper presents a programmable analog current-mode circuit used to calculate the distance between two vectors of currents, following two distance measures. The Euclidean (L2) distance is commonly used. However, in many situations it can be replaced with the Manhattan (L1) one, which is computationally less intensive whose realization comes with less power dissipation and lower hardware complexity. The presented circuit can easily be reprogrammed to operate with one of these distances. The circuit is one of the components of an analog Winner Takes All neural network (NN) implemented in the CMOS (complementary metal oxide semiconductor) 0.18  $\mu\text{m}$  technology. The learning process of the realized NN has been successfully verified by the laboratory tests of the fabricated chip. The proposed distance calculation circuit (DCC) features a simple structure, which makes it suitable for networks with a relatively large number of neurons realized in hardware and operating in parallel. For example, the network with three inputs occupies a relatively small area of 3900  $\mu\text{m}^2$ . When operating in the L2 mode, the circuit dissipates 85  $\mu\text{W}$  of power from the 1.5 V voltage supply, at maximum data rate of 10 MHz. In the L1 mode, an average dissipated power is reduced to 55  $\mu\text{W}$  from 1.2 V voltage supply, while data rate is 12 MHz in this case. The given data rates are provided for the worst case scenario, where input currents differ by 1-2 % only. In this case the settling time of the comparators used in the DCC is quite long. However, that kind of situation is very rare in the overall learning process.

**Keywords**—Distance calculation circuit, Euclidean and Manhattan distance, parallel data processing, CMOS implementation, asynchronous circuits, self-organizing neural networks

## I. INTRODUCTION

**D**ISTANCE calculation circuits (DCCs) computing distances vectors of signals are used in numerous applications, mostly in pattern recognition and signal analysis [1], [2], [3], [4], [5], [6], [7] but also in artificial neural networks (ANNs) [8]. As ANNs, so far, have only rarely been implemented in hardware, the majority of the reported DCCs

T. Talaška and M. Kolasa are with the Faculty of Telecommunications, Computer Science and Electrical Engineering, UTP University of Science and Technology, ul. Kaliskiego 7, 85-796, Bydgoszcz, Poland, (e-mail: talaška@utp.edu.pl)

R. Długosz is with DELPHI Automotive Company, ul. Podgórci Tynieckie 2, 30-399, Kraków, Poland, with the Faculty of Telecommunications, Computer Science and Electrical Engineering, UTP University of Science and Technology and with the Institute of Microengineering, Swiss Federal Institute of Technology in Lausanne, Rue de la Maladière 71B, CH-2002, Neuchâtel, Switzerland (e-mail: rafal.dlugosz@gmail.com)

W. Pedrycz is with the Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB Canada & Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University Jeddah, 21589, Saudi Arabia and Systems Research Institute, Polish Academy of Sciences Warsaw, Poland (e-mail: wpedrycz@ualberta.ca).

were optimized for the applications of the first type. In general, when viewing the circuit from the formal point of view, in both cases the DCC performs a similar task. However, there are also certain differences which make the structure as well as the development process of such circuits substantially different in particular cases.

In this paper, we present a novel programmable DCC designed for the application in low power, low chip area, self-organizing neural networks (NNs) realized as application specific integrated circuits (ASICs) in the CMOS technology [9], [10], [11]. The proposed DCC has been used in a prototype analog Winner Takes All (WTA) NN realized in the TSMC (Taiwan Semiconductor Manufacturing Company) CMOS 0.18  $\mu\text{m}$  technology, but this circuit can also be used in self-organizing NNs trained with other algorithms. The realized WTA NN comprises also the following modules: a winner selection circuit (WSC) that determines which of the neurons becomes the winner for a given learning pattern  $X$ , a conscience mechanism that is used to eliminate or at least to reduce the problem of the, so called, dead neurons [9] and an adaptation mechanism (ADM) that is used to adjust the weights of the neurons, serving also as an analog memory for the calculated weights [10]. The NN is also equipped with some initialization mechanism that enables an initial polarization of the weights of the neurons.

In this study, we are motivated by the possibility of using such NNs in practical applications in various areas of engineering that embrace image processing [12], control and robotics [13], [14], electrical engineering [15] and health care systems [16]. All such applications require modifications and an optimization of existing learning algorithms to facilitate their hardware implementation [12], [17]. In such applications larger learning systems combining various learning techniques are often required that include semi-supervised and deep learning [18], [19].

ANNs realized as ASIC can be used in portable low power devices, for example, as components of nodes in Wireless Body Area Networks (WBANs) used in medical diagnostics. A demand for such systems is high [16], [20], [21]. Throughout a constant process of monitoring patients, the systems of this type would prove beneficial in numerous examples of health issues, including heart attacks or strokes, Alzheimer disease, Parkinson disease, prevention of sudden infant death syndrome, and others [16], [22], [23]. The currently used WBAN systems involve relatively simple sensors that usually perform only basic tasks, such as a collection of biomed-

ical data, analog-to-digital conversion (ADC), simple data preprocessing and conditioning (filtering) [24], [25], [26]. Finally, the collected data throughout the radio frequency (RF) communication block are being transmitted to a base station for a further processing and analysis [27]. One of the main problems encountered in such systems today is a substantial loss of energy (even up to 95 % of a total energy) during the RF transmission of data that significantly reduces the battery lifespan [28]. This is one of the barriers in the development of truly wearable medical systems that would be convenient for patients. The application of ANNs in such systems can bring about a real breakthrough. In the literature, one can find numerous examples of employing ANNs, realized in software, in the analysis of various biomedical signals e.g., Electrocardiographic (ECG) signals [29], [30], [31]. However, in WBANs, in which low power dissipation and low chip area of the wireless nodes are paramount features, ANNs implemented on standard programmable devices are not particularly useful.

In our former investigations we carried out a comparison of self-organizing maps (SOM) implemented as ASICs with similar systems realized with the use of Field Programmable Gate Arrays (FPGAs) [32]. These studies showed, that ASIC realizations can offer even two orders of magnitude better performance in terms of the Figure-of-Merit (FOM) defined as attainable data rate over dissipated power. This is due to the possibility of a very good match of the architecture of the circuit and the number of the transistors used to the function performed by the circuit. In this context, the development of small and ultra-low power ANNs can be a major step forward towards the realization of “intelligent” and thus more autonomous sensors that will be able to realize data analysis and classification of biomedical data. Such sensors contact a base station only in urgent situations, which, in turn, reduces the amount of energy consumed during the RF data transmission.

The paper is structured as follows: the next section presents a state-of-the-art background necessary to put the proposed circuit in a proper perspective. Since the overall WTA NN has been realized as the analog circuit, our interest was in analog DCCs [33]. However, to make this study exhaustive, we present also some comparative digital circuits that can be potentially interesting for fully digital implementations of NNs. While presenting the state-of-the-art study, it was appropriate to distinguish DCCs considered as separate blocks as well as larger systems – data classifiers – that use DCCs as one of the components.

In Section III, we present the design process of the proposed circuit. First, we offer a brief overview of other components used in the realized WTA NN. This is important, as the DCC contains the components whose role is to enable an interaction with other blocks of the NN. Next, an underlying idea of the circuit is described. In the sequel, we discuss the negative aspects that need to be taken into account when designing such circuits. Finally, we present the CMOS implementation of the DCC.

In Section IV we report on an experimental assessment of the realized circuit. Both transistor level simulation results and the laboratory measurements of the fabricated circuit

are presented. In Section V, we discuss the obtained results with the emphasis on the accuracy of the realized circuit and the comparison of the parameters with other state-of-the-art solutions. Finally, the conclusions are drawn in Section VI.

## II. STATE-OF-THE-ART BACKGROUND

The state-of-the-art overview has been arranged into two parts. First, we focus on simpler circuits that are counterparts to the proposed DCC considered as a separate functional module. The reported circuits often offer different features and functions that impact the complexity, parameters and performance of the circuit. This means that direct comparison is not that straightforward. For this reason, in the second part of the presentation, we provide an overview of larger systems – data classifiers – that use DCCs as one of their components. To facilitate this comparison, we consider classifiers that use blocks with functionality similar to that of the WTA NN discussed here. However, as the reported classifiers do not offer the adaptation function, therefore to facilitate a direct comparison we additionally estimate parameters of our system, as if it does not have the adaptation block.

### A. Analog vs. digital approach

One of the important issues is whether the circuit should be realized using digital, analog or a mixed approach. The answer will depend on the application for which the system is being designed. In larger systems with predominance of digital blocks, in which the NN would be one of the stages in the overall signal processing chain, it is more convenient to use digital NNs. In this case, digital NNs can process directly signals provided by other blocks, without the use of intermediate digital-to-analog conversion.

On the other hand, the advantage of the analog approach is visible when the NN directly processes analog input signals. In self-organizing NNs weights of particular neurons are hidden signals, not directly used in further data analysis. Usually the only signals seen at the outputs of the NN are 1-bit signals, one per neuron, that bear the information that a given neuron becomes the winner for a given input pattern  $X$ . In analog NNs all these signals can be calculated without the necessity of using analog-to-digital converters (ADCs) at the inputs of the NN. As a result, the overall system can occupy smaller area and dissipate relatively less power. In digital NNs, on the other hand, all calculations are performed on multi-bit signals, whose processing requires more complex circuits. For example, the summation and subtraction operations in analog current-mode circuits are performed in a junction, while in digital approach they require multi-bit adders and subtractors. Nevertheless, analog DCCs can be influenced by various negative physical phenomena, discussed in Section III-C, and therefore they need a careful design and optimization. All this shows that there are various trade-offs that have to be taken into account while selecting one of the described techniques.

### B. Comparison of DCCs viewed as separate blocks

In literature, one can find both the voltage mode (V-Mode) [7], [8], [34] and the current mode (I-mode) analog DCCs

[1], [2], [3], [4], [5], [6], [35]. Circuits that work in the I-mode offer simpler structure, while the achieved precision is usually comparable. One of the main advantages of the I-mode approach is that it facilitates the summation (SUM) and subtraction (SUB) operations of the signals. These operations, fundamental in DCC, are in the I-mode performed simply in junction. For this reason DCCs operating fully in the V-mode are not common, as the SUM and SUB operations are more complex in this case and require capacitors and active elements [36]. One of the possibilities is the mixed mode that enables working with the input voltage signals, while the SUM/SUB operations are realized in the I-mode [34]. The calculated signal can be then converted back to the voltage signal [1]. In [2], [6], [8], on the other hand, the input signals are voltages, while the output signal is current.

Although many DCCs have been reported in the literature, they do not offer sufficient functionality to make them suitable for the application in ANNs. For this reason, we proposed a new analog DCC, also working in the I-mode, that delivers several additional functions that are not required and thus not available in circuits used in pattern recognition. On the other hand, some blocks which are important only in pattern recognition applications have been omitted in our circuit. The only circuit that is a direct counterpart of the proposed DCC is the circuit reported in [8].

The DCCs that are used in pattern recognition usually aim to calculate with high degree of precision the real distance between two patterns, which is the final outcome in this case. The real distance is the distance calculated in accordance with the Euclidean (L2) measure that requires both squaring and rooting blocks [1], [4], [6]. In case of DCCs used in ANNs, on the other hand, the calculation of exact values of the distances between a given input pattern  $X$  and weight vectors  $W$  of particular neurons is less important than a precise determination of the neuron that is located the closest to this pattern. As discussed further in the paper, this allows omitting the rooting block and in many cases also the squaring circuits. Moreover, as in the NN realized by us the DCC cooperates closely with the adaptation mechanism (ADM), it has to provide some additional signals used in the ADM block [10]. Such functions are not required in circuits used in the pattern recognition.

The proposed DCC offers an additional feature, which is not common in other solutions of this type. The circuit can easily be programmed (using 1 bit only) to operate with two different distances, i.e. the Euclidean (L2) and the Manhattan (L1) one. For the comparison, the circuit proposed in [8], also designed for the applications in ANN, enables calculation of these two distances but is not reconfigurable. In this solution a given distance has to be determined during the circuit design through proper sizing of selected transistors. However, the ability to operate with both these measures in a single chip is a useful feature. System level simulations we have carried out for a model of a self-organized NN show that using simpler (L1) measure does not distort the learning process. This is discussed in more detail in Section III-B.

One of the components of the proposed DCC, the absolute (ABS) function block, can be classified as a I-mode rectifier

which, in this case, is based on a I-mode comparator and a simple switching network. Many circuits of this type have been reported over the past years [37], [38]. We have proposed our own solution, as the existing rectifiers are not of much use in the realized WTA NN. A typical rectifier calculates a rectified value of a single input signal. In contrast, in the proposed solution the circuit calculates a rectified value of a difference between two signals. Additionally, the comparator used in this case provides information which of these signals is greater. This 1-bit signal that controls selected switches in the ADM block allows to simplify the structure of this block [10].

In 2012 Fernandez *et al.* in [35] reported a field programmable analog array (FPAA) that contains an Euclidean distance operator as well as an I-mode multiplier that can be compared with particular components of our DCC. This circuit performs the operations  $I_{\text{out}} = \sqrt{I_X^2 + I_Y^2}$  that are insufficient in self organizing NNs. To compare, our circuit additionally calculates the absolute value of an  $x_i - w_{li}$  term, in which  $x_i$  is an  $i^{\text{th}}$  input signal, while  $w_{li}$  is an  $i^{\text{th}}$  weight of an  $l^{\text{th}}$  neuron in the NN. It also determines which of the  $x$  and the  $w$  signals is greater that is signaled by a 1-bit digital signal that is used to control the adaptation block. It also calculates the  $\eta \cdot |x_i - w_{li}|$  term that is directly used by the ADM block to adjust the weights of the winning neuron ( $\eta$  is the learning rate). The Euclidean operator proposed in [35], for two input signals, consumes the power (static power without the input signals) of  $733 \mu\text{W}$ , while the one-quadrant current-mode multiplier the power of  $275 \mu\text{W}$  for the measured bandwidth of 20 MSamples/s. For the comparison, our DCC for three input signals consumes the power of  $85 \mu\text{W}$ , including the power consumed by the squarer, at data rate of 10 MSamples/s.

To facilitate a comparison between particular DCC solutions, a Figure-of-Merit (FOM) can be defined as data rate ( $f_S$  [MHz]) over power dissipation ( $P$  [mW]) in relation to the number of the calculation channels (NC):

$$\text{FOM}_1 = \frac{\text{NC} \cdot f_S}{P} \quad (1)$$

The comparison of particular solutions, based on Eq. 1 is provided in Table I.

### C. Comparison of data classifiers based on WTA structure

A fully digital circuit, such as our NN, designed in the  $0.18 \mu\text{m}$  process is reported in [39]. This circuit is a complex system, able to calculate either the Hamming or the Manhattan distance (a counterpart to our DCC) between a given input pattern and thirty two words. The words, which are the 64-bit signals, are counterparts to neuron weight vectors in our WTA NN. The circuit described in [39] determines which of these words is the most similar to a given input pattern. This operation is performed in a block that is a counterpart to our WSC. In the Manhattan mode (L1-mode), each 64-bits input signal is divided into nine components, which are equivalent signal to the input  $x_i$  signals in our WTA NN. Particular components are in this case 7-bits signals, encoded in the thermometer code, i.e. 3-bits in the binary code. If the circuit described in [39] operates in the L1-mode, it can be viewed

as a digital WTA NN (or classifier) with 32 neurons, without the adaptation block. For the comparison, our proposed circuit allows for the adaptation of the neuron weights.

A disadvantage of this solution results from the calculations performed with data encoded in the thermometer code that substantially limits the calculation precision. In case if the 64-bits words would be divided into two signals, data resolution would be limited to 5 bits only (32 bits in the thermometer code). The overall system reported in [39] dissipates 51.3 mW, at data rate of 6.3 MHz. For the comparison, our WTA NN together with the adaptation module and the conscience mechanism, for 12 calculation channels (3 inputs and 4 neurons), dissipates power of 700  $\mu$ W at data rate of 2 MHz. In our solution the speed is strongly reduced due to the adaptation phase that is not present in the circuit reported in [39]. Without the ADM block and the conscience mechanism data rate of our circuit exceeds 7 MHz. If our WTA NN would be extended to 32 neurons and 9 inputs (288 channels) with effective resolution of 8 bits, it would dissipate power of 17 mW with the ADM block and about 4 mW without this component.

To enable a quantitative comparison between particular classifiers, let us define Figure-of-Merit (FOM) as data rate over the power dissipation normalized to one pattern (or one neuron):

$$\text{FOM}_2 = \frac{\text{NU} \cdot f_s}{P} \quad (2)$$

where NU is the number of units i.e. the number of neurons in the WTA NN or the number of words in other classifiers used in pattern recognition.

In case of data classifiers, we can use also the  $\text{FOM}_1$  measure. Both these measures tell us something different about the circuit. To explain this in more detail, let us consider two example classifiers: one with two neurons and hundred inputs and the second one with hundred neurons and two inputs. The number of the calculation channels is the same in both cases, but the winner selection circuit becomes much more complex in the second case, as it has to select the minimum signal among hundred signals. In case of the first classifier, the  $\text{FOM}_1$  provides more objective results, while in the second classifier the  $\text{FOM}_2$  is more suitable.

In case of our circuit considered without the adaptation block, the  $\text{FOM}_2$  for one neuron equals 40.87 [1/nJ]. For the comparison, the  $\text{FOM}_2$  in case of the circuit reported in [39] equals 3.9 [1/nJ]. The silicon area of our NN for 12 calculation channels equals 0.07 mm<sup>2</sup>. For 288 channels considered without the CONS mechanism, the ADM blocks and thus without most of the components of the DCC block, the chip area would equal 0.40 mm<sup>2</sup>, which is about 27 % less than in case of the circuit reported in [39] (0.55 mm<sup>2</sup>).

An interesting fully digital DCC realized in the 65 nm process, operating in the L1-mode, has been recently reported in [40]. The circuit calculates distances between an input pattern composed of 16 entries, each of them being an 8 or 16-bits signal, and 128 counterpart words. This circuit offers a similar functionality (DCC and WSC blocks) as the circuit proposed in [39], but enables processing signals with much larger resolutions, which is an advantage. One of the

components of this circuit is the Manhattan DCC, whose output signal (for the signal resolution of 8 bits) can vary in-between 0 (for an exact match between two patterns) and  $16 \cdot 256 = 4096$  in the worst case, i.e. is represented on 12 bits. The main component of the WSC block used in [40] (denoted as WTA) is a time divider, which is a kind of a counter. In the worst case scenario it requires 4096 clock cycles (at 220 MHz clock) to determine the winning signal. In this case the quantization time would equal 18.6  $\mu$ s (53 kpatterns/s). Such a scenario is rather seldom, while an average reported quantization time equals 0.796  $\mu$ s (1.25 Mpatterns/s). The circuit dissipates 3.56 mW of power for  $128 \cdot 16 = 2048$  calculation channels. For the comparison, if our WTA NN would be redesigned to such number of channels, it would dissipate power of 28 mW at data rate of 7 Mpatterns/s (without the ADM mechanism). The  $\text{FOM}_2$  in the circuit described in [40] equals 1.9 [1/nJ] and 44.96 [1/nJ], for the worst case and the average case scenarios, respectively.

A substantial difference between the worst and the average cases raises the question about how to set up the input pattern/s rate. The issue will mostly depend on the application. In our WTA NN we always consider the worst case scenario, especially at the beginning of the learning process, when distances between particular patterns  $X$  and neuron weight's vectors  $W$  are large.

Another fully digital circuit, designed in the CMOS 0.18  $\mu$ m technology which, to some extent, is comparable to our solution has been reported in [41]. An exact comparison is not possible, as this circuit is based on the calculation of the Hamming distance that requires a different approach. It allows to calculate in parallel distances between a given input pattern and 64 reference words (256 bits signals). At voltage supply of 1.8 V the circuit consumes power of 34 mW, for the search time equal to 255 ns ( $f_s \approx 4$  MHz). The FOM calculated using Eq. 2 equals in this case 7.53 [1/nJ].

### III. THE PROPOSED PROGRAMMABLE DCC

The proposed circuit closely cooperates with other blocks in the realized WTA NN. Some features have been added just because of this interaction. This is why it seems appropriate to begin this Section with a brief overview of the system. Subsequently, we present an idea of the circuit and its transistor level implementation. As the proposed circuit is the analog solution, it is important to discuss various issues that appear during the realization of such circuits and that may impact the parameters of the circuit.

#### A. An overview of the realized analog WTA NN

A block diagram of the overall WTA NN is shown in Fig. 1 (top). The NN is composed of several blocks that perform different operations in the system but simultaneously very closely cooperate with each other. Particular neurons are composed of equal channels that operate fully in parallel. A single channel contains two blocks. The first one in the chain is the DCC that calculates the following signals:  $d_{li} = |x_i - w_{li}|^2$  (or  $d_{li} = |x_i - w_{li}|$  in the L1 mode),  $d_{li}^n = \eta \cdot |x_i - w_{li}|$ ,  $s_{li} = \text{sign}(x_i - w_{li})$  and  $\overline{s_{li}}$ . If a given neuron becomes the

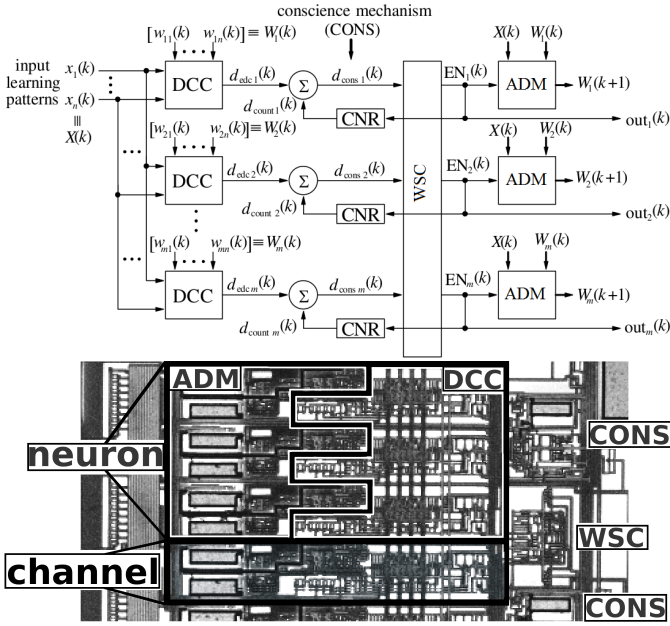


Fig. 1. WTA neural network: (top) a general block diagram [9], (bottom) microphotograph of a prototype NN with 12 calculation channels (4 neurons with 3 inputs) realized in the CMOS 0.18  $\mu\text{m}$  technology

winner, the  $d_{li}^n$  signal is used by its adaptation mechanism (ADM), which is a second component of the channel, to update its weight  $w_{li}$ . The 1-bit signals  $s_{li}$  and  $\bar{s}_{li}$  throughout a set of switches control the ADM block that performs the following operation:

$$w_{li}(k+1) = w_{li}(k) + (s_{li} - \bar{s}_{li}) \cdot d_{li}^n(k) \quad (3)$$

The usage of the  $s_{li}$  and  $\bar{s}_{li}$  signals enabled a substantial simplification of the structure of the ADM block [10].

The  $d_{li}$  signals coming from particular channels in a given neuron are summed in a junction. The resultant current,  $D_{dccl} = \sum_i d_{li}$ , is a real distance between this neuron and a given learning pattern  $X$ . Each neuron additionally contains a CONS block (conscience mechanism), which is composed of a counter (CNR) that counts the wins of this neuron, as well as a converter that converts the value stored in the counter to the output current,  $D_{consl}$ . This signal is then summed with a given  $D_{dccl}$  signal. As a result, the CONS mechanism artificially increases the real distance if a given neuron tries to dominate in the learning process. The resultant signal,  $D_l = D_{dccl} + D_{consl}$ , is then provided to the WSC, whose role is to determine which neuron is the winner.

The number of the channels in a single neuron is equal to the number of the NN inputs,  $n$ , so the total number of channels in the overall NN equals  $n \cdot m$ , where  $m$  is the number of neurons (NN outputs). The microphotograph shown in Fig. 1 (bottom) illustrates a placement of particular components in the chip realized in the TSMC CMOS 0.18  $\mu\text{m}$  technology.

### B. The proposed circuit

Block diagram of the proposed DCC is shown in Fig. 2, while particular components of this circuit are illustrated in

Fig. 3. First, a given input learning pattern  $X$  coming from a learning set is provided to the inputs of the chip and directly used as the inputs of the DCC in each neuron in the NN. The core module of this circuit is the ABS block, shown in Fig. 3 (a), that calculates the absolute value of the  $(x_i - w_{li})$  term for each neuron weight,  $w_{li}$ . This block is controlled by a simple I-mode comparator, shown in Fig. 3 (a) that compares a given input signal,  $x_i$ , with a corresponding weight,  $w_{li}$ . The  $x_i$  and the  $w_{li}$  signals are currents. A digital output signal,  $s_{li}$ , of the comparator controls the switches in the ABS block in such a way that the larger of these currents is directed as a positive signal to the path that is composed of only PMOS-type current mirror (CM) (M11-M12). The smaller signal additionally flows throughout the inverting NMOS-type CM (M13-M14). As a result, the output signals of particular ABS blocks are always equal to  $|x_i - w_{li}|$ .

The last feature is quite useful, as it allows us to realize in a simple way two distance measures in a single chip. These two measures i.e. the Euclidean (L2) and the Manhattan (L1) ones can be described, as follows:

$$D_{L2} = A \cdot \sqrt{\sum_{i=1}^n (x_i - w_{li})^2} \quad (4)$$

$$D_{L1} = A \cdot \sum_{i=1}^n |x_i - w_{li}| \quad (5)$$

In the expressions above,  $A$  is a constant that is determined by transistor sizing. Obviously, as for each,  $l^{\text{th}}$ , neuron in the NN  $D_{L2,l} \geq 0$ , therefore for any two neurons,  $\alpha$  and  $\beta$ , if  $D_{L2,\alpha} < D_{L2,\beta}$  then  $D_{L2,\alpha}^2 < D_{L2,\beta}^2$ . For this reason, both the L2 measure and a resultant L2<sup>2</sup> one lead to the same results in the process of identifying the winning neuron (viz. ranking the signals), and therefore the rooting operation can be omitted. The resultant L2<sup>2</sup> measure can be described as follows:

$$D_{L2^2} = A \cdot \sum_{i=1}^n (x_i - w_{li})^2 = A \cdot \sum_{i=1}^n |x_i - w_{li}|^2 \quad (6)$$

In the case of the L2<sup>2</sup> measure, the rooting operation that substantially contributes to the complexity of the overall DCC is not required. Note that in (5) and (6) the same term,  $|x_i - w_{li}|$ , is used. This allows for an easy transition between both distances. In the proposed circuit the transition requires switching only four switches per each channel (in Fig. 2 only two are shown for simplicity). As a result, in the ‘‘L1’’ mode the output signal of the ABS block bypasses the squaring circuit (SQR), shown in Fig. 3 (b). All switches are controlled by a single bit only ( $L1 = \bar{L}2$ ), while the mode can be changed even during the learning process of the NN. This comes as a significant advantage over the circuit proposed in [8], in which a given measure had to be determined in advance during the design process of the chip.

The ability to work with two different measures is an important feature. The system level investigations show that it is beneficial to use the L2 or the L2<sup>2</sup> measure, for example, at the beginning of the learning process, while after the first,

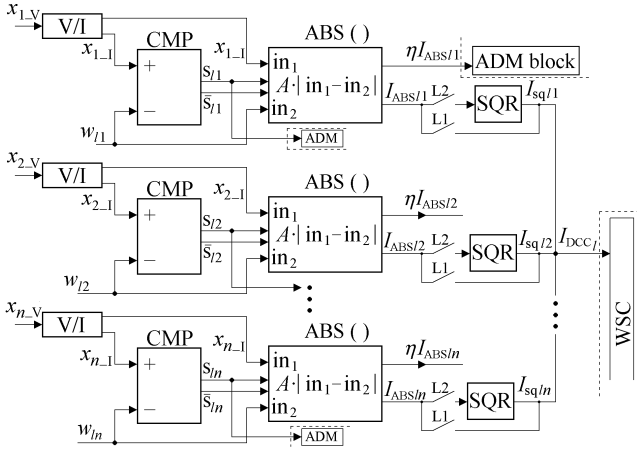


Fig. 2. Block diagram of the proposed analog, current-mode, distance calculation circuit (DCC).

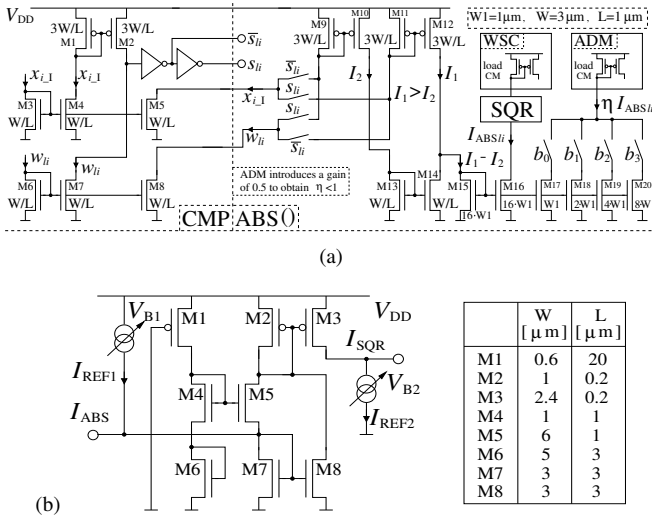


Fig. 3. Main components of the proposed DCC: (a) current mode comparator (CMP) and absolute function block (ABS), (b) one-quadrant currents squaring circuit (SQR) [6] with additional currents  $I_{REF1}$  and  $I_{REF2}$ .

rough, phase the learning process can be continued with the L1 measure. This speeds up the overall learning process and, additionally in the hardware realization, reduces the power dissipation. This issue has been studied in [32]. Selected system level simulation results are presented in Fig. 4. The figure presents quantization error after completing the learning process. The value  $Q_{err} = 16.84e - 3$  is optimal in this case and means that all neurons became representatives of different data classes and are located in the centers of these classes. Particular learning patterns  $X$  are in this case uniformly spread around these centers. The presented results show that in case of the NNs with the number of neurons up to 400 the L1 measure enables reaching the same or better results. For larger NNs the use of the L2 measure in many cases leads to better results. The simulations of the learning process show that the determinative phase is the very beginning of the learning process. In this phase it is beneficial to use the L2 mode, while later the NN can be trained with the L1 distance measure.

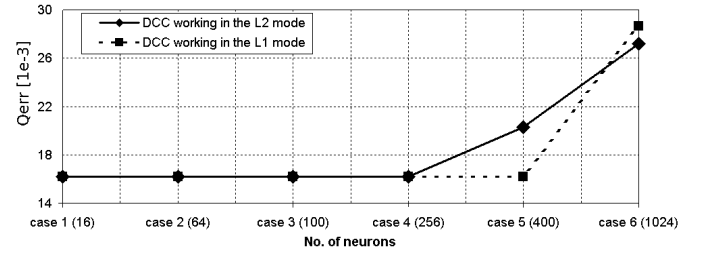


Fig. 4. Quantization error after completing the learning phase of the NN with different numbers of neurons. Comparative results for two distance measures.

Another advantage of having the positive value at the output of the ABS block is the possibility of using a simple one-quadrant SQR circuit to obtain the  $L2^2$  distance measure. In the proposed solution, we have used an SQR circuit proposed in [6], shown in Fig. 3 (b). The circuit, originally implemented in the CMOS  $2 \mu\text{m}$  technology, has been redesigned and optimized in the CMOS  $0.18 \mu\text{m}$  process. To make it more flexible and to improve its performance (the shape of the output parabola), we have added two current sources,  $I_{REF1}$  and  $I_{REF2}$ , controlled by biasing voltages,  $V_{B1}$  and  $V_{B2}$ . The  $I_{REF1}$  DC current enables locating the input signal in an optimal region, in which the values of the output signal are close to theoretical values. The  $I_{REF2}$  current, on the other hand, enables adjusting the level of the signal provided to the Winner Selection Circuit (WSC) that in the implemented WTA NN is the next block in the signal processing path. This feature is very useful if the number of the inputs of the NN is one of the parameters. It is worth to note that the minimal values of the inputs and the weights currents are usually larger than  $1 \mu\text{A}$  to avoid the situation in which transistors in the SQR block operate in the underthreshold region. In this case, the circuit is very slow. The performance of the SQR circuit is presented in next section.

The  $I_{ABS}$  currents, coming from particular ABS blocks in a given neuron, are summed in a junction. The resultant sum is proportional to a given distance measure. In Fig. 2 the distance is represented by the current  $I_{DCCi}$  defined as:

$$I_{DCCi} = \sum_{i=1}^n I_{sqli}(k) \quad (7)$$

or

$$I_{DCCi} = \begin{cases} A \cdot \sum_{i=1}^n |x_i(k) - w_{li}(k)|^2, & \text{for L2} \\ A \cdot \sum_{i=1}^n |x_i(k) - w_{li}(k)|, & \text{for L1} \end{cases} \quad (8)$$

where:  $w_{li}(k)$  is the  $i$ -th weight of an  $l$ -th neuron in the  $k$ -th learning cycle.

The proposed DCC has several outputs, as shown in Figs. 2 and 3 (a). One of them is the  $I_{DCCi}$  current provided to the WSC block, as described above. Additionally, particular  $I_{ABSi}$  currents multiplied by a learning rate  $\eta$ , which in this case is controlled by four 1-bits signals ( $b_0, \dots, b_3$ ), are provided to corresponding ADM blocks along with particular

digital 1-bit  $s_{li}$  signals. Multiplication by  $\eta$  is performed in a digital-to-analog converter (DAC), which is a multi-output current mirror (CM) with binary weighted output transistors, controlled by particular bits of the  $\eta$  parameter (transistors M17-M20 in Fig. 3 (a)). The values of  $\eta$  vary in the range in-between 0 and 15/16 with a step of 1/16 ( $\eta = b_0/16 + b_1/8 + b_2/4 + b_3/2$ ).

One of the advantages of the proposed circuit is the parallel operation of particular ABS blocks. For the NN with  $n$  inputs, each neuron contains  $n$  ABS blocks. A total number of these blocks in the overall NN equals  $m \cdot n$ , where  $m$  is the number of neurons. The proposed DCC operates without using a controlling clock that substantially simplifies the structure of this circuit and reduces the calculation time.

### C. Implementation issues in the proposed DCC

The problem important for large NNs containing hundreds neurons is the chip area. Each neuron contains its own DCC, while the number of the ABS blocks in each DCC is equal to the number of the inputs of the NN. For this reason, the structure of the DCC has to be relatively simple to keep the area small and the power dissipation of the overall NN low. On the other hand, the simplification of the circuit should not affect negatively the learning process of the NN. During the design and optimization process of the DCC, different negative effects have to be taken into account. They include, for example, the mismatch effect, the body effect and the channel length modulation (CLM) effect.

1) *An influence of the body effect on the circuit precision:* In several reported cases [1], [36], [42] the body effect has been considered to be the factor that reduces the circuit precision, with providing the methods that enable reducing its impact on the behavior of the circuit. In [1], for example, all PMOS transistors with the source terminals not connected to  $V_{DD}$  voltage have been located on separate n-wells coupled with these terminals, to ensure that the source and bulk (substrate) voltage ( $V_{SB}$ ) equals 0. In such a configuration the voltage across the reversed-biased well-to-substrate capacitance depends upon the value of the gate to source ( $V_{GS}$ ) voltage of the transistor. As a result, as this voltage is varying over time, the well-to-substrate capacitance is being dynamically recharged that slows down the circuit [1].

In contrast to the solution proposed in [1], to avoid the body effect, we have designed the overall DCC in such a way that sources of all transistors in the NMOS and PMOS CMs are connected to corresponding supply voltages, and thus their  $V_{SB} = 0$ . To make it possible we have used simple CMs instead of the cascaded ones. An additional positive effect of this is a relatively small value of the headroom voltage that enables the circuit to work with smaller supply voltages. A disadvantage of using simple CMs is small output impedance that enlarges the influence of the CLM effect. Fortunately, in the proposed solution this influence is relatively small, as the loads of particular CMs are input transistors of subsequent mirrors which are diode-connected. Let us consider, for example, the mirror composed of the M15 and M17–M20 transistors in the ABS block, shown in Fig. 3 (a).

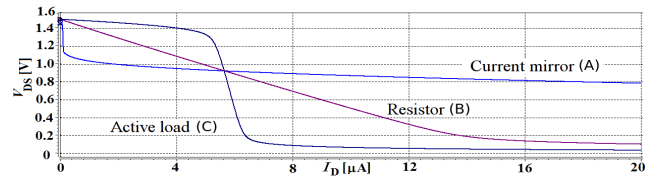


Fig. 5.  $V_{DS}$  voltage across DAC output transistors as a function of input current to the DAC (see Fig. 3 (a)), plotted for various typical loads.

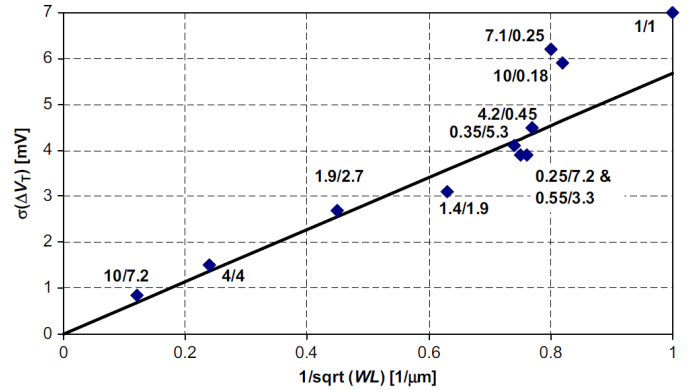


Fig. 6. An example illustration of a dependency between transistor sizes of the CMOS transistor and the  $\sigma\Delta V_{TH}$  parameter [44].

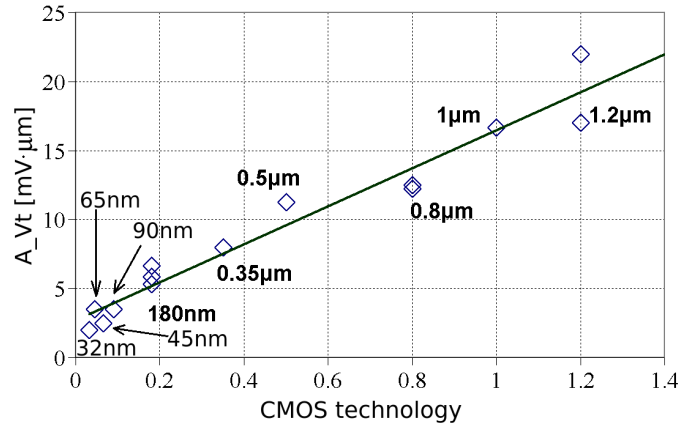


Fig. 7. A review of the threshold voltage mismatch in different CMOS technologies – a comparison made on the basis of the Pelgrom plots.

As the current  $\eta \cdot I_{ABSli}$  increases, the  $V_{GS}$  voltage of the loading PMOS CM, located in the subsequent ADM block (schematically shown in Fig. 3 (a)), increases as well, resulting in the output  $V_{DS}$  voltage varying only moderately, as shown in Fig. 5 (curve A). For the comparison, we present also the  $V_{DS}$  voltage for the resistive (B) and the active loads (C). As in case of the curve (A) the most stable values of the  $V_{DS}$  voltage are for  $I_D > 1$  to  $2 \mu A$ , the designed WTA NN usually operates in this range, with some exceptions.

2) *Impact of the mismatch effect:* Another problem, common to current mode circuits, is the mismatch effect that modifies the gain of the CM. We have studied this problem in more detail in the context of the ADM block described in [10], but as that circuit uses similar CMs, the conclusions

are applicable also in case of the proposed DCC. One of the common ways to present dependency between the mismatch effect and transistor sizes for a given CMOS technology is a Pelgrom plot [43]. An example diagram of this type is shown in Fig. 6. In this case standard deviation of a given parameter ( $\sigma\Delta V_{TH}$  in this example) is plotted versus  $1/\sqrt{W \cdot L}$  [44], [45], [46], [47], [48], where  $W \cdot L$  is the area of the transistor gate:

$$\sigma\Delta V_{TH}[\text{mV}] = f(1/\sqrt{W \cdot L}) [1/\mu\text{m}], \quad (9)$$

Waveforms visualized on Pelgrom plots are linear functions with a constant slope denoted as  $A_{VT}$ . One can notice that the mismatch depends on transistor sizes. The larger transistors are used, the smaller is the  $1/\sqrt{W \cdot L}$  factor and, in turn, the smaller is the mismatch. In each CMOS technology there are some minimum allowed sizes ( $W/L$ ) of transistors. In the TSMC CMOS 0.18  $\mu\text{m}$  process they equal 600/180 [nm]. Such transistors should not be used in current-mode analog circuits, as the gain error of the CM could become in this case as high as 100 % [10]. For this reason we oversized the transistors to suppress this error. When increasing the sizes of the transistors, it is necessary to bear in mind the values of the currents that flow through the CM. For a given value of the currents, if we increase transistor sizes, we improve the mismatch properties that reduces the gain error of the CM, but we also decrease the gate-to-source voltage ( $V_{GS}$ ) of the transistors. If the  $V_{GS}$  voltage decreases, the gain error of the CM increases [10], so we have two contradictory phenomena. As a result, for each current value optimum sizes should be found. In case of the proposed DCC working with currents at the level of 2 – 10  $\mu\text{A}$  the optimum sizes ( $W/L$ ) of the NMOS transistors equal 3/1 – 1/1  $\mu\text{m}$ , while of the PMOS transistors equal 9/1 – 3/1  $\mu\text{m}$ . For such sizes of the transistors and given values of the signals, a theoretical value of the gain error of a single CM reaches a local minimum that does not exceed 2 % [10].

One of the optimization techniques we used in case of the proposed DCC relies on minimizing the number of CMs on the path in-between the inputs and the output of the DCC, as theoretically the error accumulates along the path. The usage of the comparator and the switches allowed us to reduce the number of CMs on a single path (excluding the SQR circuit) to maximum 3. For the comparison, in the circuit proposed in [1] the number of CMs on the corresponding path equals 4 or 5. Some error is introduced by the SQR circuit, as shown in Fig. 10, but this error occurs only when the circuit operates in the  $L^2$  mode.

What is interesting is how the circuit parameters depend on the technology. Fig. 7 presents a comparative study between different CMOS technologies ranging from 32 nm to 1.2  $\mu\text{m}$ . On vertical axis are the values of the  $\sigma\Delta V_{TH}$  parameter for transistor gate areas normalized to 1  $\mu\text{m}^2$  i.e. directly the values of the  $A_{VT}$  parameter. The proposed circuit has been realized in the CMOS 0.18  $\mu\text{m}$  process, as this technology provides a reasonable price to parameters ratio. If the project would be redesigned in the never process, the chip area could be reduced without compromising on the precision. For example, comparing the 65 nm and the 0.18  $\mu\text{m}$  technologies

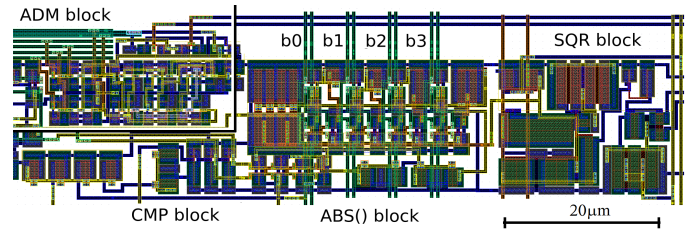


Fig. 8. Layout of a single channel of the proposed DCC. The area of this block equals 1280  $\mu\text{m}^2$ . The area of the overall DCC composed of three equal channels equals 3840  $\mu\text{m}^2$

one can notice that the values of  $A_{VT}$  equal 3.5 and 5.8, respectively. This means that in the 65 nm process a given value of  $\sigma\Delta V_{TH}$  will be achieved for the  $1/\sqrt{W \cdot L}$  factor 5.8/3.5 times larger than in the 0.18  $\mu\text{m}$  technology, i.e. for the area of the transistors  $(3.5/5.8)^2 = 0.36$  times smaller.

#### D. Realization of the proposed DCC in the CMOS technology

The proposed DCC has been implemented in the CMOS 0.18  $\mu\text{m}$  technology, as a component of a fully analog WTA NN with three inputs, containing four neurons i.e., twelve equal signal processing channels. Each channel represents a single  $x_i - w_{l,i}$  pair. The number of the neurons in this prototype NN is small, but as each of its building blocks has been designed by us from the ground up, the overall system, as well as particular components had to be carefully verified before a larger system could be realized. The outputs of particular channels are shorted in a junction. This approach enables an easy realization of a modular NN with programmed number of inputs, in which particular channels could be shared between different neurons.

The layout of a portion of a single channel that contains the realized DCC is shown in Fig. 8. The chip area of the DCC equals 1280  $\mu\text{m}^2$ . The remaining area is occupied by the ADM block. The b0 – b3 digital signals in the ABS block control the value of the learning rate  $\eta$ , as described in details in Section III-B. We comment on these results in Section V.

## IV. EXPERIMENTAL VERIFICATION OF THE PROPOSED CIRCUIT

In this section, we present selected postlayout simulation, as well as the measurement results that illustrate the behaviour and parameters of the realized circuit. Due to the large number of signals that had to be measured or introduced to the chip, only selected internal signals could be connected to external pads. Some signals were not connected in order to avoid their distortion by the measurements. In this situation, simulations enable observing those signals which are not available at the outputs of the chip.

In the case of the measurements, the input signals were voltages converted on resistors outside the chip to currents introduced then to the chip. Inside the chip the currents throughout a current mirror were provided to DCCs of particular neurons. The NN output signals were also currents, but to measure them we used 10 k $\Omega$  resistors of class 0.1 % connected in series with the output pads. Thus the signals have



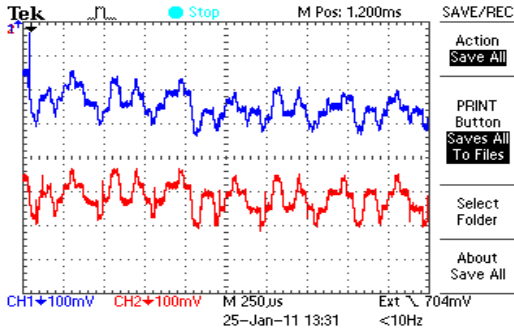


Fig. 9. Measurement results illustrating output currents from two selected DCCs of the WTA NN implemented in the CMOS 0.18  $\mu\text{m}$  technology working in the  $L^2$  mode.

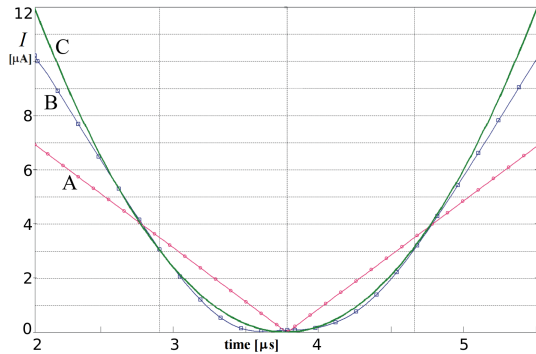


Fig. 10. Operation of the squaring circuit: (A) an input current, (B) the resultant output current, (C) an ideal output signal.

been measured as voltages on the oscilloscopes, as shown in Fig. IV-B (the  $L^2$  case). The presented results were typically observed during the measurements of 15 samples of the chip.

#### A. Transistor level simulation results

Simulations enable the observation of the details of the behavior of particular components of the NN seen as separate blocks. One of the blocks, that has been verified in this way, is the SQR circuit. Its performance is illustrated in Fig. 10. The resultant curve differs slightly from the theoretical waveform, but these differences are seen mainly for large input signals. From the point of view of the classification problem far more important are the results obtained for small input signals, because they occur in neurons that are located in the proximity of the corresponding learning patterns. It is also worth noting, that Fig. 10 does not show the static input-output characteristic. We applied a triangular signal at the input of the circuit to observe its dynamic properties. A certain asymmetry visible in the vicinity of the zero value results from the resultant transient state.

To illustrate the performance of the overall DCC, selected postlayout simulation results are presented in Fig. 11. In this case the simulations enable a direct observation of the output signals of the comparators, which is not possible in laboratory tests. The  $I_x$  and  $I_w$  signals used in the tests have been selected in such a way to observe properties of the circuit in various scenarios, including the worst case scenario, in which the

TABLE I  
PERFORMANCE COMPARISON BETWEEN REPORTED DCCS

Ref. (Techn.)	No. inputs	A [ $\text{mm}^2$ ] (1 input)	$f_S$ [MHz]	P [ $\text{mW}$ ] (1 input)	Error [%]	FOM1 [ $1/\text{nJ}$ ]
[1]	4x5	10.4 (0.52)	1	14.95 (0.75)	1	1.33
[8]	2	0.0061 (0.00305)	ND	0.2 (0.1)	5–15	–
[7]	16x16	1.2 (0.0047)	0.33	0.7 (0.00273)	ND	120
[36]	4x6	0.624 (0.026)	1	ND (–)	0.4	–
[35]	2	ND (–)	20	0.733 (0.367)	3	27.3
This work	3	0.0038 (0.00128)	12 ( $L^1$ ) 10 ( $L^2$ )	0.055 ( $L^1$ ) 0.085 ( $L^2$ )	1 ( $L^1$ ) 4 ( $L^2$ )	656 ( $L^1$ ) 353 ( $L^2$ )

input signals of particular ABS blocks are almost equal. In the period from 0 to 2.7  $\mu\text{s}$  they often differ by 1 % only (for particular ABS blocks). Additionally, their values have been distributed over relatively large range in-between 2 and 7  $\mu\text{A}$ . The signals in this period are staircase waveforms. This allows for dynamic properties of the circuit to be observed. Additionally, after settling the outputs of the comparators such signals allow to determine the static input-output properties of the ABS block, as well as of the overall DCC. For the input signals with almost equal values the settling time of the comparators is the longest. In practice, such situation is rather rare, as the signals in a real learning process usually differ by much more than 1 %. Average power dissipation equals 55  $\mu\text{W}$ . This means that energy consumption per a single distance calculation in the worst case scenario (settling time  $\approx 85$  ns) equals 4.5 pJ.

#### B. Measurement results

Selected measurement results are shown in Figs. 9 and 12 for the DCC working in the  $L^2$  mode. In the presented case the NN performs a typical learning process, and therefore it is not possible to precisely control the waveforms of the weights, as in the simulations of a separate DCC, shown in Fig. 11. In the  $L^2$  mode the power dissipation was equal to 85  $\mu\text{W}$ . This value has been determined on the basis of the postlayout simulations (as in the  $L^1$  case) for a separate DCC.

### V. DISCUSSION OF THE OBTAINED RESULTS

#### A. Assessment of the circuit precision

An important issue is to estimate how accurate the circuit calculates the output signal for given input signals. It can be done by the assessment of an error at the output of the DCC that can be calculated as follows:

$$\text{ERROR} = \left| \frac{I_{\text{out}_t} - I_{\text{out}_r}}{I_{\text{max}}} \right| \cdot 100 \text{ [%]}, \quad (10)$$

where  $I_{\text{out}_t}$  is a theoretical value of the output current, while  $I_{\text{out}_r}$  is its real value – measured in this case.

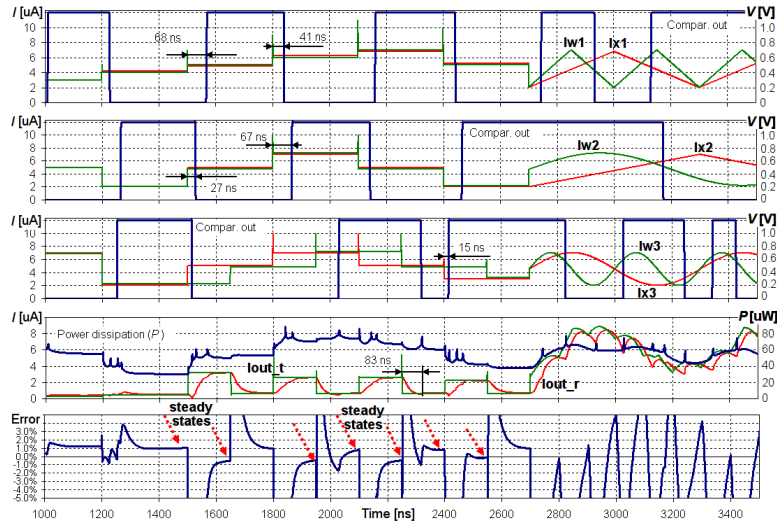


Fig. 11. Postlayout simulations of the DCC for input currents in the range of 2–7  $\mu\text{A}$  for the L1 mode: (a, b, c) input ( $I_x$ ), weight ( $I_w$ ) and the output signals from the comparators, (d) the theoretical ( $I_{\text{out}_t}$ ) and real ( $I_{\text{out}_r}$ ) DCC output currents and the power dissipation, (e) error:  $I_{\text{out}_t} - I_{\text{out}_r}$  in reference to the maximum value of the  $I_{\text{out}}$  current. The output current is normalized i.e. is divided by the number of channels.

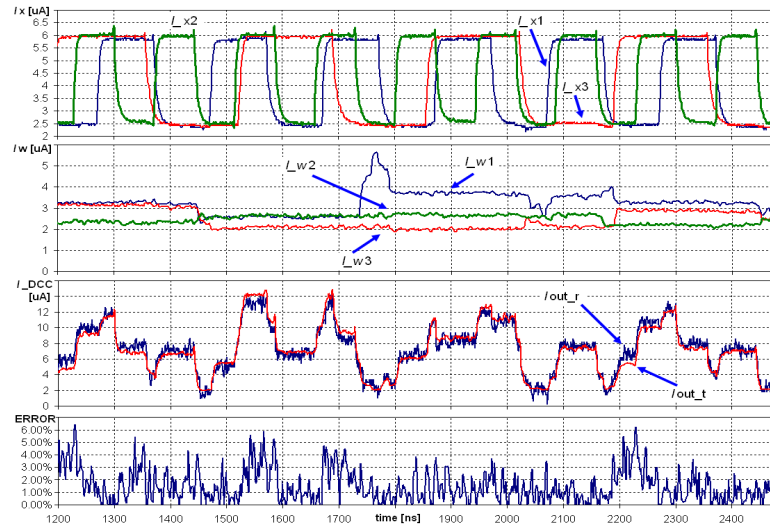


Fig. 12. Measurement results for a single neuron operating in the  $L2^2$  mode: (a) input training signals ( $I_x$ ), (b) corresponding weights ( $I_w$ ), (c) theoretical ( $I_{\text{out}_t}$ ) and real ( $I_{\text{out}_r}$ ) DCC output currents, (d) the error defined as in Eq. 10.

The static error can be accurately determined only for constant input signals (steady state) visible, for example, in the period in between 0 and 2.7  $\mu\text{s}$  in Fig. 11. In this case, for the circuit working in the L1 mode the steady state error does not exceed 1 %. In the measurements, shown in Fig. 12 for the  $L2^2$  mode, the steady state error is usually at the level of 2 %, sometimes exceeding 3–4 %. The main source of the increased error in this case is the SQR block (see Fig. 10), but this block can be replaced by a more precise one if needed. Some portion of the error results also from the initial V-I and the final I-V conversion performed outside the chip, as well as from the noise visible in the signal. The presented results show that the DCC in the L1 mode is more accurate, which is an advantage, as NN mostly operates in this mode. For the comparison, in the circuit proposed in [8], also designed for

the application in ANN, the error often exceeds the level of 5–10 %, for the output currents one order of magnitude larger.

A general observation confirms that the values of the signals affect both the accuracy and speed of the circuit. We observe that the delay is, more or less, linearly proportional to an average level of the input currents. We performed simulations of the circuit working in the L1 mode for small currents in the range in between 100 nA and 600 nA. For such values the steady state error was usually larger than 2.5 %, while a delay introduced by the DCC was equal to 400 ns for an average value of the input currents of 600 nA. For the currents at the level of 200 nA the delay was 650 ns. However, the power dissipation is smaller in this case, as the speed is lower, and, therefore, the energy per a single calculation cycle decreases only moderately, while the errors are larger. This confirms that it is not reasonable to use very small input signals.

TABLE II  
PERFORMANCE COMPARISON BETWEEN REPORTED DATA CLASSIFIERS AND THE REALIZED WTA NN

Ref. (Techn.)	Realization	No. words	Vector length	Distance measure	A [mm <sup>2</sup> ] (per 1 unit*)	f <sub>s</sub> [MHz]	P [mW] (per 1 channel)	FOM1 [1/nJ]	FOM2 [1/nJ]
[39] (0.18μm)	digital	32	9	Hamming / L1	0.55 (1.91e-3)	6.3	51.3 (0.178)	35	3.92
[40] (65nm)	digital	128	16	L1	1.48 (0.72e-3)	0.053# 1.25	3.56 (0.00174)	30.48# 719	1.9# 44.94
[41] (0.18μm)	digital	64	4	Hamming	3.08 (12.03e-3)	3.92# 20	36.5 (0.143)	27.5# 140.3	6.88# 35
This work (0.18μm)	analog (I-mode)	4	3	L1 / L2 <sup>2</sup>	0.02* (1.67e-3)	7*·#	0.22* (0.0183)	368#	40.87#

\* Data for the circuit without the adaptation block and the conscience mechanism

# Data for the worst case scenario

### B. Comparative study with other state-of-the-art solutions

As described in Section II, a direct quantitative comparison of the proposed DCC with other circuits of this type is not straightforward, as particular solutions were designed for different applications and therefore offer different functions, and have different numbers of inputs. To facilitate the comparison we calculate the Figure-of-Merit (FOM) using Eq. 1. The results for selected circuits are presented in Table I. In several reported circuits, certain data are not present, which makes the calculation of the FOM impossible for these cases.

The proposed circuit has been designed in newer technology than other solutions, but to minimize the mismatch effect, we had to strongly oversize transistors used in the circuit. That did not allow us to fully utilize the advantages of the newer technology used in this case. The circuit reported in [7] features the FOM comparable to our DCC. However, this circuit does not offer the functions available in our solution, which are required in ANNs, as discussed in Section II.

In our opinion, it is more effective to compare larger systems of similar functionality that use DCC components. In Table II our prototype WTA NN is compared with the recently reported data classifiers which, to a certain degree, are similar to our NN. Such systems are composed of many DCCs working in parallel, and the WSC block that determines which of the DCCs provides the smallest signal. The main difference relies on the lack of the adaptation block in the solutions described in [39], [40], [41]. All these solutions are digital circuits that theoretically make it possible to use transistors with minimal sizes for a given technology. On the other hand, in the case of digital realization of the given function, a larger number of transistors will be required. As a result, the chip area in our solution per one calculation unit is the smallest among the circuits designed in the CMOS 0.18 μm technology. The ‘unit’ is an equivalent to the calculation channel in our NN (shown in Fig. 1 bottom). The number of units can be calculated as number of words (equivalent to neurons in our NN) times the vector length (No. of the inputs of the NN).

In comparison with the circuit reported in [40], designed in the CMOS 65 nm technology, the chip area of our NN (without the ADM and the CONS blocks) per one unit is only two times larger. Nevertheless, considering the results shown in Fig. 7, one can notice that  $A_{VT}$  in the CMOS 65 nm technology (3.5 [mV·μm] [49] (page 515)) is 1.65 times smaller than in the

0.18 μm process (5.8 [mV·μm]). As a result, if our NN would be ported to 65 nm technology, the sizes of the DCC could be reduced theoretically by 64 %, without compromising on precision. This problem has been discussed in more detail in Section III-C2.

It is worth noting that the presented circuit is a proof of concept, designed to verify the general idea of the proposed solution, as well as its performance. For this reason the chip area was not the subject of a strong optimization (some blank areas are visible). Taking it into account a further reduction of the area even by 25 % is possible. Our circuit was designed to enable a choice between two distance measures, which is a feature not available in other solutions. However, the simulation results show that the SQR block that occupies 40 % of the area currently occupied by the DCC can be eliminated in most cases, as discussed in Section III-B.

Comparing the FOM (Eq. 2), it can be observed that the proposed WTA NN offers better parameters than other classifiers. As regards our circuit, we always consider the worst case scenario, in which the settling time of the comparators is the longest. This is mandatory, if viewed from the point of view of the learning algorithm, in which the winning neuron has to be indicated properly. For the comparison, in [40], [41] an average search time can be taken into account that is acceptable in case of the application of the circuit in pattern recognition. For this reason, in the case of the circuit reported in [40] the FOM for an average search time is larger than in our DCC, but for the worst case scenario it is even one order of magnitude smaller.

The realized prototype chip contains a small number of neurons (12 channels in 4 neurons). We did not design larger NN, as all its components were designed from the ground up, and thus the focus was on the optimization of their parameters rather than on increasing the number of neurons. Nevertheless, the size of the NN can easily be increased by duplicating the channels. In many practical applications the number of neurons can be limited to a relatively small value. For example, in [50], [51] it has been shown that to classify the ECG complexes the number of neurons can be limited to 30–130 only.

## VI. CONCLUSIONS

The paper presents a novel distance calculation circuit (DCC) suitable for low power self-organizing neural networks

implemented at the transistor level in the CMOS technology. The circuit is programmable, i.e. it can operate with two commonly used distance measures, the Manhattan and the Euclidean ones. This is one of the advantages. The circuit features a simple structure, resulting in a relatively small silicon area. A single channel composed of the comparator, the abs() function and the squaring blocks occupies the area of  $1280 \mu\text{m}^2$ . The comparison of the proposed DCC as well as the overall NN with other counterpart circuits of this type shows that the power dissipation related to attainable data rate is very low as well. These parameters strongly depend on the values of input signals as well as on the distance measure used in a given situation.

In comparison with other DCCs, the precision of the proposed circuit is relatively high. When reporting measurements, the precision was slightly lower due, mostly, to V-I and I-V conversions performed outside the chip using resistors. Another source of the nonidealities is the SQR block that in the following prototype chip will be either improved or eliminated. The last conclusion results from the simulations carried out by means of the software model of the NN that show that the L1 measure in most cases allows to obtain comparable results of the learning process and, at the same time, requires substantially simpler hardware.

#### REFERENCES

- [1] Bin-Da Liu, Chuen-Yau Chen, Ju-Ying Tsao, "A modular current-mode classifier circuit for template matching application", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 47, No. 2, 2000, pp.145–151
- [2] S. Vlassis, G. Fikos, S. Siskos, "A floating gate CMOS Euclidean distance calculator and its application to hand-written digit recognition", *International Conference on Image Processing*, Vol. 3, 2001, pp: 350-353
- [3] C. Popa, "CMOS current-mode Euclidean distance circuit using floating-gate MOS transistors", *International Conference on Microelectronics*, Vol. 2, 2004, pp.585–588
- [4] C. Popa, "Current-mode Euclidean distance circuit independent on technological parameters", *International Semiconductor Conference*, Vol. 2, pp: 459-462, 2005
- [5] O. Landolt, E. Vittoz, P. Heim, "CMOS Selfbiased Euclidean distance computing circuit with high dynamic range", *Electronics Letters*, Vol. 28, No. 4, pp: 352-354, 1992
- [6] K. Bult, H. Wallinga, "A Class of Analog CMOS Circuits Based on the Square-Law Characteristic of an MOS Transistors in Saturation", *IEEE Journal of Solid-State Circuits*, Vol. SC-22, No.3, 1987, pp. 357–365
- [7] G. Cauwenberghs, V. Pedroni, "A Low-Power CMOS Analog Vector Quantizer", *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 8, August 1997, pp.1278–1283
- [8] A. Gopalan, A. H. Titus, "A New Wide Range Euclidean Distance Circuit for Neural Network Hardware Implementations", *IEEE Transactions on Neural Networks*, Vol. 14, No. 5, Sep. 2003
- [9] R. Długosz, T. Talaška, W. Pedrycz, R. Wojtyna "Realization of the Conscience Mechanism in CMOS Implementation of Winner-Takes-All Self-Organizing Neural Networks", *IEEE Transactions on Neural Networks*, Vol. 21, Iss. 6, pp.961–971, June 2010
- [10] R. Długosz, T. Talaška, W. Pedrycz, "Current-Mode Analog Adaptive Mechanism for Ultra-Low Power Neural Networks", *IEEE Trans. on Circuits and Systems-II: Express Briefs*, Vol.58, Iss.1, pp.31–35, Jan.2011
- [11] R. Długosz, M. Kolasa, W. Pedrycz, M. Szulc, "Parallel Programmable Asynchronous Neighborhood Mechanism for Kohonen SOM Implemented in CMOS Technology", *IEEE Transactions on Neural Networks*, Vol. 22, Iss. 12, pp. 2091–2104, Dec. 2011
- [12] A. wietlicka, "Trained stochastic model of biological neural network used in image processing task", *Applied Mathematics and Computation*, DOI: 10.1016/j.amc.2014.12.082, 2015
- [13] P. Skruch, "An educational tool for teaching vehicle electronic system architecture", *International Journal of Electrical Engineering Education*, Vol. 48, No. 2, 2011, pp. 174–183
- [14] P. Skruch, "Feedback stabilization of a class of nonlinear second-order systems", *Nonlinear Dynamics*, Vol. 59, No. 4, 2010, pp. 681–692
- [15] D. Borkowski, A. Wetula, A. Bien, "Contactless Measurement of Substation Busbars Voltages and Waveforms Reconstruction Using Electric Field Sensors and Artificial Neural Network", *IEEE Transactions on Smart Grid* Vol.PP, Issue: 99, DOI: 10.1109/TSG.2014.2363294, 2015
- [16] B. Latré, B. Braem, I. Moerman, C. Blondia, P. Demeester, "A survey on wireless body area networks", *Wireless Networks*, Vol. 17, No. 1, 2011, pp. 1–18
- [17] K. Gugaa, A. wietlicka, M. Burdajewicz, A. Rybarczyk, "Random number generation system improving simulations of stochastic models of neural cells", *Computing*, Vol. 95, No. 1, 2013, pp. 259-275
- [18] Jin Xu, Haibo He, Hong Man, "DCPE Co-Training for Classification", *Neurocomputing* Vol. 86, 2012, pp. 75–85
- [19] J. Weston, F. Ratle, and R. Collobert, "Deep learning via semi-supervised embedding", in *Proceedings of the 25 International Conference on Machine Learning*, (ICML), New York, NY, USA, 2008, pp.1168–1175
- [20] F.C. Morabito, D. Labate, A. Bramanti, F. La Foresta, G. Morabito, I. Palamara, H.H. Szu, "Enhanced Compressibility of EEG Signal in Alzheimer's Disease Patients Sensors", *IEEE Sensors Journal*, Vol.13, Iss. 9, Sept 2013, pp. 3255–3262
- [21] Zhilin Zhang, Tzyy-Ping Jung, Scott Makeig, Bhaskar D Rao, "Compressed Sensing of EEG for Wireless Telemonitoring with Low Energy Consumption and Inexpensive Hardware", *IEEE Transactions on Biomedical Engineering*, Vol. 60, Iss. 1, 2013, pp.221–224
- [22] Hong Ying, M. Schlosser, A. Schnitzer, T. Schafer, M.E. Schlafke, S. Leonhardt, M. Schiek, "Distributed Intelligent Sensor Network for the Rehabilitation of Parkinson's Patients", *IEEE Transactions on Information Technology in Biomedicine*, 2011, pp:268-276
- [23] M. Avvenuti, C. Baker, J. Light, D. Tulpan, A. Vecchio, "Non-intrusive Patient Monitoring of Alzheimer's Disease Subjects Using Wireless Sensor Networks", *World Congress on Privacy, Security, Trust and the Management of e-Business (CONGRESS '09)*, 2009, pp:161-165
- [24] K. Lorincz, D. J. Malan, T. R.F. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, and M. Welsh, "Sensor networks for emergency response: challenges and opportunities", *IEEE Pervasive Computing*, Vol. 3, No. 4, 2004, pp:16-23
- [25] Lu Shilong, Xi Huang, Li Cui, Ze Zhao, Li Dong, "Design and implementation of an ASIC-based sensor device for WSN applications", *IEEE Transactions on Consumer Electronics*, Vol. 55, No. 4, 2009, pp:1959-1967
- [26] S. Mandal, L. Turicchia, R. Sarpeshkar, "A Low-Power, Battery-Free Tag for Body Sensor Networks", *IEEE Pervasive Computing*, 2010, pp:71-77
- [27] Cao Huasong, V. Leung, C. Chow, H. Chan, "em Enabling technologies for wireless body area networks: A survey and outlook", *IEEE Communications Magazine*, Vol. 47, No. 12, 2009, pp:84-93
- [28] A. Bereketli and O. Akan, "Communication coverage in wireless passive sensor networks", *IEEE Communications Letters*, Vol. 13, no. 2, 2009, pp. 133–135
- [29] C.R. Leite, D.L. Martin, G.R. Sizilio, K.E. Dos Santos, B.G. de Araujo, R.A.Valentim, A.D.Neto, J.D.de Melo, A.M.Guerreiro, "Classification of cardiac arrhythmias using competitive networks", *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, 2010, pp.1386–1389
- [30] G.Valenza, A.Lanata, M.Ferro, E.P.Scilingo, "Real-time discrimination of multiple cardiac arrhythmias for wearable systems based on neural networks", *Computers in Cardiology*, Vol. 35, 2008, pp. 1053–1056
- [31] B. Tighiouart, P. Rubel, M. Bedda, "Improvement of QRS boundary recognition by means of unsupervised learning", *Computers in Cardiology*, Vol. 30, 2003, pp. 49–52
- [32] M. Kolasa, R. Długosz, W. Pedrycz, M. Szulc, "Programmable Triangular Neighborhood Function for Kohonen Self-Organizing Map Implemented on Chip", *Neural Networks*, Elsevier, Vol. 25, pp.146–160, (January 2012)
- [33] T. Talaška and R. Długosz, "Current Mode Euclidean Distance Calculation Circuit for Kohonen's Neural Network Implemented in CMOS  $0.18\mu\text{m}$  Technology", *Canadian Conference on Electrical and Computer Engineering*, pp.437–440, April 2007
- [34] Shen-Iuan Liu and Cheng-Chieh Chang, "A CMOS Square-Law Vector Summation Circuit", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 43, No. 7, 1996, pp.520–524
- [35] D. Fernández, L. Martínez-Alvarado, J. Madrenas, "A Translinear, Log-Domain FPAA on Standard CMOS Technology", *IEEE Journal of Solid-State Circuits*, Vol. 47, No. 2, Feb. 2012, pp.490–503
- [36] Yu-Cherng Hung, Bin-Da Liu, "1-V CMOS similarity measurement chip for binary pattern identification", *International Workshop on Cellular Neural Networks and Their Applications*, 2005, pp.36–39

- [37] S. Khucharoensin, V. Kasemsuwan, "A High Performance CMOS Current-Mode Precision Full-Wave Rectifier (PFWR)", *International Symposium on Circuits and Systems*, Vol.1, 2003, pp. 1-41-1-44
- [38] B. Boonchu, W. Surakampontom, "A CMOS current-mode squarer/rectifier circuit", *International Symposium on Circuits and Systems (ISCAS)*, Vol.1, 25-28, 2003, pp.1-405-1-408
- [39] Y. Oike, M. Ikeda, K. Asada, "A High-Speed and Low-Voltage Associative Co-Processor with Exact Hamming/Manhattan-Distance Estimation Using Word-Parallel and Hierarchical Search Architecture", *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 8, Aug. 2004, pp. 1383-1387
- [40] S. Sasaki, M. Yasuda, H.J. Mattausch, "Digital Associative Memory for Word-Parallel Manhattan-Distance-Based Vector Quantization" 38<sup>th</sup> *European Solid-State Circuit conference (ESSCIRC 2012)*, France, Sept. 2012, pp. 185-188
- [41] H.J. Mattausch, W. Imafuku, A. Kawabata, T. Ansari, M. Yasuda, T. Koide, "Associative Memory for Nearest-Hamming-Distance Search Based on Frequency Mapping" *IEEE Journal of Solid-State Circuits*, Vol. 47, No. 6, June 2012, pp. 1448-1459
- [42] E. Farshidi, N. Alaei-sheini, "A micropower current-mode pattern-matching classifier circuit using FG-MOS transistors", *IEEE Signal Processing and Communications Applications Conference*, 9-11 April 2009, pp.860-863
- [43] M.J.M. Pelgrom, H.P. Tuinhout and M. Vertregt, "Transistor matching in analog CMOS applications", *IEEE International Electron Devices Meeting*, December 1998, pp.915-918
- [44] J.A. Croon, Maarten Rosmeulen, Stefaan Decoutere, Willy Sansen, Herman E. Maes, "An Easy-to-Use Mismatch Model for the MOS Transistor", *IEEE Journal of Solid State Circuits*, Vol. 37, No. 8, August 2002
- [45] Xiaobin Yuan, Takashi Shimizu, Umashankar Mahalingam, Jeffrey S. Brown, Kazi Z. Habib, Daniel G. Tekleab, Tai-Chi Su, Sarkar Satadru, C. Michael Olsen, Hyunwoo Lee, Li-Hong Pan, Terence B. Hook, Jin-Ping Han, Jae-Eun Park, Myung-Hee Na, and Ken Rim, "Transistor Mismatch Properties in Deep-Submicrometer CMOS Technologies", *IEEE Transactions on Electron Devices*, Vol. 58, No. 2, February 2011
- [46] Terence B. Hook, Jeffrey B. Johnson, Jin-Ping Han, Andrew Pond, Takashi Shimizu, and Gen Tsutsui, "Channel Length and Threshold Voltage Dependence of Transistor Mismatch in a 32-nm HKMG Technology", *IEEE Transactions on Electron Devices*, Vol. 57, No. 10, October 2010
- [47] X. Yuan, Q. Zhang, H. Tran, S. Fox and M. Sherony "Effect of SiGe channel on pFET variability in 32 nm technology", *Electronics Letters*, 1<sup>st</sup> March 2012, Vol. 48, No. 5, pp.273-274
- [48] Augustin Cathignol, Krysten Rochereau, Samuel Bordez, Grard Ghibaudo, "Improved Methodology for Better Accuracy on Transistors Matching Characterization" 2006 *International Conference on Microelectronic Test Structures*, pp.173-178
- [49] Marcel J.M. Pelgrom, *Analog-to-Digital Conversion*, Springer December 6, 2012, ISBN-10: 1461413702, ISBN-13: 978-1461413707, Edition: 2<sup>nd</sup> ed. 2013,
- [50] O. Inan, L. Giovanrandi, and G. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features", *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2507-2515, Dec. 2006
- [51] L. He, W. Hou, X. Zhen, and C. Peng, "Recognition of ECG patterns using artificial neural network", *6th International Conference Intelligent Systems, Design Appl.*, vol. 2. Jinan, China, Oct. 2006, pp. 477-481



**Tomasz Talaska** received the M.Sc. degree and the Ph.D. degree in telecommunication from the Faculty of Telecommunication, Computer Science and Electrical Engineering, UTP University of Sciences and Technology, Bydgoszcz, Poland, in 2002 and 2009, respectively. Currently, he is an Assistant Professor at the UTP University of Sciences and Technology, Bydgoszcz, Poland. His research areas are ultra low power analog and analog-digital integrated circuits and hardware implementation of the artificial neural networks especially self-organized neural networks.

He is the fellow of two research grants granted by European Union and by Polish Government. He is coauthor of more than 50 research papers.



especially analog-digital application-specific integrated circuit self-organized neural networks.

**Marta Kolasa** received the M.Sc. degree and the Ph.D. degree in telecommunications from the Faculty of Telecommunications, Computer Science and Electrical Engineering UTP University of Science and Technology, Bydgoszcz, Poland, in 2005 and 2012, respectively. She is currently an Assistant Professor with the UTP University of Science and Technology. She is co-author of more than 40 research papers. Her current research interests include energy efficient analog-digital integrated circuits, artificial neural networks and their hardware implementation,



**Rafał Długosz** received the M.Sc. degree in Control and Robotics and the Ph.D. degree in Telecommunications (with distinctions) from Poznan University of Technology (PUT) in Poland, in 1996 and 2004, respectively. Currently, he is Active Safety R&D Team Leader in Delphi Poland S.A. (Krakow) and Assistant Professor at the University of Sciences and Technology, Bydgoszcz, Poland. He was the fellow of several scientific fellowships that include Kolumb scholarship granted by Foundation for Polish Science in Poland, Marie Curie Outgoing International Fellowship granted by European Union and scholarship granted by Deutscher Akademischer Austausch Dienst (DAAD). During these fellowships, between 2005 and 2008, he was with the Department of Electrical and Computer Engineering, University of Alberta, Canada. Then he joined Electronics and Signal Processing Laboratory (ESPLAB) Institute of Microtechnology (IMT) at Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. He spent also three months in Innovations for High Performance (IHP) Microelectronics Institute in Frankfurt (Oder) in Germany. He has published over 120 research papers and book chapters. His main research areas comprise ultra-low power reconfigurable analog and mixed analog-digital circuits such as analog and digital filters, analog-to-digital converters (ADCs), artificial neural networks, fuzzy logic systems and other specialized circuits.



**Witold Pedrycz** is Professor and Canada Research Chair (CRC) in Computational Intelligence in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. He holds an appointment of special professorship in the School of Computer Science, University of Nottingham, UK. In 2009 Dr. Pedrycz was elected a foreign member of the Polish Academy of Sciences. In 2012 he was elected a Fellow of the Royal Society of Canada. Witold Pedrycz has been a member of numerous program committees of IEEE conferences in the area of fuzzy sets and neurocomputing. In 2007 he received a prestigious Norbert Wiener award from the IEEE Systems, Man, and Cybernetics Council. He is a recipient of the IEEE Canada Computer Engineering Medal 2008. In 2009 he has received a Cajastur Prize for Soft Computing from the European Centre for Soft Computing for pioneering and multifaceted contributions to Granular Computing. In 2013 has was awarded a Killam Prize. In the same year he received a Fuzzy Pioneer Award 2013 from the IEEE Computational Intelligence Society.

His main research directions involve Computational Intelligence, fuzzy modeling and Granular Computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and Software Engineering. He has published numerous papers in this area. He is also an author of 15 research monographs covering various aspects of Computational Intelligence, data mining, and Software Engineering.