

Low-Voltage CMOS Circuits for Analog VLSI Programmable Neural Networks

دوائر قياسية منخفضة الجهد ذات كثافة عالية للشبكات العصبية
المبرمجة

Mohy A. Abo El-soud¹, Roshdy A. AbdelRassoul², Senior Members IEEE, Hassan H. Soliman¹ and Laila M. El-ghanam¹.

¹ Electronics & Comm. Eng. Dept., Faculty of Eng., Mansoura University,

² Arabic Academy for Science & Technology and Maritime Transport, Alexandria.

الملخص العربي:

في هذا البحث تم تصميم دوائر قياسية ذات كثافة عالية لأحد أنواع الشبكات العصبية الصناعية وهو النوع متعدد الطبقات ذي التغذية الأمامية منخفض الجهد (low-voltage multilayer feedforward artificial neural network) وذلك باستخدام تكنولوجيا ترانزستور أكسيد معدن شبه الموصل المكمل (CMOS technology). فالتقدم في هذه التكنولوجيا جعل جهد تشغيل الدوائر المتكاملة أكثر انخفاضا. فكلما نقص الجهد المطلوب كلما صغرت أبعاد الترانزستور وقلت القدرة المستهلكة. ووحدة البناء الأساسية للنظام هي الخلية العصبية (neuron) والتي تنقسم بدورها إلى ثلاث أجزاء أساسية وهم الأوزان العصبية (synaptic weights)، وعنصر المعالجة (processing element)، والدالة الزائدية اللاخطية (nonlinear Tanh activation function). وبالتالي كانت الخطوة الأولى هي تصميم الخلية العصبية بكل أجزائها كل على حده والذي تم تقديمه في هذا البحث، وكذلك تم استخدام الخلية المقترحة في بناء شبكة عصبية عامة للتعرف على الأنماط المختلفة للكتابة.

Abstract:

This paper presents an overview of designing an analog VLSI system used for handwritten recognition; namely a programmable neural network in linear as well as subthreshold CMOS technology. Synaptic weights are designed in the triode region. In addition, the processing element (PE) and the Tanh activation function are designed in subthreshold region. Such subthreshold CMOS technology has some interesting features, such as high integration density, exponential transfer characteristics and low-power consumption. The proposed system is realized in a standard 0.8µm CMOS technology and operated with a ± 1V power supply.

Keywords: Low-voltage analog VLSI, programmable neural networks, handwritten character recognition.

1. Introduction

The main benefit of the analog approach is that a wide range of operations can be performed using a small number of transistors. This means simple basic blocks and

connections, which lead to a small area and thus larger networks can be built on a single chip. Analog VLSI has been identified as a major technology for future information processing. This is primarily because some of the traditional analog designs limitations such as accurate absolute component values, device matching, and precise time constants are not major concerns in artificial neural networks (ANNs). This is due to the fact that computation precision of individual neurons does not seem to be of great importance.

Many authors use several approaches for the analog implementations of ANNs in VLSI technology. Among them are MOS transistors working in continuous-time [1], in weak [2], or strong inversion [3], and linear multiplier as synaptic weight [4].

This paper is organized as follows: section 2 introduces the proposed neuron description. First, we present three-MOS transistors as synaptic weight using switched-resistor (SR) technique. Next, we describe the PE as well as Tanh activation function operating in subthreshold region. Simulation results are introduced in section 3, and section 4 gives the application of handwritten character recognition. Finally section 5 presents the conclusions.

2. Proposed Neuron Description

An artificial neuron can be modeled by a set of weights, which multiply the input elements to the neuron. The neuron sums the weighted inputs and passes the result through a hard limiter or sigmoidal function.

2.1 Synaptic Weight Circuit

Fig. 1 shows an adaptive weight circuit. Such circuit consists of simple three transistor synapses similar to that in [1] but using SR technique for transistor M_3 with capacitor C , and a buffer stage.

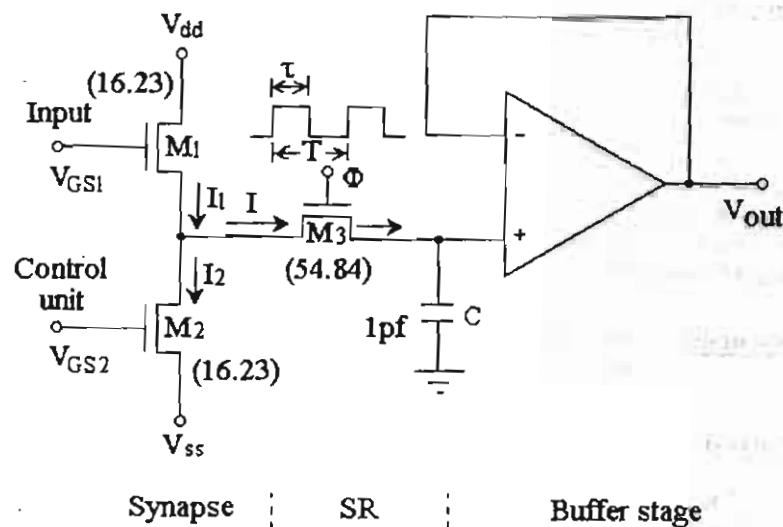


Fig. 1 Design of the programmable synaptic weight.

Then the output current I is applied to the drain of M_3 , which has an equivalent resistance using SR technique given by [5]:

$$R_{eq3} = R_{on3}/d \quad (1)$$

Where R_{on3} is the *on*-resistance of M_3 , and d is the duty cycle of the clock signal Φ . It is assumed that a positive synaptic weight is realized when $V_{GS1} < V_{GS2}$, and the negative weight when $V_{GS1} = V_{GS2}$, the case of zero output current. The capacitor C converts the output current of M_3 to a voltage change at the processing node. The buffer stage is used to match between the synaptic weight and the PE.

2.2 Processing Element

In subthreshold MOS technology, processing (summing) circuit as well as Tanh activation function are built since, it offers advantages of high integration density, low-power dissipation and exponential transfer characteristics [6]. For an NMOS transistor operating in subthreshold (weak inversion) region, the drain-to-source current depends exponentially on gate-source voltage and is given by [7]:

$$I_{DS} = I_0 e^{kVG - VS} \quad (2)$$

Where $I_0 = 10^{-12}$ A, k is constant, V_G is the gate voltage, and V_S is the source voltage.

Fig. 3 shows the PE as a two-stage CMOS operational amplifier (op-amp) with aspect ratios of MOS transistors. The differential input transistors M_4 - M_5 amplify the differential input signal and operate in weak inversion. Thus, the total transconductance g_m would be constant and rather small [8]. The current mirror formed by M_{11} and M_8 supplies the differential pair with the biasing current. The W/L ratio of M_8 is selected to yield the desired input-stage bias. The input differential pair is actively loaded with the current mirror formed by M_6 and M_7 . The *second gain stage* consists of M_9 , which is a common-source amplifier actively loaded with the current-source transistor M_{10} [9]. The maximum output voltage swing of the output stage is limited by: $V_{ss} + V_{ds(sat)n} < V_{out} < V_{dd} - V_{ds(sat)p}$ [10]. The frequency compensation is implemented using a Miller-feedback capacitor C_C [9].

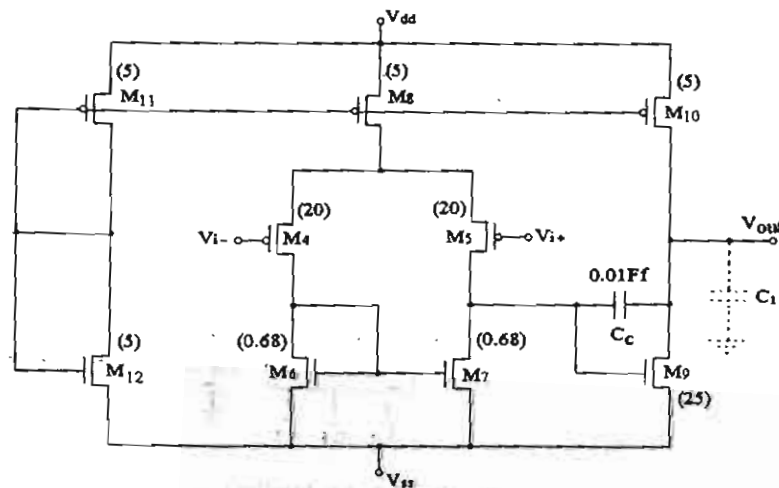


Fig. 3 Processing element as a two-stage CMOS op-amp.

2.3 Tanh Activation Function

Fig. 4 shows the Tanh activation function. Such function is basically obtained from the input differential pair M_{13} - M_{14} operates in subthreshold region (i.e. $V_{GS} < V_T$). The drain of M_{14} is connected to the current mirror formed by M_{18} and M_{19} . Similarly, the drain of M_{13} is connected to the current mirror formed by M_{17} and M_{16} . Thus, the currents coming out of M_{16} and M_{19} are just the two halves of the current in the differential pair. M_{19} current is then reflected through M_{21} and M_{20} , and is subtracted from I_{16} to form the output. Thus, the output current is just the difference between I_{13} and I_{14} . Hence this circuit generates an output current that is proportional to the difference between the two drain currents of the differential pair that represents Tanh function [6] as seen below.

The manner in which I_{15} is divided between M_{13} and M_{14} is sensitive function of the difference between V_1 and V_2 . From equation (2), the difference current, which represents the Tanh function is given by [7]

$$I_{13} - I_{14} = I_{15} \tanh \frac{k(V_1 - V_2)}{2} \quad (3)$$

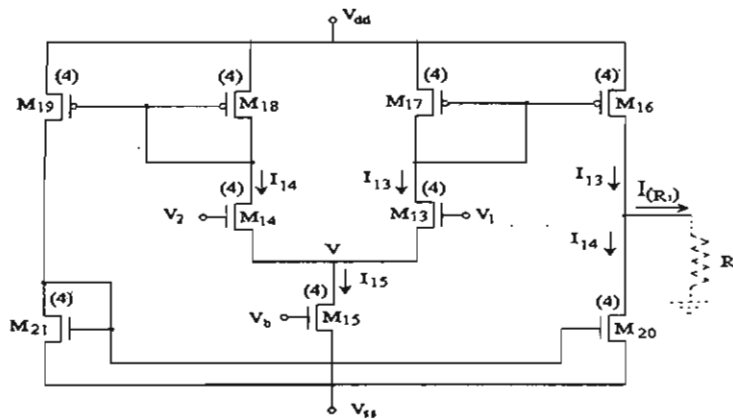


Fig. 4 The suggested Tanh configuration.

Finally, the complete design of the proposed neuron is shown in Fig. 5.

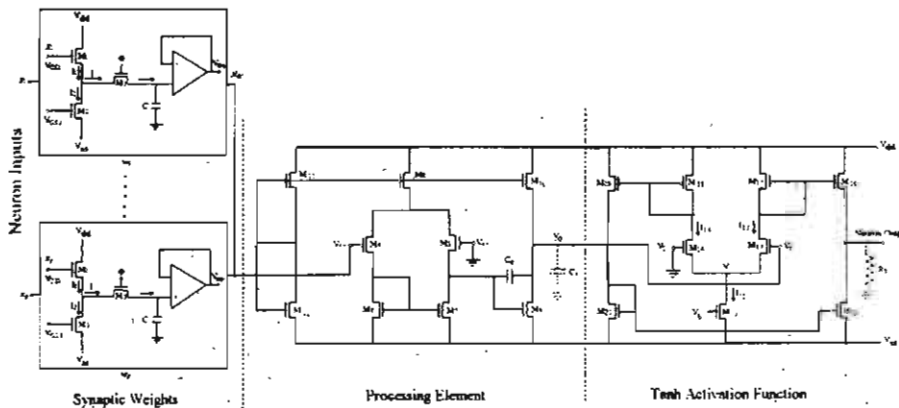


Fig. 5 Final design of the suggested neuron.

3. Simulation Results

The simulation results of the designed system will be introduced through this section. The system is simulated using the SPICE program with $0.8\mu\text{m}$ CMOS technology and $\pm 1\text{V}$ power supply.

3.1 Simulation Results of Synaptic Weight

The simulation results of the synaptic weight shown in Fig. 1 are given in Fig. 6. It is clear that, an approximately linear range of the output current from $-48.107\mu\text{A}$ to $-168.889\mu\text{A}$ is obtained over the range of the differential input voltage extends from 400mV to 1V .

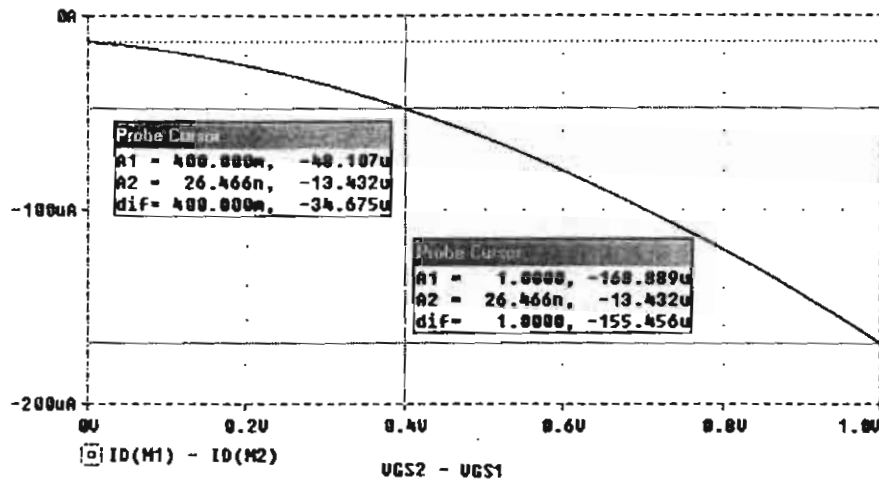


Fig. 6 Transfer characteristics of the synaptic weight.

3.2 Simulation Results of Processing Element

The simulation results of the PE shown in Fig. 3 using SPICE program are introduced here. PE has been tested using 1pF capacitor load and 0.01fF compensation capacitor.

3.2.1 DC Analysis of CMOS Op Amp

The first analysis requested is a DC sweep of the input differential voltage V_d between the two supply limits with the input common-mode level V_{CM} set to 0V . The resulting large-signal differential-mode transfer characteristics are shown in Fig. 7.

From the graph several characteristics of the op amp could be observed. First, the linear region of the amplifier is bounded between the input voltage level of -200mV and $+200\text{mV}$. Conversely, this corresponds to a maximum output voltage swing bounded in the negative direction by -0.9874V and in the positive direction by $+1\text{V}$. Also, it can be noticed that the input-referred offset voltage is $+212.887\mu\text{V}$.

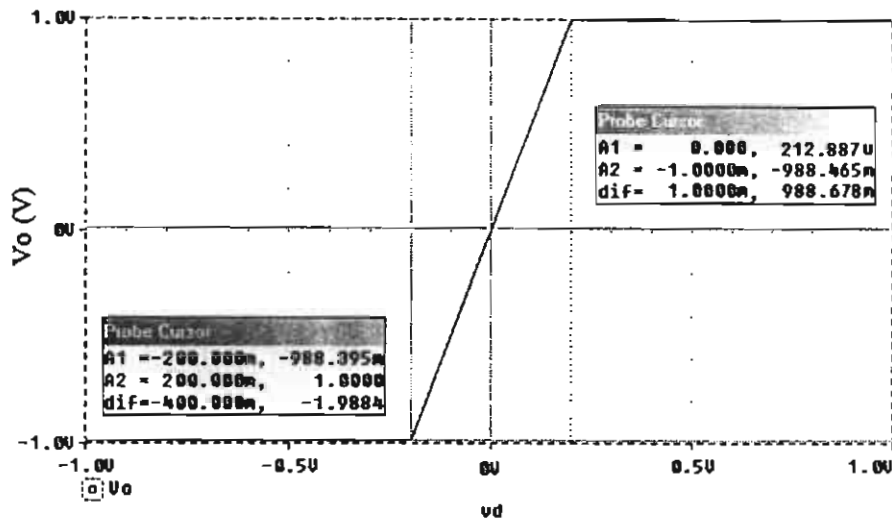


Fig. 7 Differential mode characteristics.

3.2.2 Frequency Response of CMOS Op Amp

To compute the differential magnitude of the op amp using SPICE, a 1V amplitude AC voltage is commonly chosen for the input signal because the output voltage then directly represents the transfer function of the circuit, and the input DC offset voltage remains at $-212.887\mu\text{V}$ in order to keep the op amp in its linear region.

The results of the frequency-domain analysis are shown in Fig. 8. The unity-gain frequency f_i is read directly from the graph using probe as 5.5688 MHz.

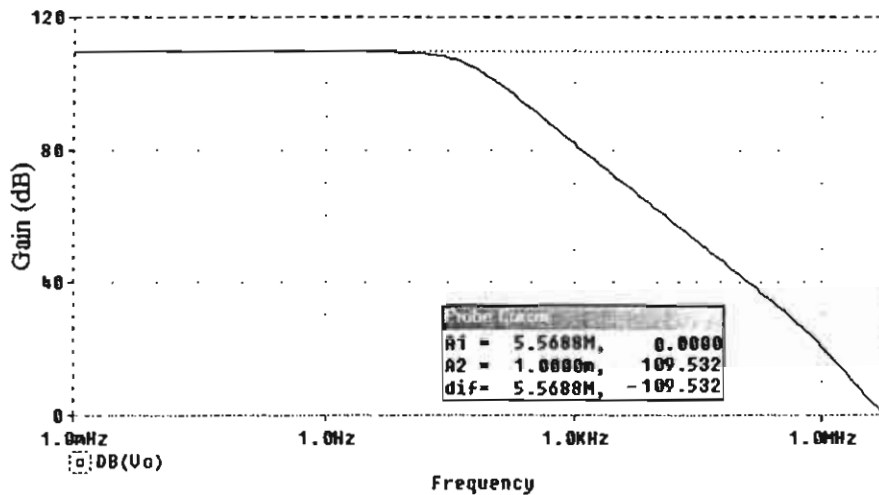


Fig. 8 Op amp frequency response in dB.

3.3 Simulation Results of Tanh Activation Function

Finally, the simulation results of the Tanh activation function shown in Fig. 4 are presented in Fig. 9.

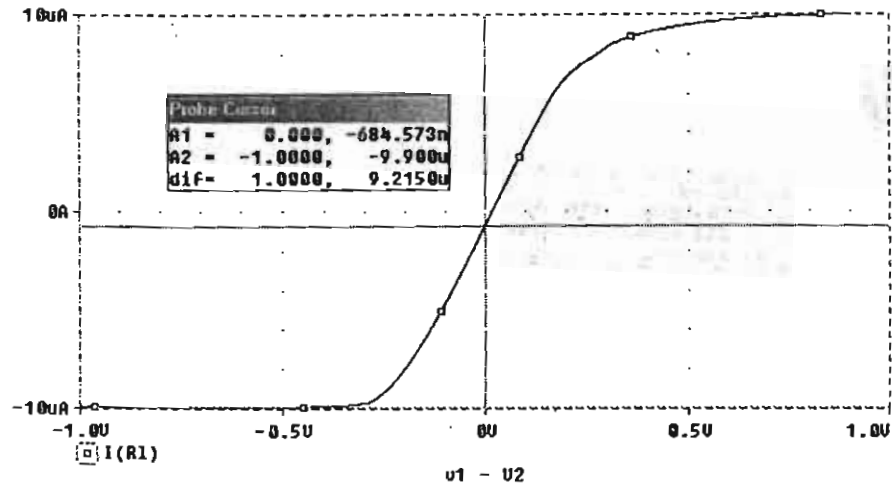


Fig. 9 Transfer characteristics of Tanh.

4. Handwritten Character Recognition using Multilayer Neural Network

Handwriting is a natural means of communication which nearly everyone learns at an early age. In many situations, writing by hand is the fastest and most convenient way to communicate with another person. The recognition of natural writing with totally unconstrained words remains a very challenging problem in pattern recognition. It is due to the large variety of handwriting styles, ambiguous boundaries of character strings and considerable big variation in letter shape [11]. So, it is useful to examine this field more closely and to identify its several areas.

4.1 Image Resizing

The underlying concept of image resizing is to use a simple preprocessor to reduce the dimensionality of the inputs to the neural network. The preprocessor takes the original images as inputs and groups neighboring pixels in blocks. The sum of the number of "on" pixels in each block is used to create a pixel in a resulting image. The resulting image is then fed into the neural network. In essence, the preprocessor removes the finer details of the image and creates a new image based on the general locality of the pixels.

CEDAR CDROM1 databases taken from the Internet [12] are used in this paper. Chosen examples from these databases are shown in Fig. 10, which includes 5 shapes of character A and 10 shapes of digit 7. The images are first resized and the features of the resized images are given in Tables 1 and 2.

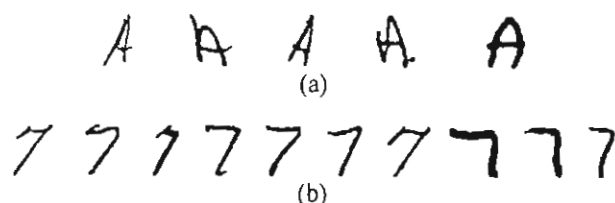


Fig. 10 Chosen sample handwritten images.

Table 1 Features of the resized character A images.

A	A	A	A	A
0.1294	0.902	0.1294	0.0157	0.0941
0.1059	0.1373	0.1882	0.1255	0
0.0510	0	0.0039	0.1882	0.0392
0.1294	0	0	0.0588	0

Table 2 Features of the resized digit 7 images.

7	7	7	7	7	7	7	7	7	7
0.1059	0.1333	0.0039	0.1059	0.1333	0.1137	0.1608	0.1216	0.1098	0.0118
0.1725	0.1333	0	0.1725	0	0.1333	0	0.1686	0.1216	0
0.1059	0.1333	0	0.1059	0	0.1137	0.1137	0.1333	0.1098	0.0118
0.1059	0.1333	0.1137	0.1059	0.1333	0.1137	0.1137	0.0706	0.0980	0.0118

Then, the two studied cases are trained and tested using MATLAB 6 library by introducing them to a multilayer feedforward neural network, which is suggested to have three layers; an input layer, a hidden layer, and an output layer. The input layer consists of 4 neurons, the hidden layer has 3 neurons, and the output layer has 2 neurons corresponding to the number of the input cases required to be recognized. Classification results depend on training and testing processes of input classes. The specifications used in the training process are listed in Table 3. The "hold-half-out" method is used to estimate the probability of error of the classifier. The train and test sizes are each 50 percent of the total data.

Table 3 Training specifications.

Training Algorithm	Back-Propagation (BP)
Performance Function	Mean-Square Error
Performance Goal	$1e^{-3}$
Max. Number of Epochs	10000
Learning Rate	0.5

Classification results are given in Table 4, which shows the resulting percentage of both training and testing sets. It is concluded from the results that training and testing are done with one hundred percentage.

Table 4 Classification results for the two studied cases.

Studied Case		Network Architecture	Training Set	Testing Set
Case I	5 shapes of character A	(4:3:4)	100%	100%
Case II	10 shapes of digit 7			

5. Conclusions

In this paper, the hardware implementation of a feedforward neural network has been presented. The implemented circuit of the synapse has been realized using an analog circuit and SR technique. Furthermore, the PE and the Tanh activation function have been designed in weak inversion. Such system has been used for recognizing handwritten alphanumeric characters taken from CEDAR databases.

References

- [1] Alan Murray and Lionel Tarassenko, "Analogue Neural VLSI-A Pulse Stream Approach", Chapman & Hall Publishing Company, 1994.
- [2] E. A. Vittoz, "Analog VLSI implementation of neural networks," in IEEE Int. Symp. Circuits Syst. Proc., 1990, PP. 2524-2527.
- [3] J. B. Lont and W. Guggenbuhl, "Analog CMOS implementation of multilayer perceptron with nonlinear synapses," IEEE Trans. on neural networks, Vol. 3, No. 3, PP. 457-462, May 1992.
- [4] T. Morishita et al, "A BiCMOS analog neural network with dynamically updated weights," in IEEE Int. Solid-State Circuits Conf. Proc., 1990, PP. 142-143.
- [5] M. A. Abo-El-soud, "Analog circuits for electronic neural network," The Proc. of 35th Midwest Symp. on Circuits and Syst., PP. 5-8, Aug. 1992.
- [6] Andreas G. Andreou, Kwabena A. Boahen, Philippe O. Pouliquen, Aleksandra Pavasovic, Robert E. Jenkins and Kim Strohhenn, "Current-Mode Subthreshold MOS Circuits for Analog VLSI Neural Systems," IEEE Trans. on neural networks, Vol. 2, No. 2, March 1991.
- [7] Mead C., "Analog VLSI and Neural Systems", Addison-Wesley Publishing Company, 1989.
- [8] Johan H. Huijsing, Ron Hogervorst, and Klass-Jan de Langen, "Low-power Low-voltage VLSI Operational Amplifier Cells," IEEE Trans. on Circuits and Syst.-I: Fundamental theory and applications, Vol. 42, No. 11, November 1995.
- [9] Adel S. Sedra and Kenneth C. Smith, "Microelectronic Circuits", Fourth Edition, Oxford University Press, Inc., New York, 1998.
- [10] Ismail, M. and T. Fiez, "Analog VLSI: Signal and Information Processing", First Edition, Mc-Graw-Hill Companies, 1993.
- [11] Bunke, H., et al., "Off-Line Cursive Handwriting Recognition using Hidden Markov Models," *Pattern Recognition*, 28 (9), PP. 1399-1413, 1995
- [12] Available from:
<http://www.cedar.buffalo.edu/Databases/CDROM1/>