DEVELOPMENT OF RECONFIGURABLE NONLINEAR CIRCUITS FOR NEURAL NETWORKS

By

ABDULLAH B. SHWEHNEH

A Thesis Presented to the

DEANSHIP OF GRADUATE STUDIES

In Partial Fulfillment of the Requirements

for the degree

MASTER OF SCIENCE

IN

ELECTRICAL ENGINEERING

KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS

Dhahran, Saudi Arabia

June 2005

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by Abdullah B. Shwehneh under the direction of his thesis advisor and approved by his thesis committee, has been presented to and accepted by Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN ELECTRICAL ENGINEERING.

Thesis Committee

Dr. M. T. Abuelma'atti (Advisor)

Dr. Hussain Al-Zaher (Member)

Dr. S. M. Al-Shahrani (Member)

J. M. Bakhashwain

Dr. Jamil M. Bakhashwain
(Department Chairman)

Dr. Mohammad A. Al-Ohali
(Dean of Graduate Studies)

[Date] 13-6-2005

To my Family …

ACKNOWLEDGMENT

All praise and glory is to almighty Allah who gave me the courage and patience to carry out this work, and peace and blessings of Allah be upon his last prophet Muhammad.

Acknowledgement is due to King Fahd University of Petroleum and Minerals for supporting this research. I am grateful to the electrical engineering department, faculty and staff, for all means of support that it offered to me.

I would like to express my deep gratitude to my thesis advisor Prof. Muhammad Taher Abuelma'atti for his unconditional help, his encouragement and his valuable suggestions during the preparation of my thesis. I would also like to thank my committee members Dr. Hussain Al-Zaher and Dr. Saad Al-Shahrani for spending their time reading my thesis and for their constructive suggestions and comments.

Furthermore, I would like to express my deep and warm gratitude to my dear parents for their warm support through all these years, for the religious and scientific foundation they gave me and for the prayers they made for me. May Allah bless them and give them a long life with all good deeds. I would like also to express my deep and warm gratitude to my dear wife for her support and patience through my study.

At the end of this acknowledgement, I would like to express my appreciation to my colleagues and to people who gave me support in KFUPM and to all my friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Thesis Abstract

Name: ABDULLAH BAKRI SHWEHNEH

Title: DEVELOPMENT OF RECONFIGURABLE NONLINEAR CIRCUITS FOR NEURAL NETWORKS

Degree: MASTER OF SCIENCE

Major field: ELECTRICAL ENGINEERING

Date of degree: March 2005

In this thesis, a new generic CMOS circuit for realizing different nonlinear functions from the same topology is presented. This circuit can be very useful in Neural Networks applications as it implements the main nonlinearities required by many types of Neural Networks. With transistors operating in strong inversion the circuit can be digitally configured to realize any of the following four functions: Gaussian (Radial Basis), Sigmoid and two piecewise linear functions – Triangular and Satlin. The circuit can approximate these functions with RRMS error less than 1%. It is shown that the center, width, peak amplitude and slope of the dc transfer curve can be independently controlled. Simulation results using 0.18 CMOS process model parameters of TSMC technology are included.

Keywords: Analog Circuits, Programmable Neural Networks.

خلاصة الرسالة

الإسم : عبد الله بكري شويحنة

عنوان الرسالة: تصميم دارة الكترونية غير خطية قابلة للبرمجة من أجل تطبيقات الشبكات العصبية

الدرجة الممنوحة : ماجستير في العلوم

حقل التخصص: الهندسة الكهربائية

تاريخ منح الدرجة: **مارس 2005م**

تناقش هذه الرسالة تصميم دارة الكترونية متعددة الوظائف بتقنية الترانزستور الحقلي (CMOS) قادرة على تمثيل وظائف مختلفة غير خطية انطلاقاً من نفس التصميم. هذه الدارة يمكن أن تكون مفيدة جداً في تطبيقات الشبكات العصبية نظراً لإمكانيتها تمثيل معظم الوظائف الغير خطية التي تحتاجها الشبكات العصبية. الدارة قابلة للتحكم رقمياً لاختيار الوظيفة المراد تمثيلها. هذه الدارة قابلة لتمثيل أربع وظائف غير خطية هي: منحني غاوس (Gaussian) ، منحني S (Sigmoid) ، التابع المثلثي و التابع الخطي. تستطيع هذه الدارة تمثيل هذه الوظائف (التوابع) بخطأ تقريبي مقداره 1% . المركز و العرض و الارتفاع والميل لهذه الوظائف، هي جميعها قابلة للتحكم بشكل منفصل. المحاكاة بواسطة الكمبيوتر تستخدم تقنية TSMC لتحديد المعطيات المتعلقة بعناصر الدارة.

مفاتيح: الدارات الإلكترونية التماثلية، الشبكات العصبية القابلة للبرمجة.

درجة الماجستير في العلوم

جامعة الملك فهد للبترول و المعادن

يونيو 2005م

CHAPTER 1

Introduction and Literature Survey

Artificial Neural Networks (ANNs) have shown great promise in different applications including pattern recognition, identification, classification, speech, vision and control systems. Neural Networks (NNs) [1] are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Depending on the application, different types of NNs are used. Typical NN consists of very simple elements (Fig. 1-1), which are processing elements (neurons) and connecting elements (synapses). In general the task of a processing element is to sum up arriving signals and to perform a nonlinear operation on the sum, whereas a connecting element has to weight an arriving signal with a simple multiplication. Depending on the type of NN, the input-output characteristics of the neurons can have different shapes [1],[2]: piecewise-linear, sigmoid, or Gaussian-like shapes. The hardware realization of these processing elements will be considered in this thesis.

The field of neural networks has found solid application only in the past fifteen or twenty years [3-9], and is still developing rapidly. Today, the semiconductor industry is able to provide technologies for implementation of millions of transistors on a single chip. VLSI technology and future ULSI technologies are, and will be, a good implementation medium for artificial neural circuits. This fact, together with the development of new important network types and learning algorithms, caused a breakthrough in artificial neural network technology.

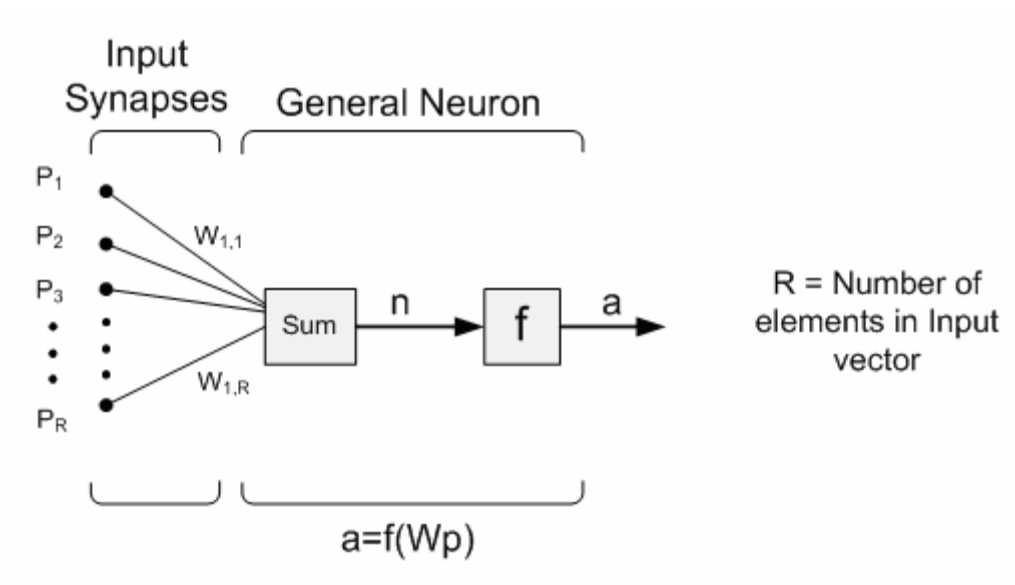


Figure 1-1 A simple model of a neuron.

Analog VLSI circuits using CMOS technology [3-5, 9-12] seem best suited for the implementation of neural networks. They usually require less power, less area and they are closer to the biological model than their digital counterparts. The continuous improvement of today's technology increases the operating frequency range of the CMOS circuits, and makes it superior with other technologies.

Analog implementations benefit from exploiting the characteristics of the devices. Just a single MOS-transistor can provide many functions:

- Generation of square, square-root, exponential and logarithmic functions
- Voltage-controlled current source
- Voltage controlled conductance, linear in a limited range
- Fixed or voltage-controlled pseudo-conductance, linear in a wide range
- Analog multiplication of voltages
- Short term and long term storage
- Light sensor
- Switch

Some of these characteristics are exploited in the implementation of different nonlinear functions used in NNs.

Neurons with different functions implementations (input-output characteristics) were implemented and reported in the literature: Piece-Wise linear [3-6] [13-23], Sigmoid [6, 9, 24-37] and Gaussian Functions [38-48]. Although all mentioned references report analog designs of these Neurons, digital counterparts are also available [49, 50]. What is common between most of the reported circuits is the one-function implementation. The differences are mainly in the mode of operation, operation range, accuracy and programmability of the implemented function. But no generic circuit has been published so far which can implement more than two of the main functions needed for NNs. Such a circuit can be very useful especially when different

functions are utilized in the same NN structure (for example, the transfer function of neurons in the hidden layer can be different from those in the output layer).

In this thesis, a new generic CMOS circuit for realizing different nonlinear functions from the same topology is proposed. This circuit can be used for NNs hardware implementation. Special attention is paid for the reconfigurability and programmability of the proposed circuit.

1. 1. Thesis Overview

This thesis consists of four chapters that deal mainly with problems associated with the design of configurable nonlinear circuits that could be used as a part of different ANNs applications.

- This first introductory chapter is focused on literature review of the main nonlinear and piece-wise linear circuits that already exist, trying to refer to their main advantages and disadvantages. The chapter concludes with the motivation toward developing a generic circuit which can implement different nonlinear functions required by ANNs applications with better programmability features.
- The second chapter starts by proposing our approach in building the generic circuit, which will be able to implement four different functions. The design will start with a circuit which implements one of the piece-wise linear functions (Triangular Basis Function). Through the rest of this chapter the

circuit will be modified to implement the other functions. Simulation results will be included for each function.

- Chapter Three discusses the proposed circuit and introduces more simulations which reveal the effect of possible variations in the parameters of transistors on the accuracy of the implemented functions. The effect of the variation in temperature is also considered.
- Chapter four includes a summary of the results obtained throughout this research and provides possible directions for future work.

1. 2. Neural Networks Transfer Functions

The simplest possible model of a NN is shown in Fig. 1-1. The job of each neuron (node) is to sum a number of weighted inputs (synapses) to generate an output signal which is a function of that sum.

Depending on each particular application, a number of these simplified neurons are connected together in different ways and programmed (learned) to recognize certain patterns by controlling the weight of each synapse. The transfer (activation) function (TF) f can be a piece-wise linear function, or can be a nonlinear function if a nonlinear function synthesis is of interest. Different types of TFs are considered next:

1. 2. 1. Hard-Limit transfer functions:

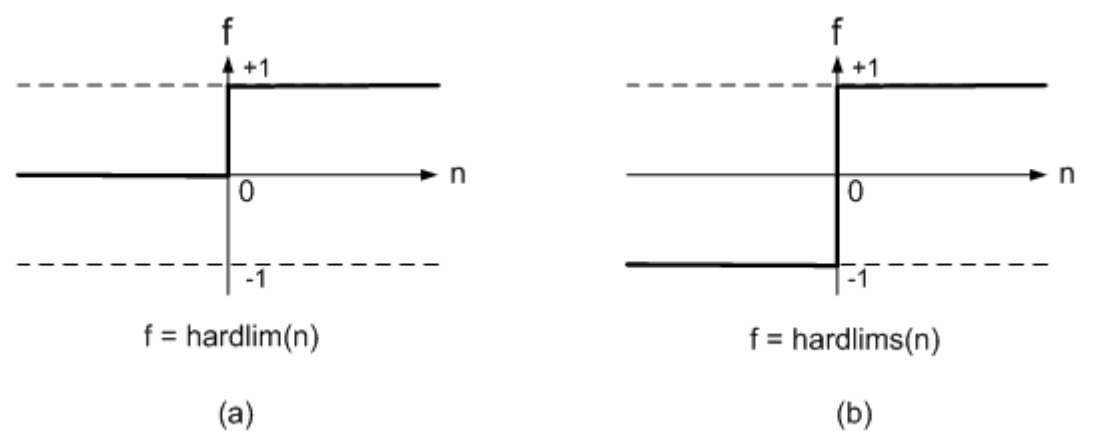


Figure 1-2 (a) Hard-Limit TF (b) Symmetric Hard-Limit TF

These TF's can be easily implemented using comparators.

1. 2. 2. Linear transfer functions:

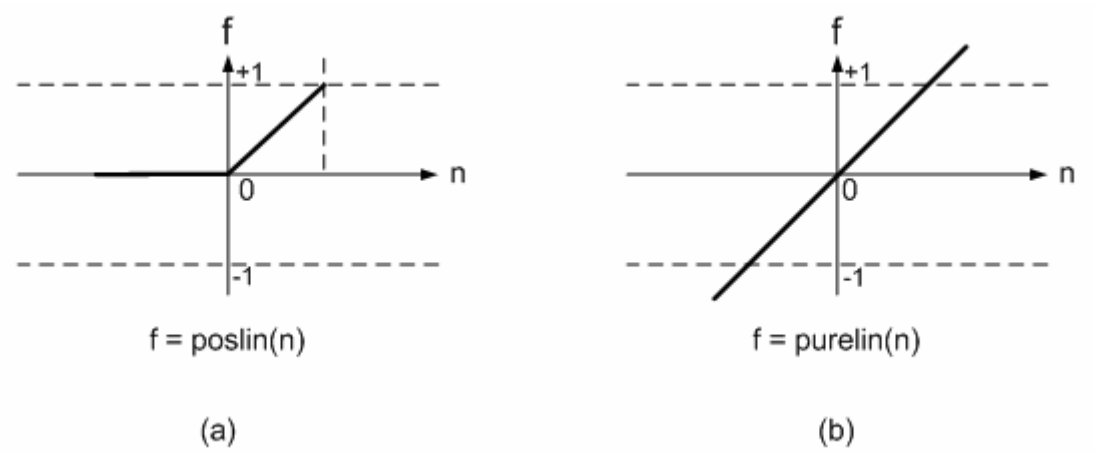


Figure 1-3 (a) Positive Linear TF (b) Linear TF

The simplest way to implement these functions is by using OpAmp configured as a noninverting amplifier. The positive or negative data interpolations will be decided by power supply values.

1. 2. 3. SatLinear transfer functions:

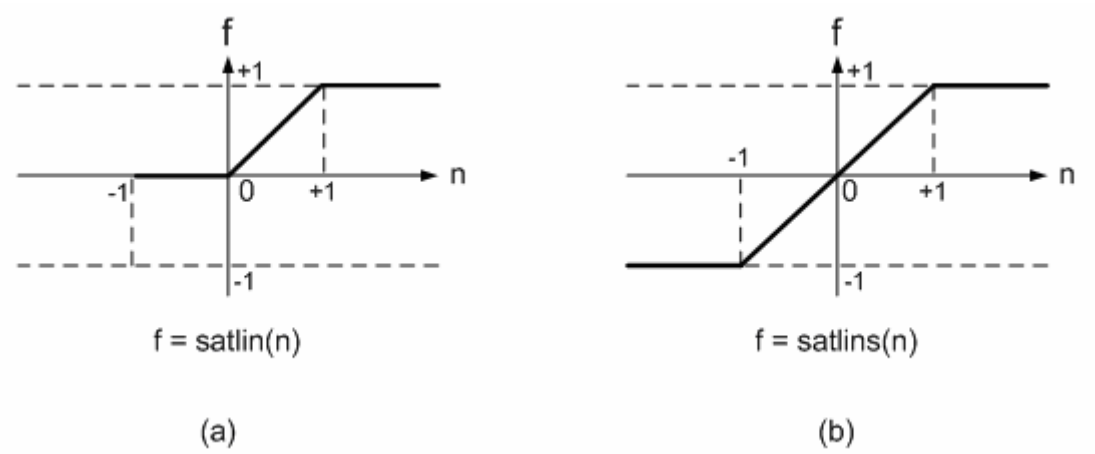


Figure 1-4 (a) Satlin TF (b) Satlins TF

These TFs can be implemented using a current conveyor [19,20] or a linear amplifier with desired saturation levels.

1. 2. 4. Sigmoid Functions:

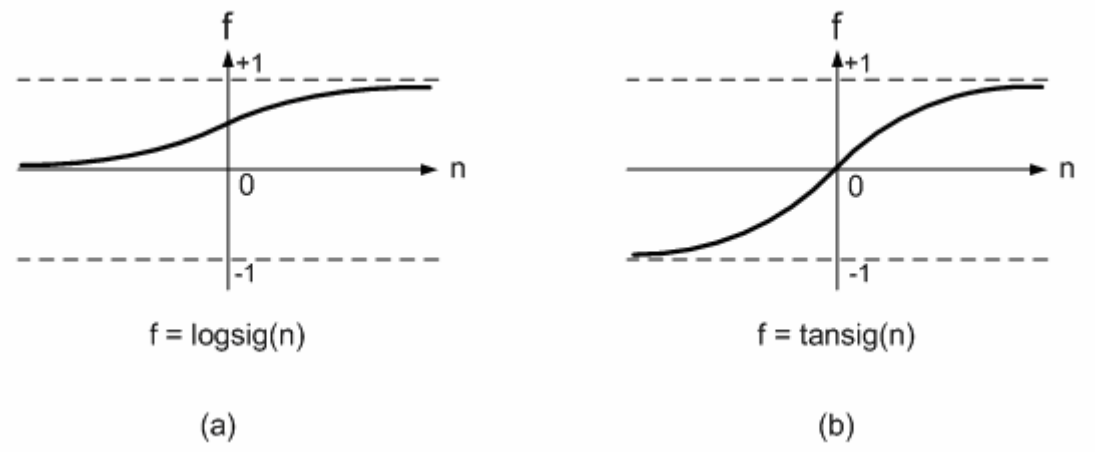


Figure 1-5 (a) Log-Sigmoid TF (b) Tan-Sigmoid TF

The Log-Sigmoid and the Tan-Sigmoid functions can be expressed as

$$\text{logsig}(n) = 1 / (1 + \exp(-n)) \tag{1-1}$$

and

tansig(n) = 2/(1+exp(-2*n))-1 (1-2)

The function tansig is very close to tanh function.

Another more detailed expression for the Sigmoid function is given by [1]:

$$f(n) = \frac{1}{1+e^{-\alpha(n+\theta)}} \tag{1-3}$$

were α is the gain factor of the function, and θ is the threshold of the function.

These functions are important if nonlinear response of the NN is of interest. They can be realized by cascading simple inverting amplifiers with input and feedback resistors [3,4,25] (see Figure 1-12), or by using CMOS differential pair [29,33,34].

1. 2. 5. Triangular and Radial Basis Functions:

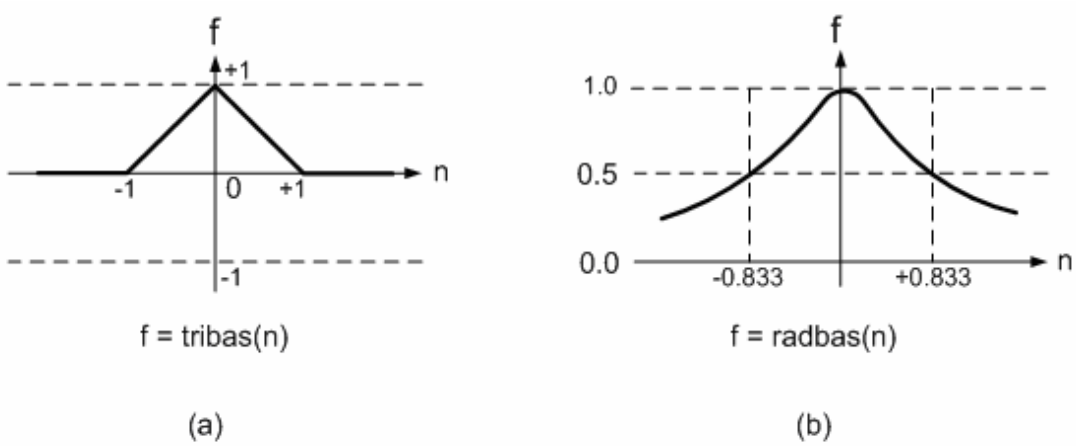


Figure 1-6 (a) Triangular Basis Function (b) Radial Basis Function

The Triangular Basis Function (TBF) is given by:

$$tribas(n) = 1 - |n|, \quad \text{if } -1 \leq n \leq 1; \tag{1-4}$$

$$= 0, \quad \text{otherwise.}$$

And the Gaussian or Radial Basis Function (RBF) is given by:

$$radbas(n) = exp(-n^2); \tag{1-5}$$

More detailed expression for the RBF function is given by [40]:

$$f(n) = e^{-\frac{(n-m_n)^2}{2\sigma^2}} \tag{1-6}$$

Were m_n is the mean of the RBF function, and σ is the Standard Deviation (STD).

A simple example of TBF implementation is an input class AB stage with additional biasing circuitry at the output [21] . The RBF can be generated using a simple CMOS circuit, biased in weak inversion, called 'correlator' [42].

1. 3. Literature Review

As ANNs applications advance rapidly, many circuits emulating different parts of NNs were published in the literature. A special attention was paid to the compatibility of these circuits with VLSI implementation. Following is a brief review of different existing implementations. For CMOS circuits, floating bulk terminals are assumed to be connected to either V_{DD} (PMOS transistors) or V_{SS} (NMOS transistors), unless otherwise is shown. A review of piece-wise linear functions circuits is given next.

1. 3. 1. Linear Conversion Circuits

Different circuits with linear current-to-voltage or voltage-to-current conversion are proposed in literature [3-5, 13, 14]. Linear current-to-voltage conversion occurs through the transresistance amplifier consisting of an operational amplifier and a feedback resistor [3-5] (Fig. 1-7). Resistance R is replaced by an active resistance in [3]. For proper linear conversion, a source follower is included as an output stage [5]. Good accuracy for a wide range can be achieved but with poor programmability and relatively high power consumption.

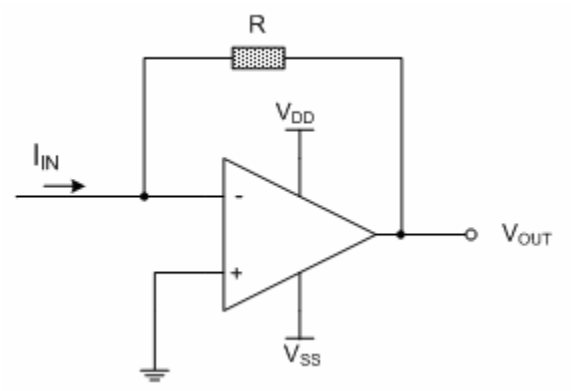


Figure 1-7 Linear current-to-voltage conversion.

In [13] ideal diodes are obtained by using commercially available diodes interconnected with OTA's (Fig. 1-8). Using several ideal diodes connected in parallel blocks, with different configurations, allows the generation of a piecewise-linear function of any shape (TBF or SatLin). The number of OTA's blocks in the circuit is equal to the number of segments of the implemented function. The slopes of the transfer characteristics can be modified by changing the OTA's transconductances.

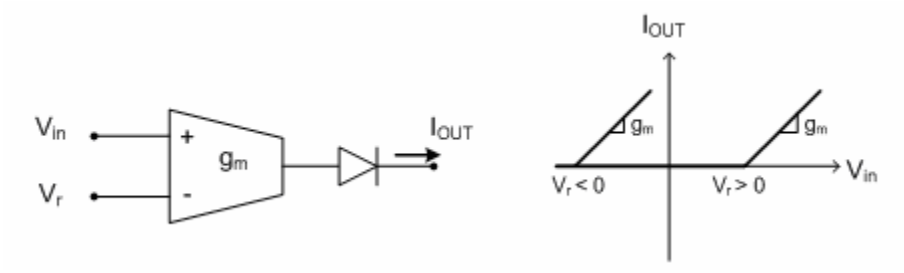


Figure 1-8 Linear voltage-to-current conversion [13].

In [14] a SatLin function circuit is proposed. It converts the voltage to a current using a modified version of the differential amplifier. This differential amplifier, depicted in Fig. 1-9, is provided with both positive and negative feedback (Fig. 1-9).

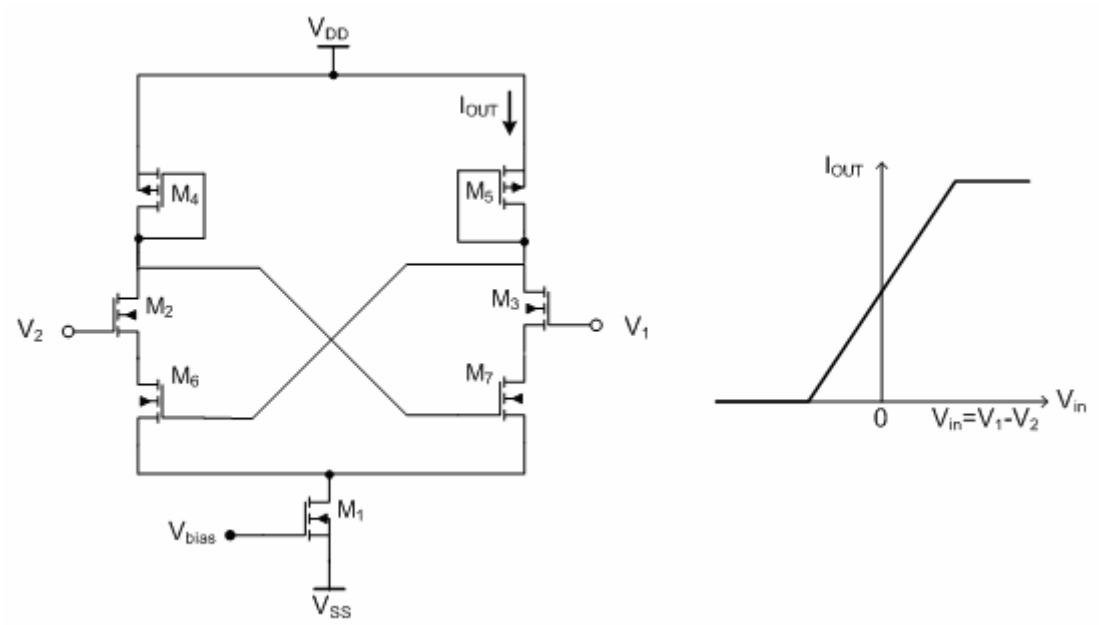


Figure 1-9 Linear voltage-to-current conversion [14].

In more detail M_6 and M_7, which are designed to work in the linear region, provide a negative series-series feedback for the differential amplifier, whose effect is the widening of the linear portion of the transconductance gain. Moreover, when M_2 (M_3)

enters the cutoff region, the drain voltage of M_2 (M_3) increases, and in turn this causes a decrease of the drain-source resistance of M_7 (M_6). The global result is a positive feedback effect that sharpens the corners of the transconductance gain. The input voltage can be applied to one input of the differential amplifier, while the second input will decide the threshold of the transfer characteristic, depending on the voltage applied to it. Changing the value of V_{BIAS} will affect the saturation level and the slope of the linear portion of the transfer function at the same time.

1. 3. 2. Current Mode Piecewise-linear functions circuits

Arbitrary function synthesis circuits: Different current mode function synthesizing circuits are proposed in literature [15-18]. The technique in [15] is based on the utilization of current mirrors as basic building blocks. As in [13], a number of building blocks equal to the number of segments of the PWL function is used. But here it is important to have a replica of the input current for each block, while in [13] the input voltage is common for all blocks. This approach is improved in [16] by utilizing current mirrors with different gain values and a diode at the output of every block for switching the current at the break points. Only one reference current is required. The networks are temperature and technology insensitive. Although a good accuracy can be achieved using the approaches proposed in [15,16], both approaches are poor in regard to programmability of the realized function parameters.

Another similar approach for implementing an arbitrary piece-wise linear (PWL) characteristic, with better programmability feature, is presented in [17]. Each of the breaking points and each slope is separately controllable (Fig. 1-10). Transistors are intended to operate in the subthreshold region. This way, the slope m_i of the gain segments can be controlled through the bias voltages V_{ia} and V_{ib} (transistors sources voltages). In particular,

$$m_i = e^{\left(\frac{V_{ia} - V_{ib}}{nU_T}\right)} \tag{1-7}$$

Where n is the slope factor of the MOS transistors, and U_T is the thermal voltage. As such, every block can be considered as a programmable gain current mirror. Almost all piecewise-linear functions can be implemented using this technique but with a small current values.

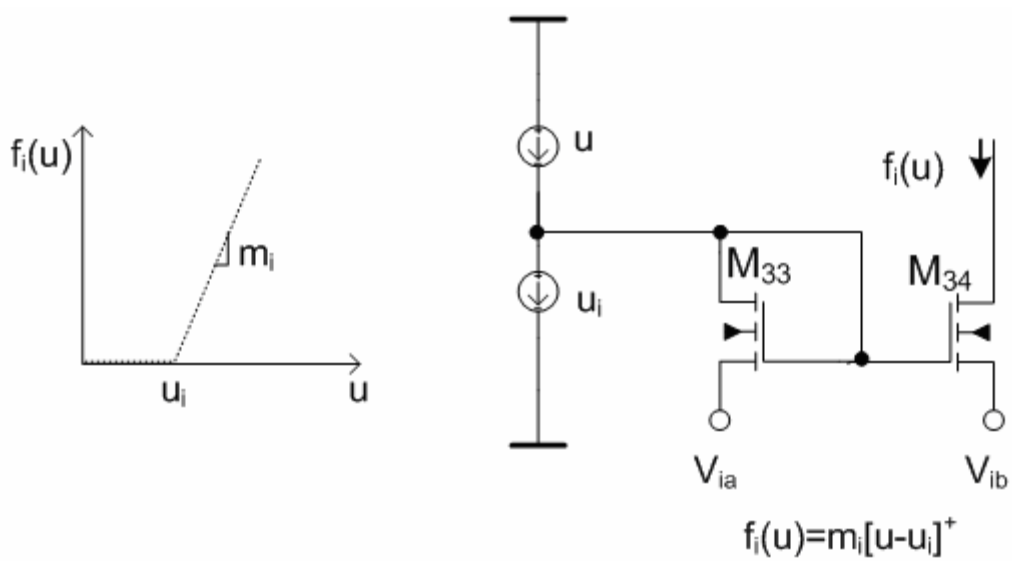


Figure 1-10 One-sided rectified function

In [18] a current-mode subthreshold CMOS PWL synapse circuit with efficient architecture is developed. It suggested that every PWL function can be represented by a polynomial. For example a TBF can be represented by:

$$y(x)=0.5\,|x + 2| - |x| - 0.5\,|x - 2| - 1 \qquad (1\text{-}8)$$

To implement this polynomial in the circuit level we need the following operations:

- Algebraic summation;
- Scalar;
- Absolute value.

Algebraic summation is achieved by current summation at a common node. A standard four transistor multiplier cell and a simple translinear cell for the absolute value are introduced. Different PWL functions can be implemented by the same analogy. The programmability feature of this circuit is poor.

SatLinear function circuits: In [6] a neuron circuit for implementing the SatLinear function is presented. The circuit consists of three PMOS bias transistors controlled by V_{BIAS}, and four NMOS transistors used as current mirrors. Its dimensions and the value of V_{BIAS} can be used to control the slope and the saturation levels, correspondingly. So, for this circuit the saturation levels are programmable.

In [19, 20] neuron circuits are based on CMOS implementation of the class AB current conveyor. In [19] additional differential amplifier is incorporated into the signal path to accomplish the transconductance programmability (slope programmability). In [20] the output currents are limited by the current I_{LIM}, it results

in limiting the output current at the same level. So the saturation levels for this circuit are programmable (by programming I_{LIM}). A relatively good accuracy of the implemented function can be achieved.

TBF circuits: The circuit proposed in [21] for implementing the TBF is presented here (Fig. 1-11). In this circuit, a class AB input stage works as a current rectifier. It provides a simple way to implement a programmable TBF function with good accuracy. I_{OUT} corresponds to a base function for positive interpolation data. Slopes of the base function are given by current mirror gains, while the position and height of the breakpoint are provided by the DC biasing currents I_{DFT} and I_{REF}, respectively.

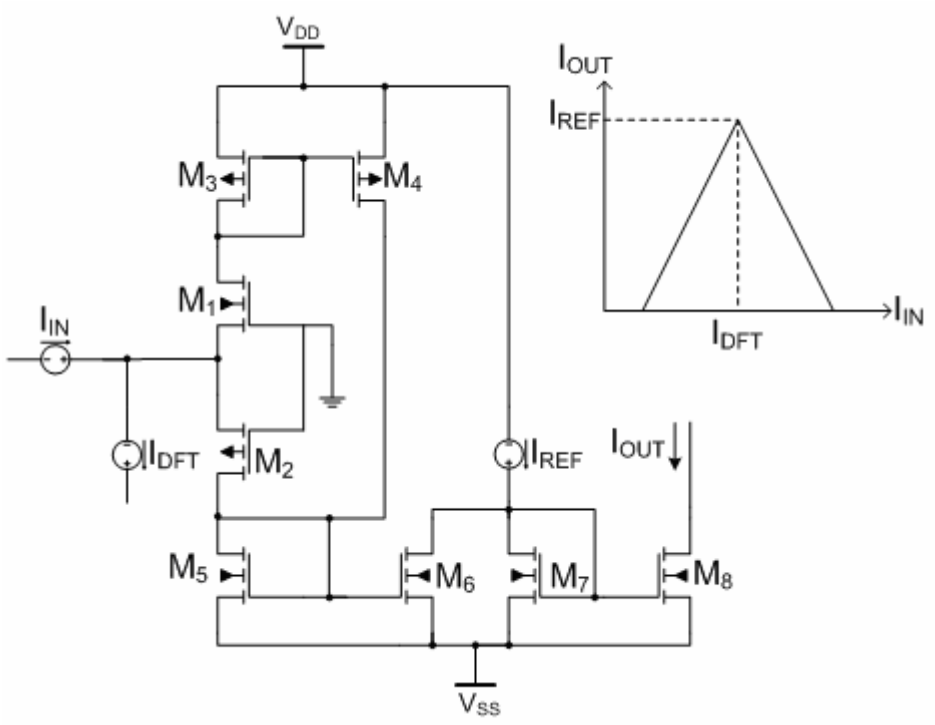


Figure 1-11 TBF implementation.

A current-mode weak-inversion circuit for analog function approximation is proposed in [22,23]. It is based on the fuzzy logic paradigm. Programming is easily achieved by sampling the target function in a grid and by encoding the sampled values in a set of digital words. The larger the number of samples, the better is the approximation. The TBF can be easily obtained using this approach but the area consumed by the circuit is relatively large.

1. 3. 3. Sigmoid function circuits

In [24] a neuron circuit is designed to differentially sum its input currents using simple transconductance amplifiers. The output voltage is proportional to the input current. To achieve the sigmoidally-shaped (S-shaped) response characteristic, two diodes are used in the feedback bath of the final stage. Programmability and accuracy are poor in this circuit. More elegant and practical way is considered in [3,4,25] where the sigmoid function is approximately realized by cascading simple inverting amplifiers with input and feedback resistors (Fig. 1-12). Gain programmability is achieved by replacing the passive resistors by voltage-controlled MOS resistances.

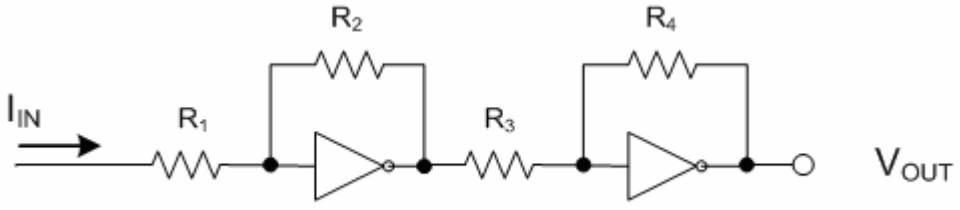


Figure 1-12 Sigmoid function circuit [3, 4, 25].

Another example of implementing the sigmoid function is proposed in [26]. The circuit uses master/slave tuning loops (Fig. 1-13), one to adjust the slope of the middle region (by varying V_{bias}), and the other to adjust the saturation level (by varying V_R). The advantages of this circuit are the good programmability and the differential treatment of the input and output signals. The main disadvantage is the moderate accuracy (the RRMS error is expected to be greater than 5%).

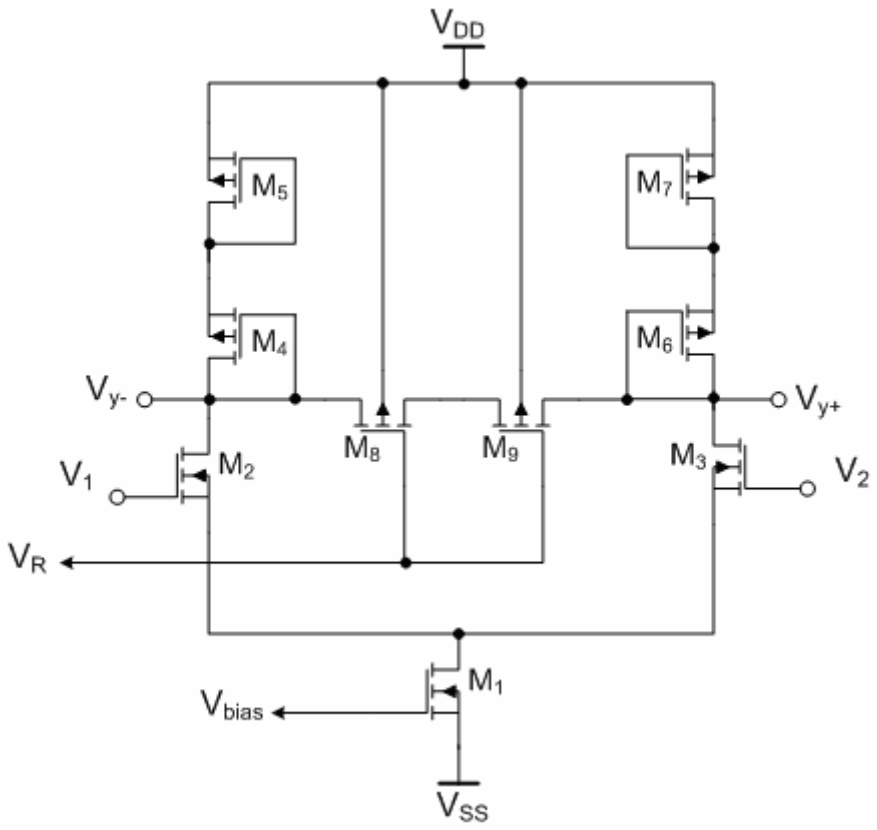


Figure 1-13 Sigmoid function implementation [26].

An I – to – V Sigmoid neuron, based on a nonlinear load (Fig 1-14a), is presented in [9]. Both lumped and distributed implementations were discussed. Nonlinear characteristic of this neuron circuit is a combination of two quadratic curves (from M_1 and M_2) and a linear transition part corresponding to R. To realize the resistor's task, some designs rely on existing parasitic/leakage impedances while the others may use a MOS device. In [27] a modified circuit with 4 MOS devices (Fig 1-14b) is presented that approximates an S-shape neural function with 4 quadratic characteristics. Two transistors are replacing the passive resistance of the circuit reported in [9] with a lightly S-shaped characteristic in the region where both transistors are off.

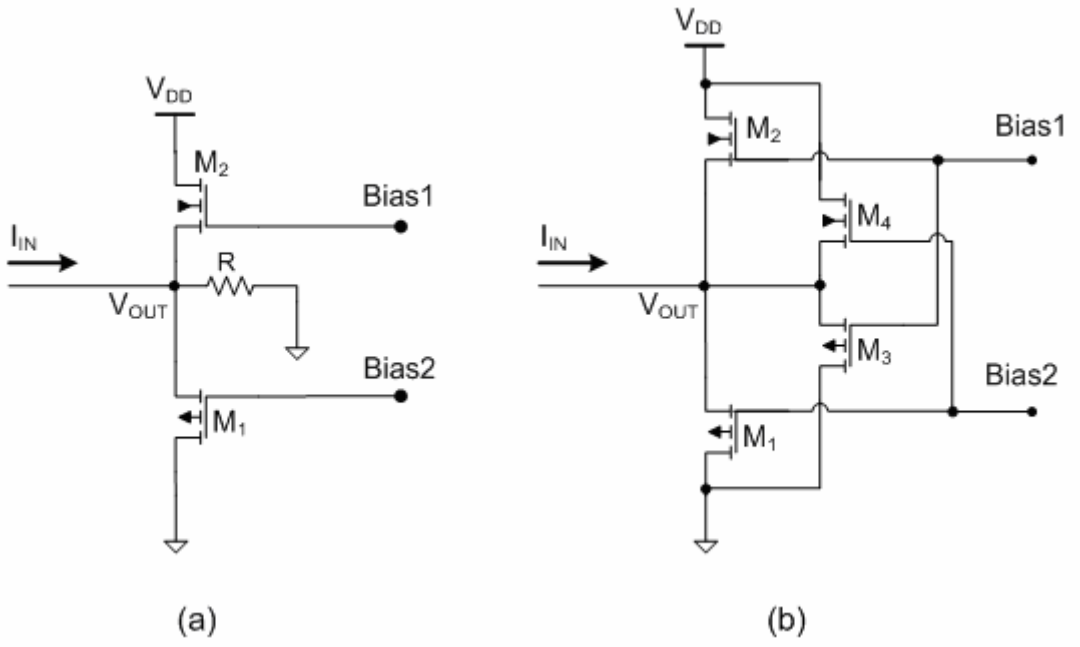


Figure 1-14 a) Sigmoid neuron [9] b) Sigmoid neuron with MOS resistor [27].

In [28] a slight modification has been done to the circuit proposed in [27]. The proposed circuit (Fig. 1-15) has an advantage, over the circuits depicted in (Fig. 1-14), of being able to treat the input and output signals differentially. The circuit is biased in weak inversion. The relation between the normalized output and input variables is $\tanh(\sinh^{-1}(x))$ which is very close to the Sigmoid function.

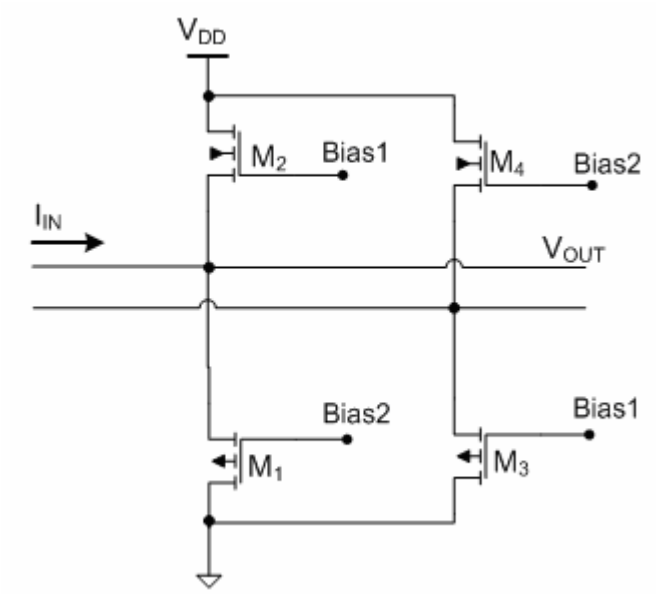


Figure 1-15 Sigmoid neuron with differential input-output [28].

The main advantage of the circuits proposed in [9, 27, 28] is the simplicity of the design, while the main disadvantage is the poor programmability.

In [29], using an optimization process, a differential (V – to – I) pair circuit with active source degeneration has been realized (Fig. 1-16). Activation functions with different gain can be realized by changing Vc. The achieved accuracy and programmability are relatively good.

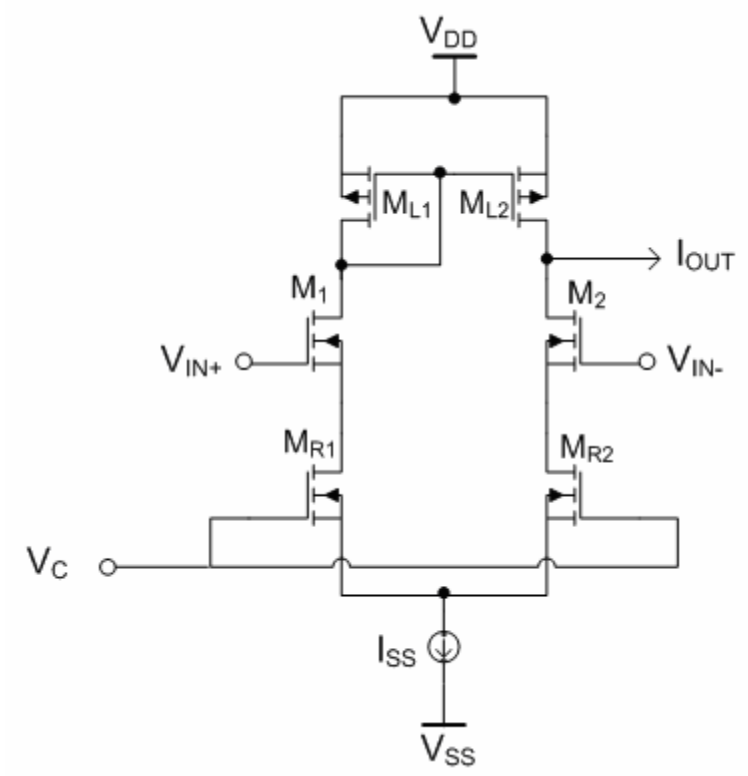


Figure 1-16 Sigmoid function implementation [29].

In [6] a current mode implementation of a sigmoid circuit, operating in a wide range of input current, is presented (Fig. 1-17). Two transistors (M_4 and M_5) form a current-voltage converter. The resulting voltage is applied to a transconductance amplifier formed of transistors M_6 to M_{11}. A Sigmoidal transfer characteristic between the input and the output currents is achieved.

A reconfigurable low-voltage low-power building block that can either function as a *synapse* or a *neuron* is proposed and analyzed in [30]. The approach of the Sigmoid function implementation is similar to the one proposed in [6]. Both circuits can

achieve a good accuracy but with poor programmability (slope and saturation levels cannot be programmed separately).

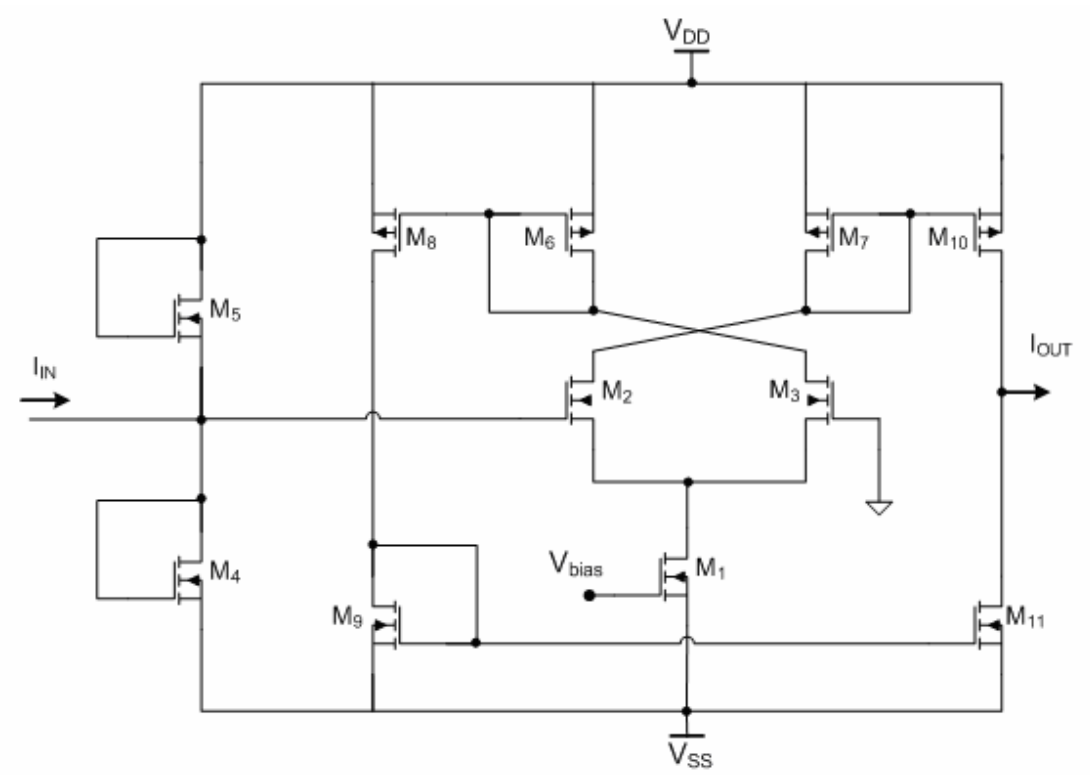


Figure 1-17 Current-mode Sigmoid function implementation [6].

1. 3. 4. Sigmoid function and its derivative circuits

The derivative of the sigmoid function is required to calculate the error through the on-chip learning process of the NN. Most circuits described in the literature, implementing the derivative, are based on two approaches. First approach is based on the fact that for the sigmoid function $f(x)$ the first derivative can be written as:

$$f'(x) = f(x)(1 - f(x)) \tag{1-8}$$

The second approach is based on the use of the approximate definition of the derivative.

$$f'(x) \cong \frac{f(x+\Delta) - f(x-\Delta)}{2\Delta} \tag{1-9}$$

In [31] the sigmoid function is realized by using a simple MOS differential pair. Gain control is added by using a voltage controlled MOS resistance. The output is fed to the input terminals of a second circuit which implements the differential of the Sigmoid. This circuit configuration is normally used as a maximum value detector. It approximates the differentiation of the sigmoid function, within 5-10% error. In [32] another realization is presented. The circuit is built from two identical differential pairs. It is capable of generating both the Sigmoid transfer function (when one differential pair is active) and its approximate derivative (when the outputs of the two active differential pairs are crosscoupled which results in an output current approximating the derivative of the Sigmoid (see eq. 1-9)). The programmability feature of this circuit is poor, and the accuracy is moderate.

Before proceeding, it is useful to present an important circuit [38] which constitutes the basis or the idea for many circuits implementing the derivative of the sigmoid or the RBF. The circuit, called a Correlator or a Bump circuit, comprises a simple differential pair (Fig. 1-18) and two additional transistors. It is biased in weak inversion region. The output current is zero whenever I_1 or I_2 is zero, and maximum whenever they are equal. For simulation, V_2 is grounded and V_1 is varied between

-110 mV and +110 mV. As can be seen from figure 1-19, I_1 approximates the sigmoid, while I_{OUT} can approximate the derivative of the sigmoid. For a proper choice of the biasing conditions and transistors dimensions, I_{OUT} can approximate the RBF. The main disadvantages of this circuit is that the derivative is not ideally symmetrical, and the programmability is poor. The same idea is exploited and slightly improved in [33,34,35]. In [34] a positive and negative interpolation data is achieved by a little modification in current mirroring.

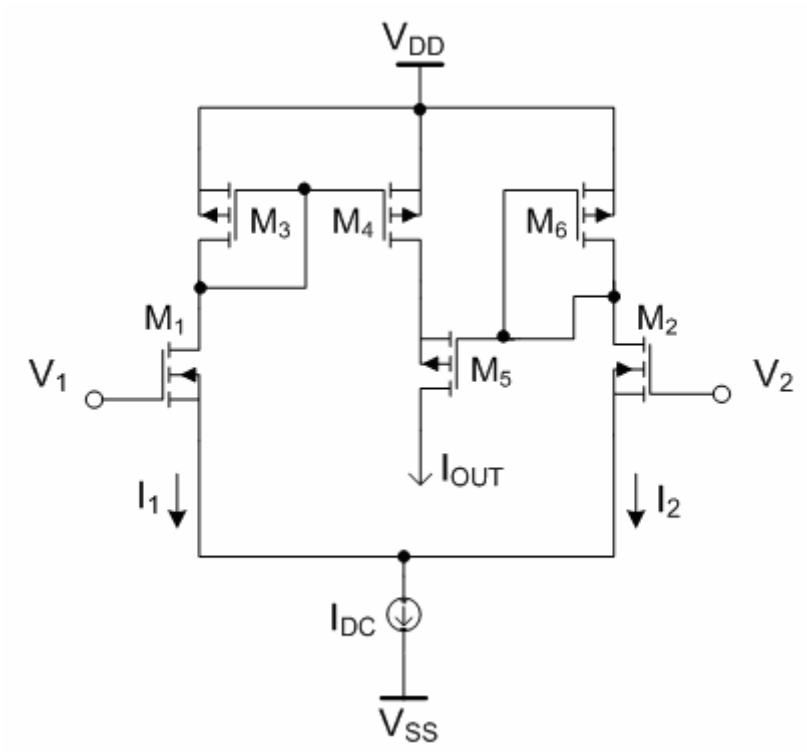


Figure 1-18 The Correlator or Bump circuit [38].

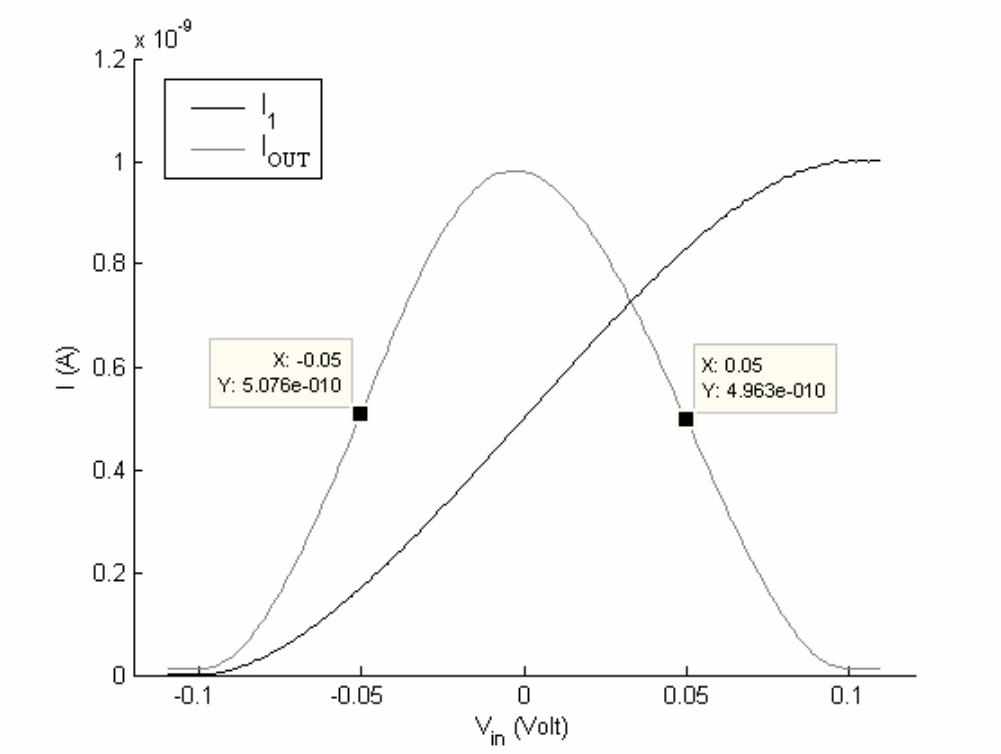


Figure 1-19 The simulation of the Correlator (or Bump) circuit [38].

In [35] to construct the derivative, the fact was exploited that outside the middle region the response of the nonlinear function $f(x)$ is either almost maximum or almost minimum. In these two regions, the derivative $f(x)(1-f(x))$ can then be approximated by the minimum of $f(x)$ and $1-f(x)$. The circuit proposed in [36] and improved in [37] uses also a simple differential pair for realizing the sigmoid function, while the derivative is achieved by exploiting the approximate definition of the derivative (see eq. 1-9). Sigmoid function with different gain factors and thresholds can be achieved leading to a good programmability.

1. 3. 5. Radial Basis function Circuits

In neural networks, backward propagation training based on Gaussian functions, or the radial basis function (RBF) networks, is more efficient than those based on sigmoid functions in the hidden layers. It is up to two or three times faster [1, 2]. And the number of interconnections is reduced, which translates into simpler VLSI implementations. Possible realizations of the RBF (Gaussian function) are reported in [38 - 48]. The circuit proposed in [39] implements the RBF using squaring element for the input and an exponential function device for the output (see eq. 1-5). The synaptic network computes the square function of the difference between the input and the weight voltages (see eq. 1-6). The synapse cell current summation and exponential nonlinearity can be achieved using a neuron which consists of a current-to-voltage converter and an exponential function circuit. This technique is again exploited in [40]. An alternative choice in [39] is to directly implement the RBF using a source degenerated differential amplifier with correlating transistors (see the Correlator circuit in Fig 1-18). This circuit enjoys a variable-height and width, and the transistors are designed to work in weak inversion. The accuracy of both approaches proposed in [39] is moderate. Different and simple approach is used in [41]. A change in a voltage applied to the gates of two connected-in-series complementary transistors will result in change in the current (flowing through them) with Gaussian-like shape. The circuit is biased to operate mostly in the subthreshold region. Accuracy and programmability are poor. Very similar approach to the one

proposed in [39] is used in [42] but with different implementation. The MOS transistors operate in strong inversion. The accuracy is improved and the Variance (Standard Deviation) is programmable. The proposed Gaussian synapses in [43] resemble the exponential nonlinearity by using piecewise approximation in the current-voltage space. The circuit consists of MOS differential pairs and several arithmetic computational units in the current-mode configuration. An improved approximation can be achieved due to symmetric handling of the signal. This circuit, operating in strong inversion, approximates the Gaussian function with error less than 2 % over the input voltage range of ± 3 V in the ideal case. The height and the center of the function are programmable. The proposed circuit in [44] is an improvement of the one in [43]. It consists of two differential transconductance amplifiers and a pair of current mirrors. The gate terminal inputs are connected in such a way that if one of the differential transconductance amplifiers is contributing a current to the output current, the other will not contribute any current to it. So the output current can be positive or negative. This synapse has several advantages over that reported in [43]. The standard deviation (width) can be changed externally by changing external biasing voltages. This Gaussian synapse does not need any differential inputs and it also uses less number of transistors. But the accuracy of this is not better than the accuracy achieved in [43]. A symmetric Gaussian function circuit is proposed in [45] based on the correlator proposed in [38]. This circuit has symmetric I-V characteristics when working in strong or in weak inversion regions.

Only center and peak amplitude (height) can be tuned. This circuit is improved in [46] to achieve the width tunability. This is achieved by using outer and inner differential pairs, to be able to tune the width without affecting the other parameters.

In [47], a new CMOS analog programmable Gaussian function, based on current mode operations and using devices operating in the subthreshold region is proposed. The function is described by an input-output relationship given by:

$$I_{out} = I_{\max} \exp\left(\frac{(I_{in} - I_u)^2}{p} \right) \tag{1-10}$$

where p is the slope coefficient. According to this equation, the first block in the design is a current subtractor followed by a programmable current squarer. The second block is an exponentiator of the obtained current. We can use the exponential dependency of the drain current on the source-gate and the back-gate voltages of a MOS transistor operating in weak inversion. Thus, a transimpedance amplifier should be inserted before the exponentiator block. The Gaussian centre can be controlled by varying the input reference current I_u. The Variance can also be controlled. In [48], a hardware implementation of the Beta neuron operating in subthreshold mode is presented. Beta Basis Function Neural Networks (BBFNNs), which can be considered as a generalization of RBFNNs, are proposed for their powerful learning and universal approximation characteristics. The Beta function is defined by:

$$\beta(x) = \left(\frac{x - x_0}{x_c - x_0}\right)^p \left(\frac{x_1 - x}{x_1 - x_c}\right)^q \tag{1-11}$$

where p>0, q>0, x_0 and x_1 are real parameters, and $x_c=(px_1+qx_0)/(p+q)$. The proposed realization is composed mainly of three stages. The first stage completes the logarithmic current-to-voltage converters. The second stage executes the subtraction and the multiplication by a coefficient. The third stage realizes the exponential voltage-to-current converter. At the end we obtain an output current that has a Beta function form. The programmability of this circuit is poor.

1. 3. 6. Summary and Conclusions

Following are tables summarizing the main features of the available circuits, for example the signal conversion type, operation mode of the circuit (strong or weak inversion), accuracy of the implemented function and programmable parameters of the implemented function. Unfortunately, for most of the previewed papers, there was no explicit mentioning about the accuracy of the implemented functions. The error of such implementations will be estimated here approximately, depending on the information mentioned in these papers, and our knowledge and expertise in this subject. For poor accuracy, the expected error is greater than 10%. For moderate accuracy, the expected error is between 5% and 10%. For good accuracy, the expected error is less than 5%. Poor programmability means a limited ability to change only one parameter or the ability to change two parameters, jointly.

Table 1-1 Summary of the piecewise-linear functions circuits.

References	Implemented Function	Conversion Type	Circuit Mode	Accuracy	Programmability
[3,4,5]	SatLins	I – to – V	Strong Inversion	----------	No
[13]	Synthesis	V – to – I	Strong Inversion	----------	Slope, BreakPoints
[14]	SatLin	V – to – I	Strong Inversion	Good	Threshold
[15]	Synthesis	I – to – I	Strong Inversion	Good	BreakPoints
[16]	Synthesis	I – to – I	Strong Inversion	Moderate	BreakPoints
[17]	Synthesis	I – to – I	Weak Inversion	Moderate	Slope, BreakPoints
[18]	Synthesis	I – to – I	Weak Inversion	Moderate	poor
[6]	SatLins	I – to – I	Strong Inversion	----------	Saturation Levels
[19]	SatLins	I – to – I	Strong Inversion	Good	Slope
[20]	SatLins	I – to – I	Strong Inversion	Good	Saturation Levels
[21]	TBF	I – to – I	Strong Inversion	Good	Threshold, peak
[22,23]	TBF	I – to – I	Weak Inversion	Moderate	Slope, Threshold

From Table 1-1 we may conclude that:

- Current mode circuits are the majority between the piecewise-linear circuits, and they are more accurate.
- There are four circuits which can synthesize different piecewise-linear functions. Two of them are working in weak inversion region. But only one of the four circuits can have programmable slope [17].

- Circuits operating in strong inversion are more accurate in general.

Table 1-2 Summary of the sigmoid function circuits.

References	Conversion Type	Operation Mode	Accuracy	Programmability
[24]	I – to – V	Strong Inversion	Poor	poor
[3,4,25]	V – to – V	Strong Inversion	Moderate	Slope(Gain)
[26]	V – to – V	Strong Inversion	Moderate	Slope, Saturation levels
[9,27]	I – to – V	Strong Inversion	poor	Poor
[28]	I – to – V	Weak Inversion	Moderate	Poor
[29]	V – to – I	Strong Inversion	Moderate	Slope
[6]	I – to – I	Strong Inversion	Moderate	poor
[30]	I – to – I	Strong Inversion	Good	Slope
[31]	V – to – V	Strong Inversion	Moderate	Slope, Threshold
[32]	V – to – I	Strong Inversion	Moderate	poor
[33,34]	V – to – I	Weak Inversion	Moderate	poor
[35]	V – to – I	Weak Inversion	Good	poor
[36,37]	I – to – V	Strong Inversion	Good (3% Error)	Threshold, Slope

From Table 1-2 we may conclude that:

- On the contrary with piecewise linear circuits, no current mode sigmoid function circuits are reported. The circuits proposed in [6, 30] are not pure current mode circuits, as the input current is linearly converted to a voltage.
- Although a mathematical expression approximating the Sigmoidal function has been found for circuits operating in weak inversion [33-35], still the accuracy level is better for circuits operating in strong inversion.
- Better programmability is achieved for circuits operating also in strong inversion.

Table 1-3 Summary of the Radial Basis function circuits.

References	Conversion Type	Operation Mode	Accuracy	Programmability
[39]	V – to – I	Weak Inversion	Moderate	Width (good range)
[39]	V – to – I	Weak Inversion	Moderate	Height, width
[41]	V – to – I	Weak Inversion	Poor	No
[42]	V – to – V	Strong Inversion	Good	Width
[43]	V – to – I	Strong Inversion	Very good (2% Error)	Height
[44]	V – to – I	Strong Inversion	Good	Height, width
[45]	V – to – I	Weak Inversion	Good	Height
[46]	V – to – I	Weak Inversion	Good	Height, width
[47]	I – to – I	Weak Inversion	Good	Width
[48]	I – to – I	Weak Inversion	Moderate	No

From Table 1-3 we may conclude that:

- Best accuracy is achieved for strong inversion mode [43, 44].
- No current mode Gaussian function circuits were reported. Circuits [47, 48] are using relatively large number of mathematical blocks to approximate the Gaussian relation between input and output currents which make them not practical for VLSI design.

1. 4. Motivation

The summary of the literature review reveals some important facts about the reported circuits:

- No generic circuit has been reported which can be configured to implement two or more NN functions.
- No circuits were reported which could achieve full programmability feature for TBF and SatLin functions.
- No current mode circuits have been reported for Sigmoid or Gaussian function implementation.

A generic circuit, able to implement the main NN functions, can be very useful in different applications especially for reconfigurable NNs which usually may include more than one nonlinear function in the same structure.

In most practical analog VLSI neural networks, flexible tuning of function parameters is a must due to the inherent learning/adaptive characteristics of neural networks. So it is very important to keep or to improve the programmability feature of the functions.

If such a circuit can be implemented using current mode technique then we can benefit from its good advantages. Besides being more suitable for piecewise-linear functions implementation it can provide improved dynamic range, speed performance (frequency range) and more practical way for doing simple mathematical calculations (current summation or subtraction).

CHAPTER 2

Problem Formulation and Proposed Circuit

2. 1. Problem Formulation and Proposed Approach

Our goal in this thesis is to develop a generic circuit which can be configured to implement four different nonlinear functions: Two piecewise-linear functions (TBF and SatLin) and two nonlinear functions (RBF and Sigmoid). The complexity of this problem comes from the fact that the nature of the output is either linear for specific range or nonlinear, depending on the choice of the function. In general, in analog electronic circuits to add nonlinearity to a linear circuit is much easier than linearizing a nonlinear circuit. So, it is easier to start by implementing the piecewise linear functions. From these functions we can implement the others by adding a source of nonlinearity to the circuit.

CMOS technology is our choice for the implementation, as it is the best in terms of lower power consumption, higher integration density, and compatibility with existing digital technologies. All these advantages make this technology the most suitable for VLSI and ULSI design.

From the summary of the literature review in chapter 1 we may conclude two important points:

- Current mode circuits or circuits with input-output current representation are more suitable for piecewise-linear functions implementation.
- Better accuracy and good programmability were achieved with circuits using MOSFETs operating in strong inversion.

After considering all these inferences it appears that the best approach for the generic circuit design is to start with a piecewise-linear circuit operating in strong inversion which has current input-output signal representation. A TBF circuit is preferable as it is easier to achieve a nonlinear response from a piecewise-linear one. Then, from these two functions we can achieve the others (Sigmoid and SatLin).

The order of the generic circuit implementation will be as follows:

- The design will start by choosing from the literature a suitable TBF circuit to be the basis of the proposed circuit design.
- An improved version of this circuit will be proposed.
- A technique will be proposed to extract the RBF from the TBF.
- A technique will be proposed to extract Sigmoid and SatLin functions from RBF and TBF functions.

A special attention will be paid for the programmability of the proposed circuit.

2. 2. TBF Implementation

After studying different circuits for this approach, it was concluded that the TBF circuit proposed in [34] is the most suitable as a starting point for the design of the proposed generic circuit. An explanation of this circuit (Fig 2-1) will follow.

2. 2. 1. Circuit Description

This circuit uses a current switch input stage. Positive input currents are drawn to the bottom device, while negative currents are drawn to the top device. These currents will finally flow through M_5, thus decreasing the current flowing through M_7 and consequently decreasing I_{out} . If I_{IN} equals to I_{DFT}, then no current will be drawn to the input stage and no current will be mirrored to M6. As such, I_{out} will be maximum and equal to I_{REF}. When I_{IN} is greater than I_{DFT} a current will be flowing through M_2 and mirrored by M_5 to M_6. This current will be drawn from the current source I_{REF} as such decreasing $I_{out.}$ When I_{IN} is less than I_{DFT} a current difference will be flowing through M_1, mirrored by M_3 to M_4 and finally mirrored by M_5 to M_6. This current will cause I_{out} again to be decreased. The increase in the difference current will eventually cause I_{out} to become zero when $I(M_5)$ is equal or greater than $I_{REF.}$

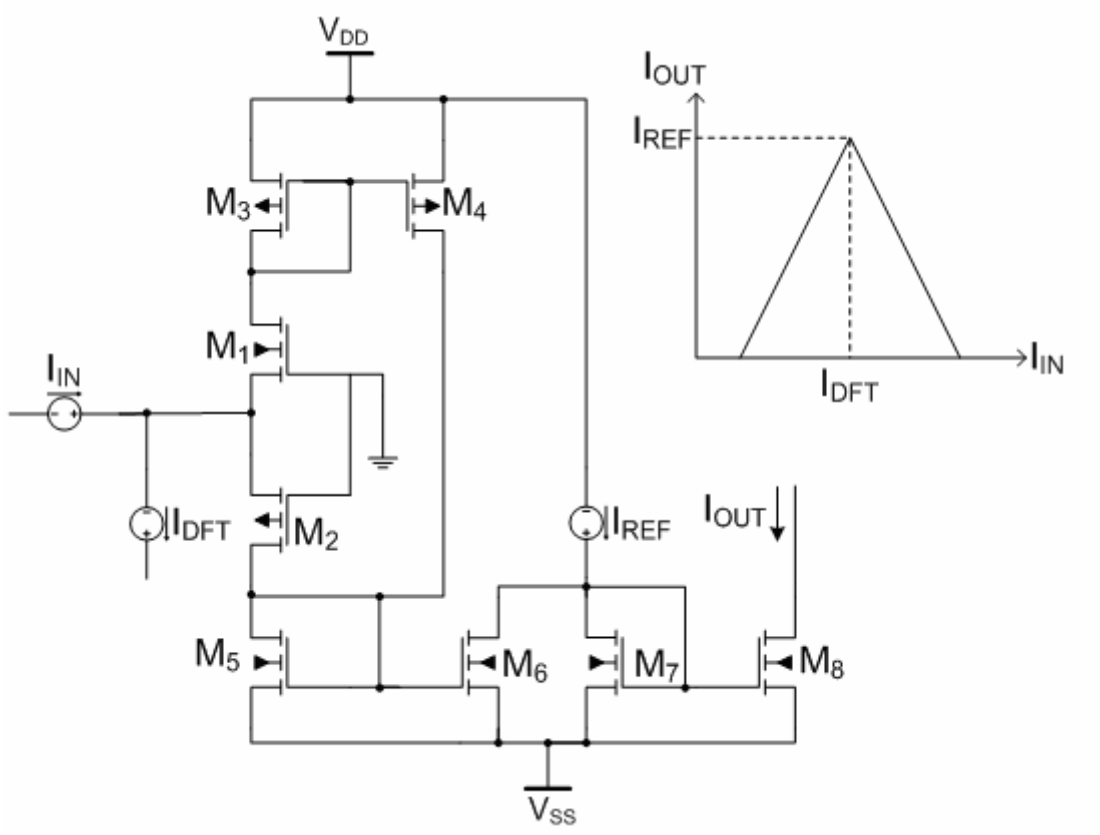


Figure 2-1 TBF implementation with positive interpolation data [34].

Slopes of the TBF are given by current mirrors gains, while the position and height of the break point are controlled by the dc biasing currents I_{DFT} and I_{REF}, respectively. I_{DFT} can be positive or negative, and so is the position of the break point (the peak). Negative interpolation data can be realized by interchanging the functions of up and bottom transistors (M_6 should be connected to M_3 and M_4 should be connected to a PMOS output current mirror).

Because of the intrinsic class AB operation, this structure allows greater operation speed compared to the others. Lower power consumption in the quiescent state is also expected. Nonetheless, it has some disadvantages like the effect of the second order effects (channel length modulation) on the accuracy of the function and also the insufficient programmability of the circuit (the slope is not programmable).

2. 2. 2. TBF Circuit Improvement

An improved version of the TBF circuit is proposed and shown in Fig. 2-2. The accuracy of this circuit is improved by using low voltage Cascode Current Mirrors [51]. Digitally controlled combination of parallel connected current mirrors, with different dimension ratios, now controls the slope of the function. By switching D_N ON or OFF we can choose between different word combinations and consequently different mirroring gain values. Five current mirrors with gains equal to 0.25, 0.5, 1, 2, and 4 are used.

2. 2. 3. Simulation results and Discussion

A voltage supply equal to $\pm$ 1 V will be used for all simulations. TSMCT16X Spice BSIM3 Version 3.1 Model with feature size 0.18 microns will be used for setting the parameters of the transistors (see appendix A) . All transistors bulk terminals are assumed to be connected to either V_{DD} (PMOS transistors) or V_{SS} (NMOS transistors). V_{B1} and V_{B2} of Fig. 2-2 will be set to zero potential. D_3 is set to zero

potential for current mirroring gain equal to one. The rest of the digital word D_n should be set to V_{SS}. Orcade Cadence 9.2 program will be used for performing the simulations. These simulation parameters are also valid for circuits simulated in the next chapters, unless otherwise is mentioned.

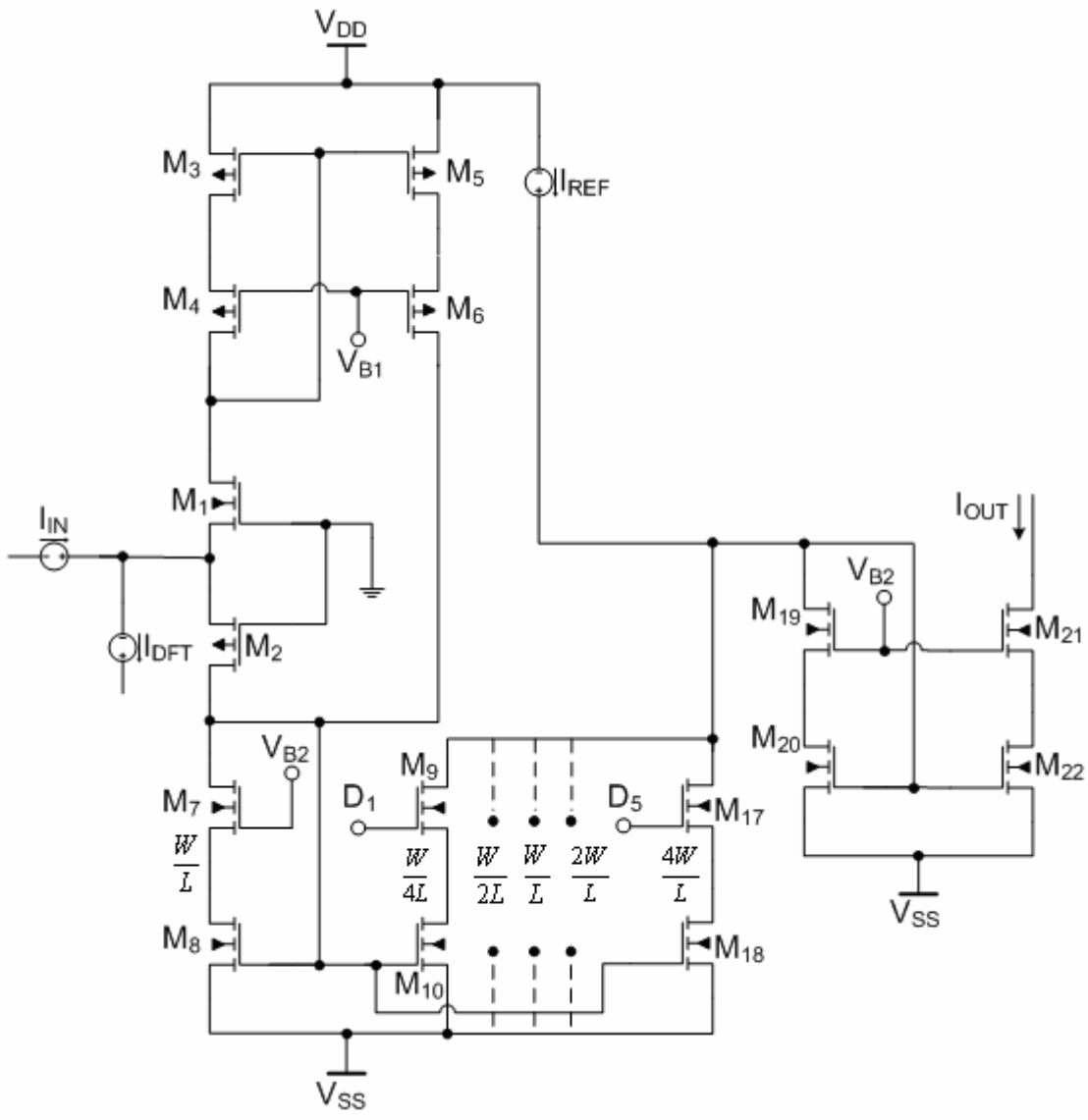
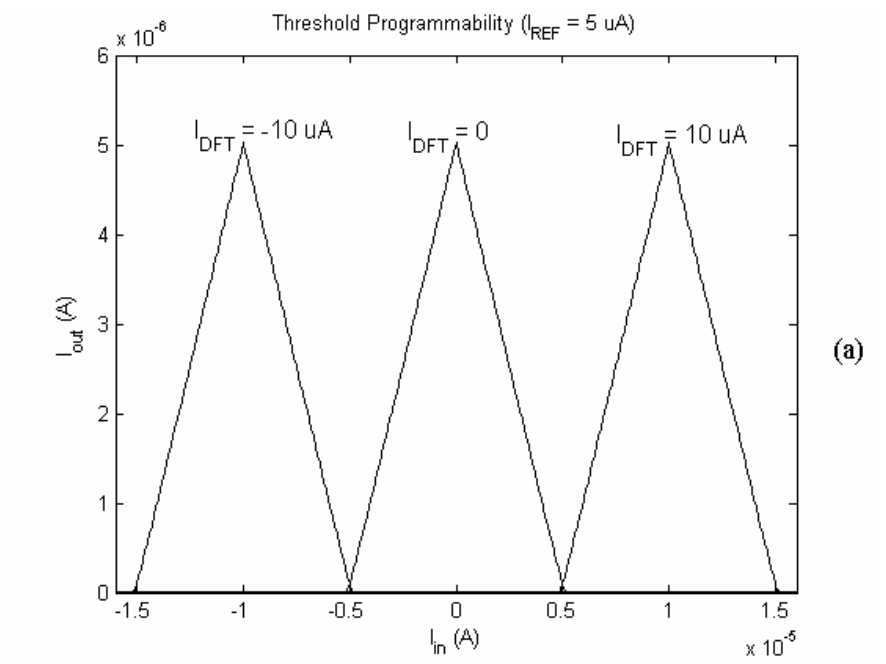
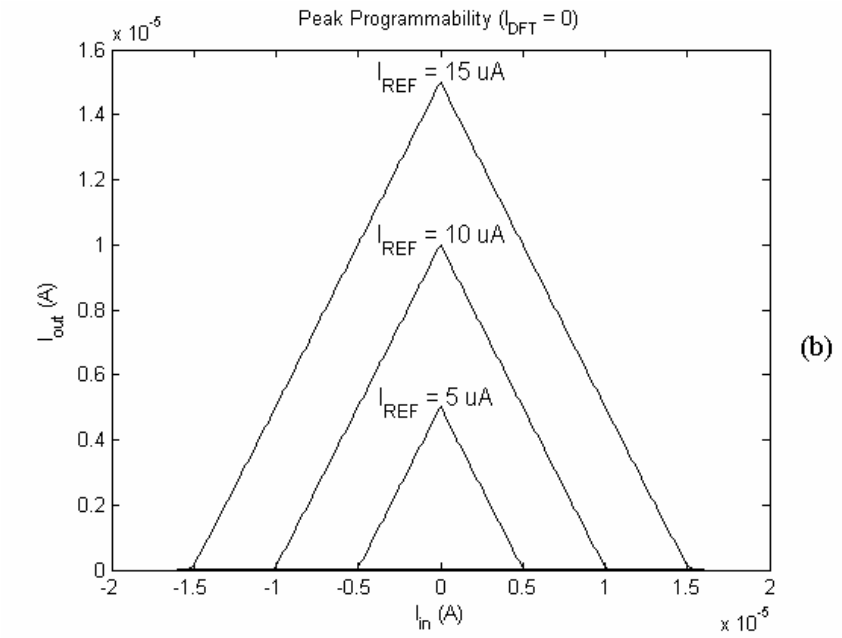


Figure 2-2 TBF implementation with improvement accuracy and programmable slope.

Figures 2-3 and 2-4 illustrate the programmability feature of the circuit in Fig. 2-2.

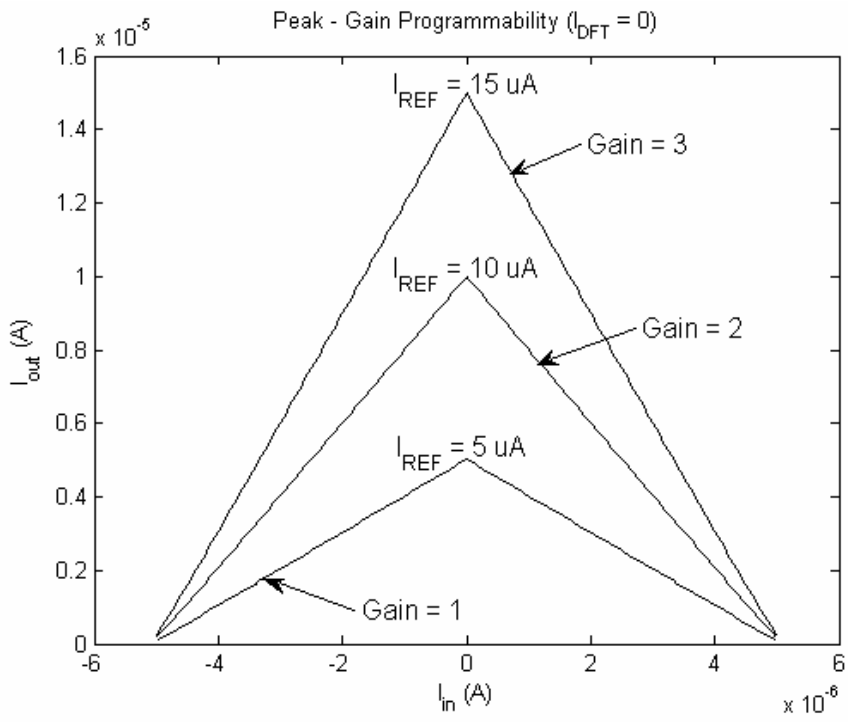


(a)



(b)

Figure 2-3 TBF programmability (a) threshold and (b) peak .



(a)

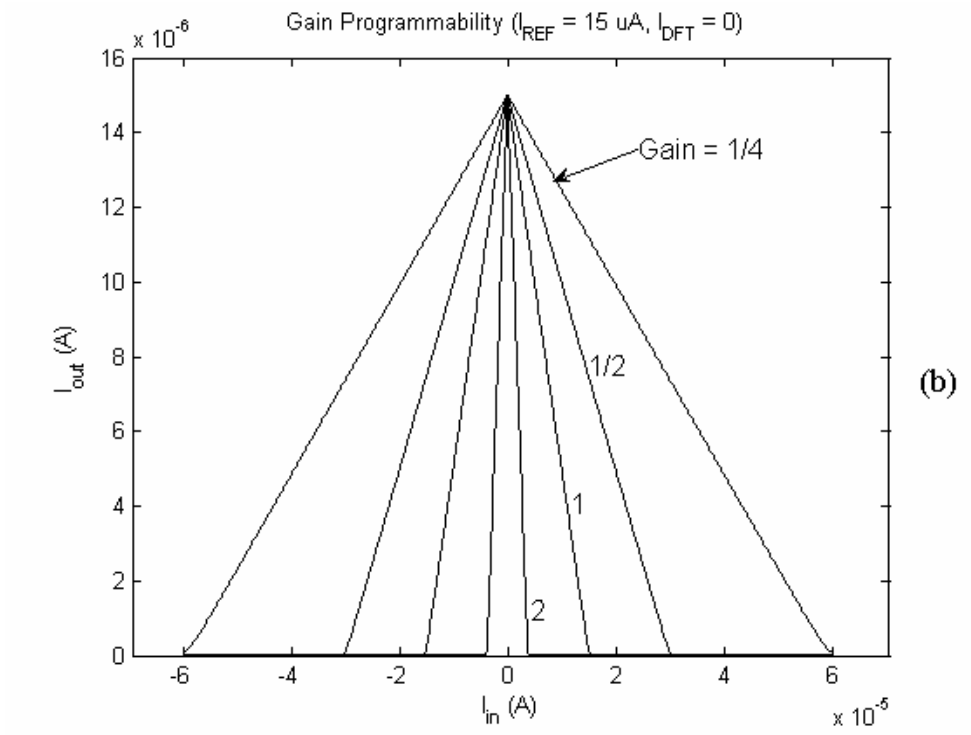


Figure 2-4 TBF programmability (a) peak-gain (b) gain.

The gain (slope) can be chosen between 0 and 7.75 with minimum step equal 0.25. Some different gain values are illustrated in Fig. 2-4b. The threshold is controlled by changing the value of I_{DFT} (Fig. 2-3a). The peak (magnitude) is controlled by changing the value of I_{REF} (Fig. 2-3b). The peak and gain can also be controlled simultaneously as illustrated in Fig. 2-4a. The input current and the break point position (threshold) can be positive or negative. The power consumption of the circuit in the quiescent state with I_{DFT} = 0 & I_{REF} = 5 uA is equal to 15 uWatts. The use of low voltage cascode current mirrors increases the output resistance of the circuit and minimizes the Channel Length Modulation effect (by forcing the mirroring transistors

to have the same Drain-Source voltage difference - V_{DS}), giving as such more accurate results.

2. 3. RBF Implementation

In this section a simple technique will be proposed to extract the RBF from the TBF circuit of Fig. 2-2.

2. 3. 1. Proposed Technique

Fig. 2-5 illustrates the idea behind this technique. As can be inferred from the graph, in order to extract the RBF from the TBF function, we need to add (or subtract) nonlinearity to (or from) the linear parts of the TBF characteristic. The subtracted and added nonlinearities should have different peaks. This can be achieved using two peaking current sources [51] with different characteristics (different peak values).

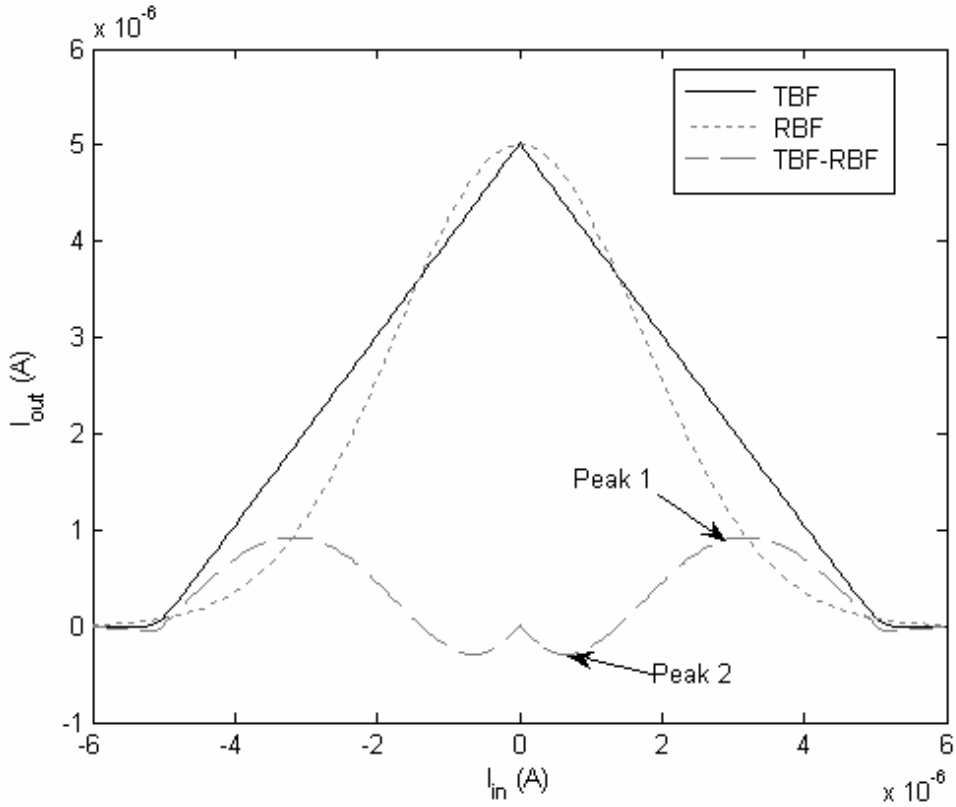


Figure 2-5 A comparison between the TBF & RBF.

Fig. 2-6 shows the circuit of the peaking current source [51] and its input-output characteristics (V_{DD} = 1 V). Inspection of the circuit in Fig. 2-6 shows that

$$V_{SG1} - I_{in}R = V_{SG2} \tag{2-1}$$

Using equation (2-1), when the input current is small, the voltage drop on the resistor is small, and $I_{OUT} \cong I_{IN}$. As the input current increases, V_{SG1} increases more slowly than the drop on the resistor. Any increase in the input current eventually causes the gate-source voltage of M_2 to decrease. The output current reaches a maximum when V_{SG2} is maximum. As can be deduced from the graph, the value of R controls the

peak of this source. Replacing R with voltage controlled MOS resistance will reduce significantly the area consumed by the circuit and will enhance the programmability feature of the circuit (Fig. 2-7).

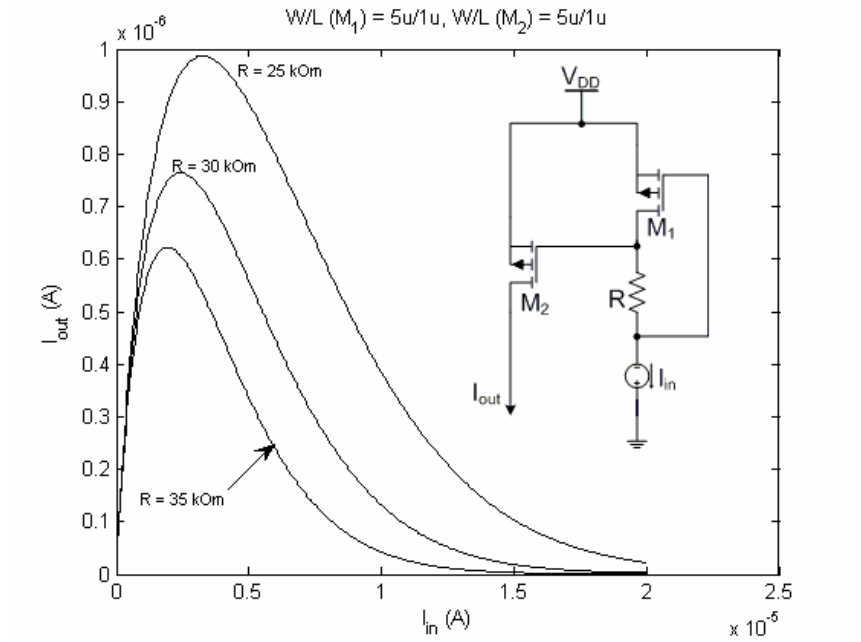


Figure 2-6 Peaking current source with MOS resistance.

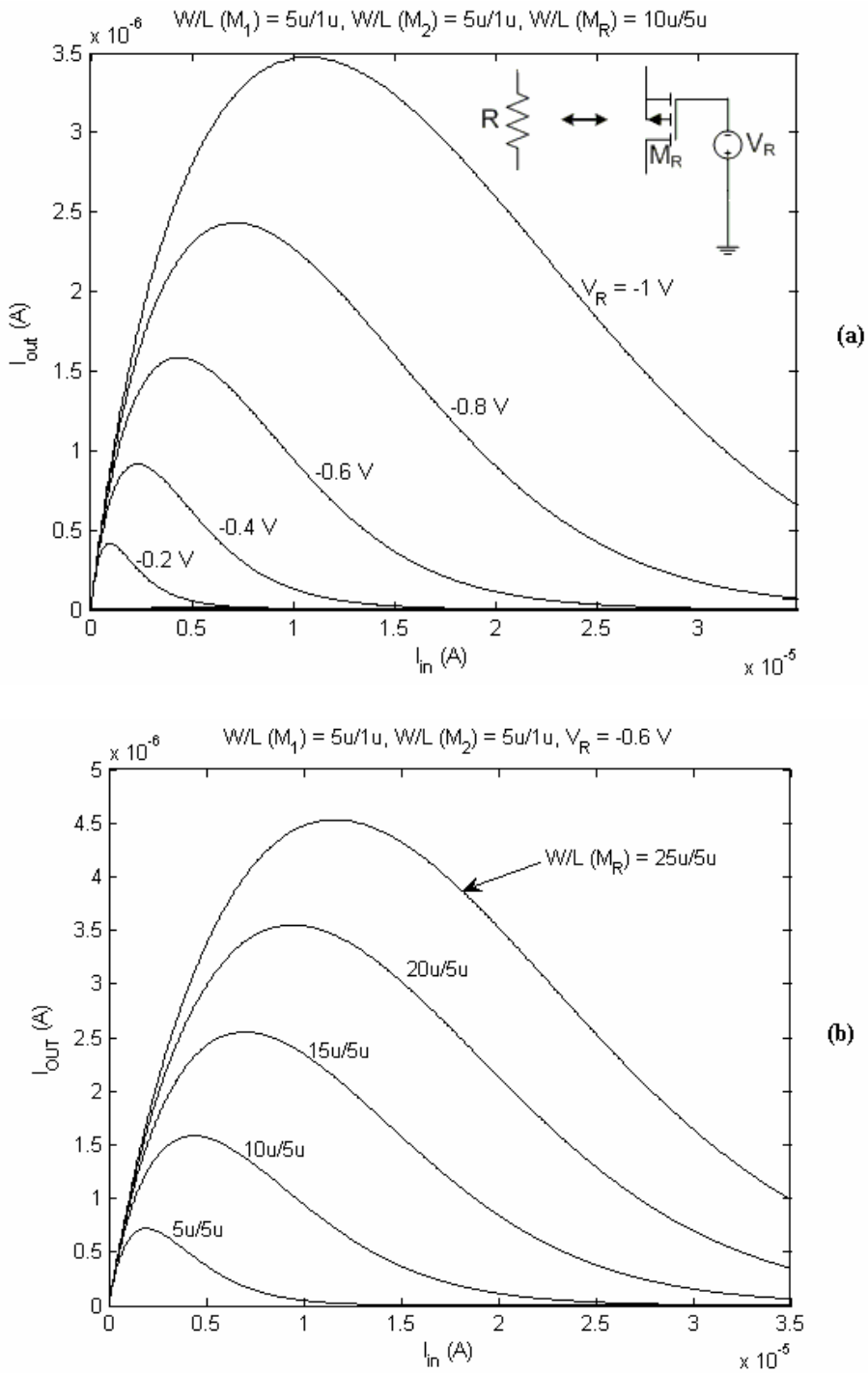


Figure 2-7 Peaking current source with MOS resistance. a) The effect of varying V_R and b) The effect of varying dimensions of transistor M_R on the output shape.

The simulation results of the peaking current source, with R replaced with voltage controlled MOS resistance, are depicted in Fig. 2-7. Fig. 2-7a shows the effect of varying V_R on the input-output characteristics of the peaking current source, while Fig. 2-7b shows the effect of using different dimensions of transistor M_R on these characteristics. It is clear from Fig. 2-7 that the output peaking current is proportional to the increase in the dimensions ratio (W/L) of transistor M_R, and inversely proportional to the decrease in the value of the voltage source V_R.

As a result, the characteristic of the peaking current-source can be controlled by proper selection of the transistors dimensions and the control voltage applied to the gate of this MOSFET.

2. 3. 2. RBF Circuit Realization and Description

The RBF can be realized by adding two peaking current sources to the TBF circuit of Fig. 2-2. Figure 2-8 shows the resulting TBF/RBF circuit implementation. The first peaking current source is formed of transistors M_{25} - M_{27} and the second peaking current source is formed of transistors M_{30} - M_{32} . In the region of peak 1 of Fig. 2-5, the input current I_{in} is relatively large (I_{DFT} = 0). This current will flow through M_2 (or through M_1 if the input current is negative) and eventually will be mirrored to M_{29}, thus forming the input current to the second peaking current-source formed of M_{30} - M_{32}. However, since it is a relatively large current, then the output current of the second peaking current-source formed of M_{30} - M_{32} will be negligible and can be

ignored. The current flowing through M_2 will be also mirrored to M_{18} where it will be subtracted from the reference current I_{REF} to form the current flowing through M_{20}.

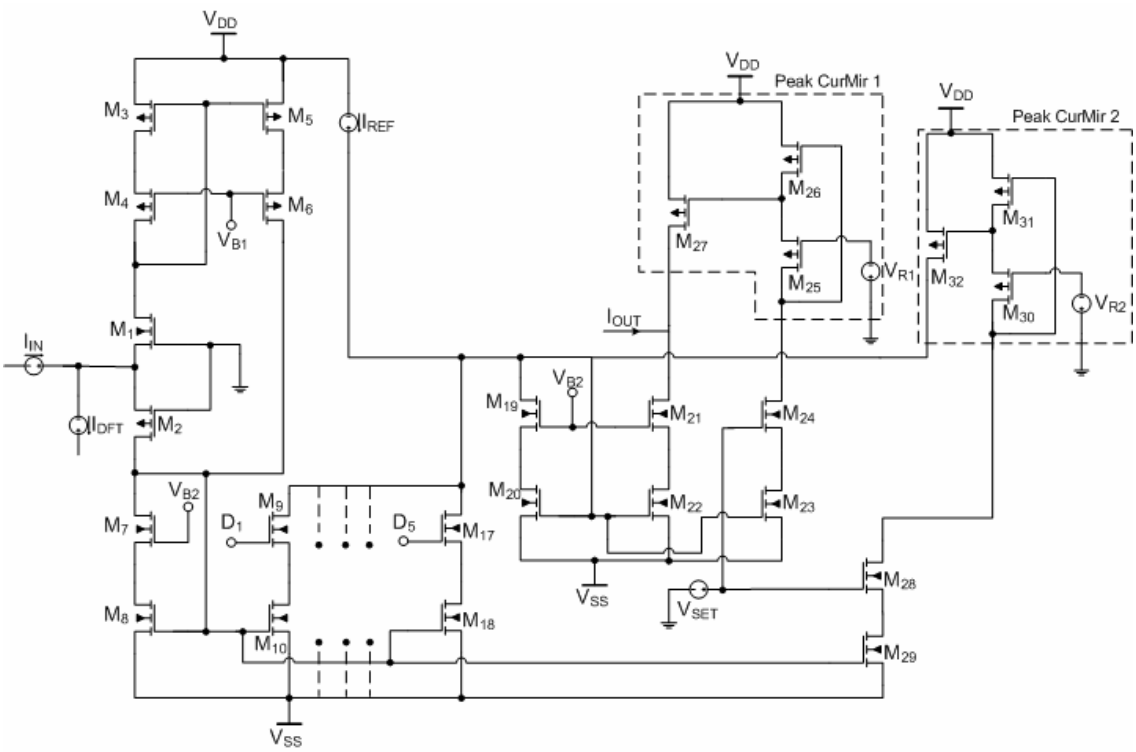


Figure 2-8 TBF/RBF circuit implementation.

This current will be mirrored to M_{22} and M_{23}. The current through M_{23} will form the input current to the first peaking current-source formed of M_{25} – M_{27}. Since it is a relatively small current, a peaking current will be obtained and will be subtracted from the current flowing through M_{22} to form the output current in the region of peak 1 of Fig. 2-5. In the region of peak 2 of Fig. 2-5, the input current is relatively small. This current will flow through M_2 and eventually will be mirrored to M_{29}, thus forming the input current to the second peaking current-source formed of M_{30} – M_{32}.

However, since it is a relatively small current, then the output current of the second peaking current-source formed of transistors M_{30} – M_{32} will not be negligible and a peaking current will be obtained. The current flowing through M_2 (or M_1 for negative input current) will be also mirrored to M_{18} where it will be subtracted from the reference current I_{REF} to form the current flowing through M_{20}. This current will be mirrored to M_{22} an M_{23}. The current through M_{23} will form the input current to the first peaking current-source formed of transistors M_{25} – M_{27}. Since it is a relatively large current, the output current of the first peaking current-source formed of transistors M_{25} – M_{27} will be negligible and can be ignored. The output current will be formed by adding the peaking current obtained from M_{32} to the difference between the reference current I_{REF} and the input current I_{in}. The resulting current will flow through M_{20} and then mirrored to M_{22} to produce the output current. This will form the output current in the region of peak 1 of Fig. 2-5. Ultimately, we will get the nonlinearity of interest.

The peaking current sources can be switched on or off by setting the value of V_{SET} to either zero potential (RBF implementation) or to V_{SS} (TBF implementation), respectively.

2. 3. 3. Simulation Results and Discussion

The circuit of Fig. 2-8 was simulated. V_{SET} was set to zero potential to activate the peaking current sources. For the RBF implementation, only transistors M_{13}, M_{14} are activated (by setting D_3 to zero potential and setting the rest of D_n to V_{SS}). Figure 2-9

illustrates the programmability feature of the circuit and shows a comparison between the simulated result and the mathematical expression of the Gaussian function (Equation 1-6 is tried, with different parameters, to fit the simulated results). The calculated normalized error of the simulated function (Fig. 2-9a) is less than 1%. The Standard Deviation (STD) of the function can be programmed by changing the value of V_{R2} as illustrated in Fig. 2-10a. Different STD values for different V_{R2} values are shown in Table 2-1. Table 2-2 Shows dimensions of the transistors used in the circuit.

Table 2-1 STD programmability

V_{R2} (V)	STD (Normalized)
-0.2	1.61
-0.4	1.925
-0.6	2.31
-0.8	2.685
-1	3.035

Table 2-2 Transistors dimensions.

Transistors	Dimensions
M_1 M_2 M_3 M_5 M_8 M_{11} M_{14} M_{24} M_{28} M_{29}	10/1
M_7 M_{13} M_{16} M_{20} M_{22} M_{25}	20/1
M_4 M_6 M_{15} M_{18} M_{19} M_{21}	40/1
M_9 M_{12} M_{27} M_{31} M_{32}	5/1
M_{10}	2.5/1
M_{26}	6/2
M_{30}	10/5
M_{23}	13.5/1
M_{17}	80/1

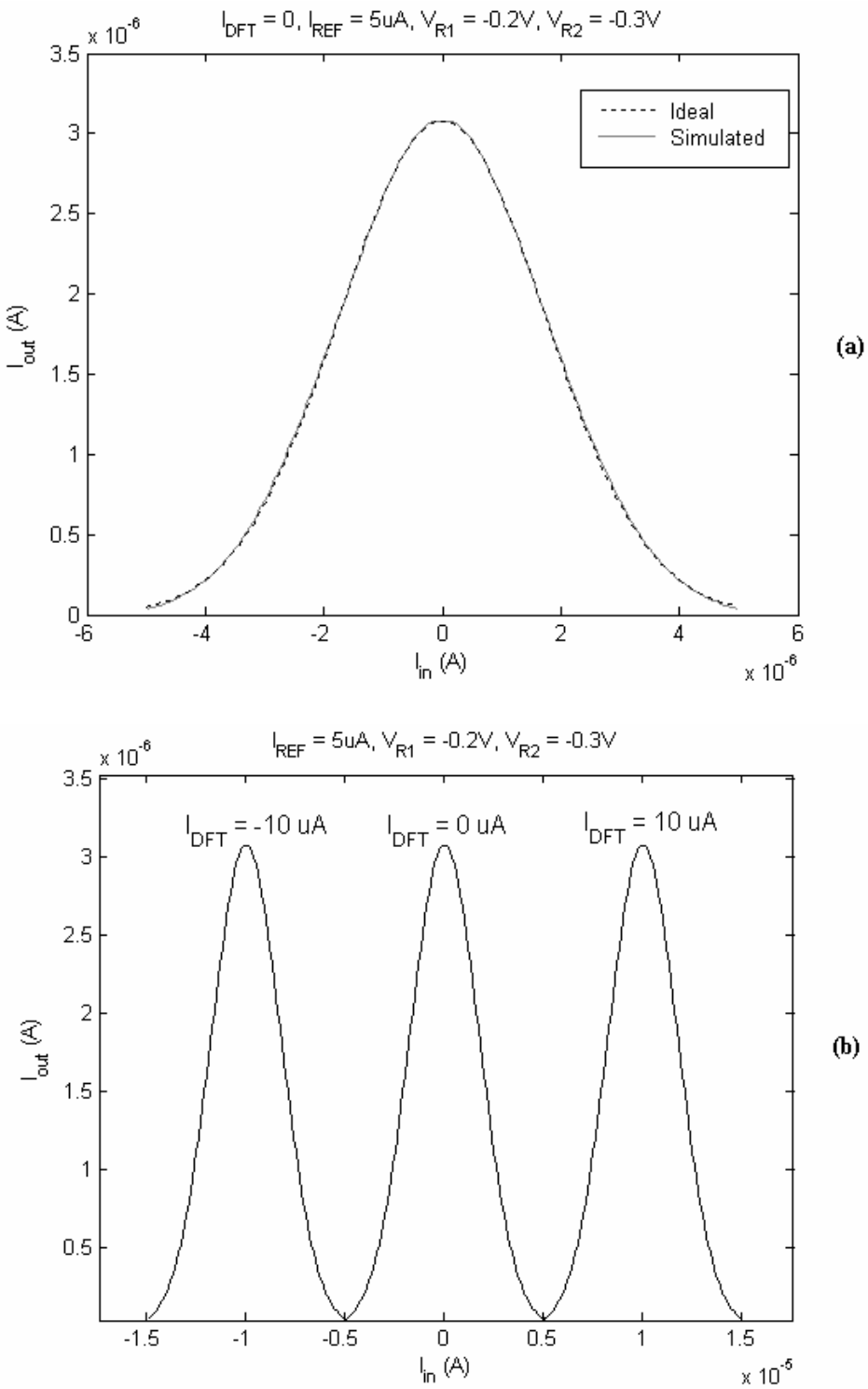


Figure 2-9 (a) Ideal & simulated RBF (b) Threshold programmability.

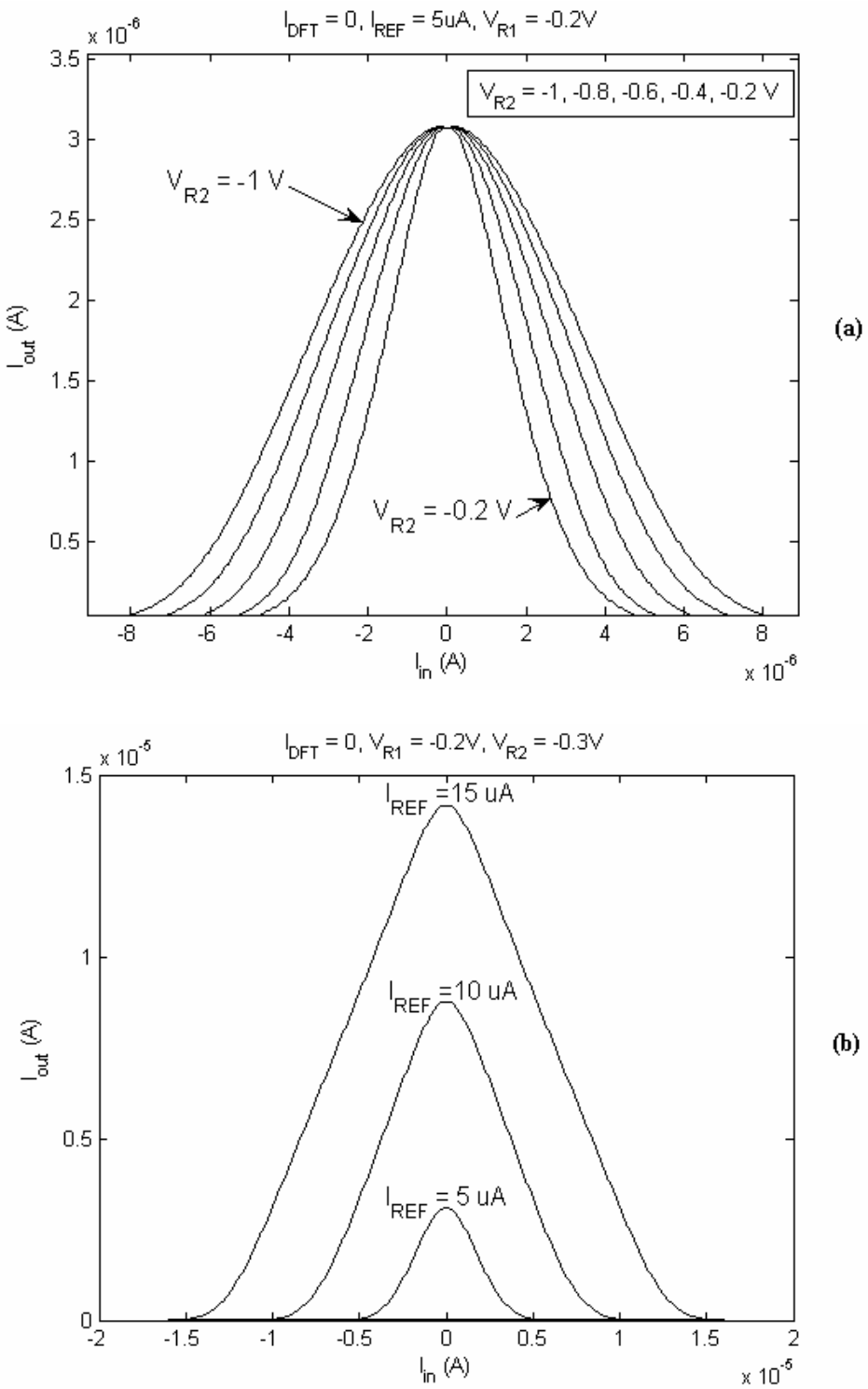


Figure 2-10 (a) STD programmability (b) Peak programmability.

Figure 2-11 shows the currents of the two peaking current sources which achieve the best accuracy for the RBF implementation. $I(M_{27})$ is the output current of the peak current mirror 1 (see Fig. 2-8), and $I(M_{32})$ is the output current of the peak current mirror 2. As can be inferred from this figure the RBF peak is reduced (compared to the TBF). The maximum value is around 3uA. This is due to the fact that the output current of the Peak current mirror 1 ($I(M_{27})$) does not go to Zero at the center of the RBF (when $I_{in} = 0$).

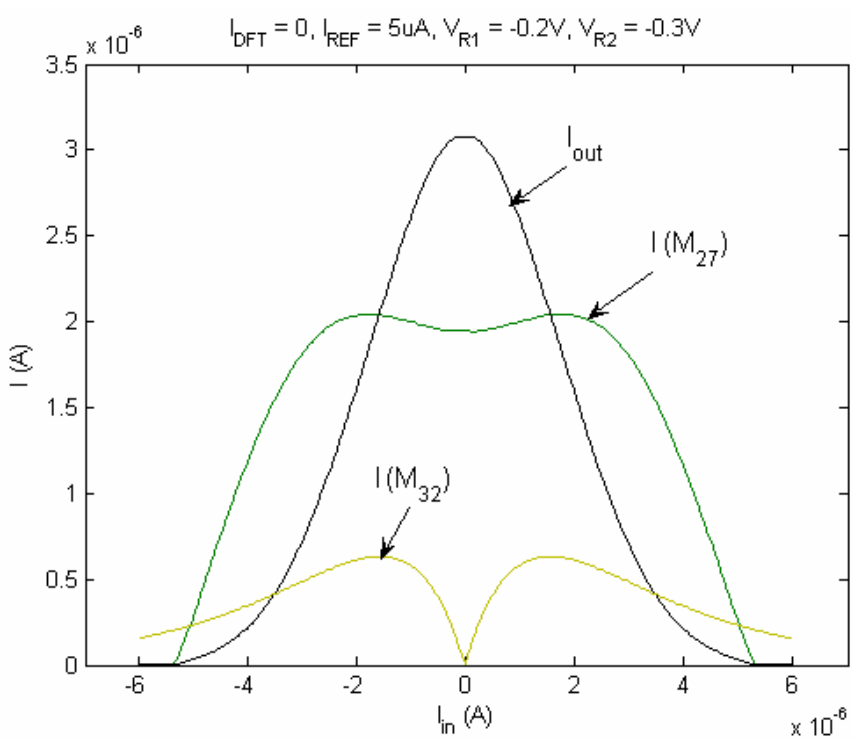


Figure 2-11 RBF & Peak current sources simulations.

2. 4. Sigmoid Function Implementation

In this section it will be shown how the Sigmoid function can be obtained from the circuits presented in previous sections.

If we refer again to figures 2-5 & 2-8, one can see that the position of the peak (center) is common between both the TBF & RBF. Both centers correspond to a situation where I_{IN} equals I_{DFT}, which means that no current will be flowing through M_7 and M_8 and consequently no current subtraction from I_{REF}. In this case the output current of the circuit will be maximum. Through the rising edges of the functions and as we move toward the center, the current flowing through M_7 and M_8 starts decreasing until it approaches zero at the center. After the peak (center) this current starts again increasing causing the falling edge. So, if after the peak we can prevent this current from being mirrored to the output then we will not have the falling edge. In other words we need to switch off the current mirroring of $I_D(M_7)$ as soon as I_{IN} exceeds I_{DFT}. As such, rather than having falling edges we will have flat characteristics after the centers which results in SatLin and Sigmoid functions approximations. To implement this, we need first to detect the excess of I_{IN} beyond I_{DFT}. This can be done using a current comparator, but the price is a more complicated circuit and a large increase in the area consumption. Instead, a voltage comparator can be used. Since the input voltage increases beyond zero when I_{IN} exceeds I_{DFT}, then we can compare this voltage with the ground. But having the other terminal of the Comparator at ground may result in oscillation in the output due to noise

contamination. The output of this comparator should cause some switches connected to the gates of the current mirror (M_8, M_{10}, M_{12}, M_{14}, M_{16}, M_{18}) to be switched on or off, depending on the connections type and the input signal. To be able to switch between different functions we suggest using a logic gate which can enable or disable the effect of the comparator (Fig. 2-12).

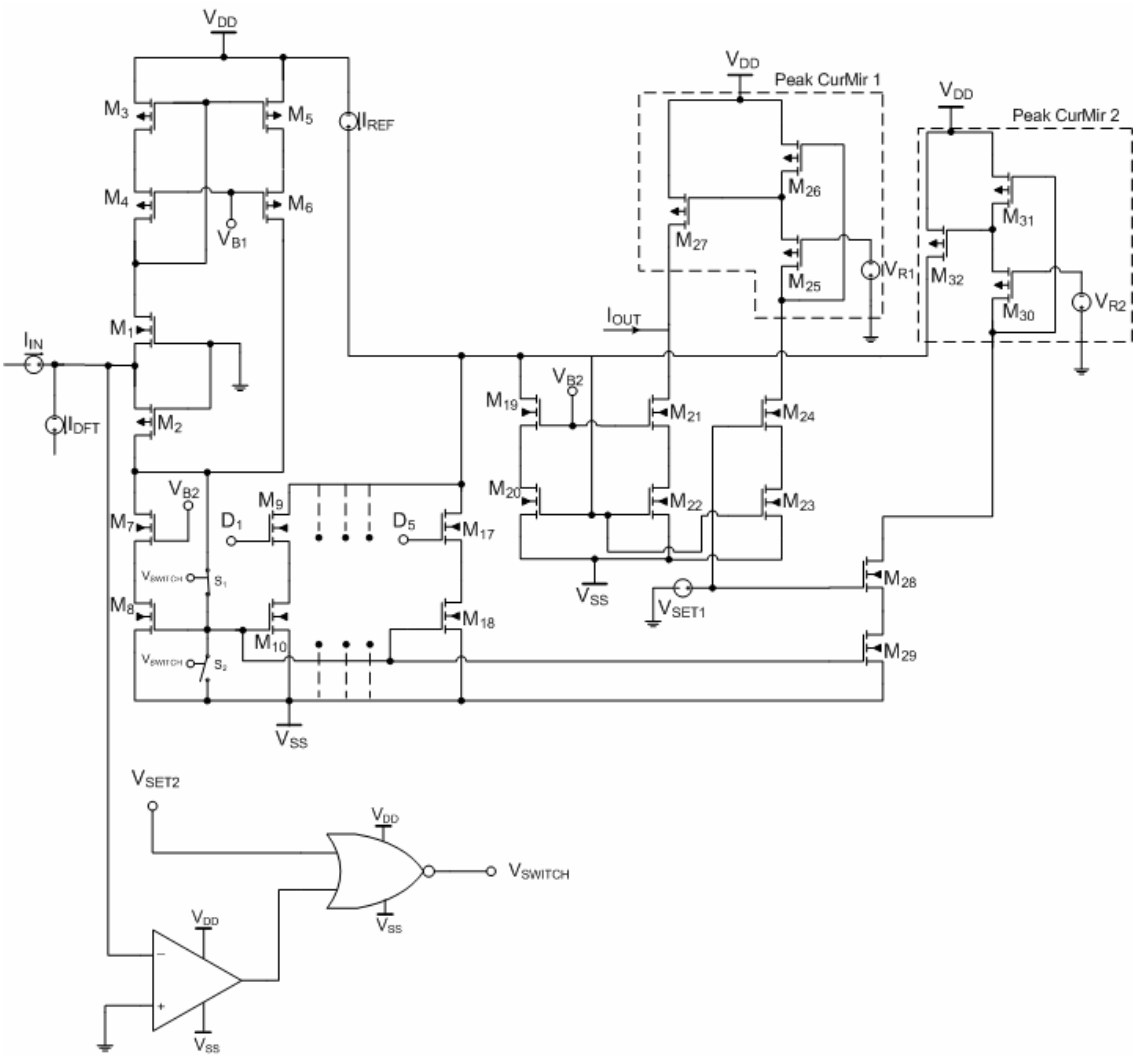


Figure 2-12 The final generic circuit.

Fig. 2-12 shows the final circuit. The voltage comparator can be implemented using the simple Comparator. A NOR-gate will be suitable for enabling ($V_{SET2} = V_{SS}$) or disabling ($V_{SET2} = V_{DD}$) the switching.

2. 4. 1. Circuit Description

Switches S_1 (PMOS transistor) and S_2 (NMOS transistor) are responsible for switching the current mirroring ON or OFF. Both transistor gates are connected to the output of the NOR-gate (V_{SWITCH}). If V_{SWITCH} is low ($V_{SWITCH} = V_{SS}$) then S_1 will be ON and S_2 will be OFF which means that the current mirroring is ON. If V_{SWITCH} is high ($V_{SWITCH} = V_{DD}$) then the mirroring transistors will be OFF as the gates of the bottom transistors will be connected to V_{SS} (S_1 is OFF and S_2 is ON).

The NOR-gate enables the control of the switching and consequently the choice of the implemented function. If V_{SET2} is set to V_{DD} then V_{SWITCH} will be low and the current mirroring will be always active, no matter what is the output of the comparator. Now, if V_{SET2} is set to V_{SS} then the value of V_{SWITCH} will be dependent on the output of the comparator. If the output of the comparator is switched to low then V_{SWITCH} will be switched to high and the current mirroring will be off. If the negative input of the comparator is connected to the input voltage of the circuit and the positive input of the comparator is connected to the ground then V_{SWITCH} will be set to high whenever I_{IN} exceeds I_{DFT} . So for enabling the current mirror switching V_{SET2} should be set to V_{SS} (the falling edge of the function will be removed from the output response). Now, provided that V_{SET2} is set to V_{SS} one can choose between the

Sigmoid and the SatLin functions by properly setting the value of V_{SET1}. Setting V_{SET1} to zero potential will activate the peaking current sources leading to a Sigmoid function implementation, while setting V_{SET1} to V_{SS} will switch them off leading to a SatLin function implementation. Figure 2-13 shows the implementation of the NOR-gate [53] and Figure 2-14 shows the implementation of the voltage comparator [52].

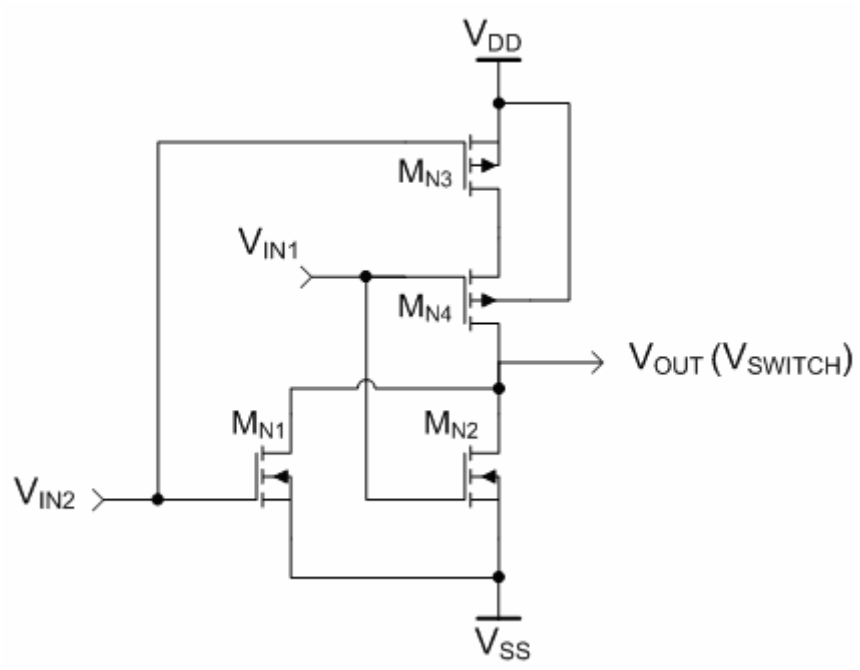


Figure 2-13 NOR-Gate implementation.

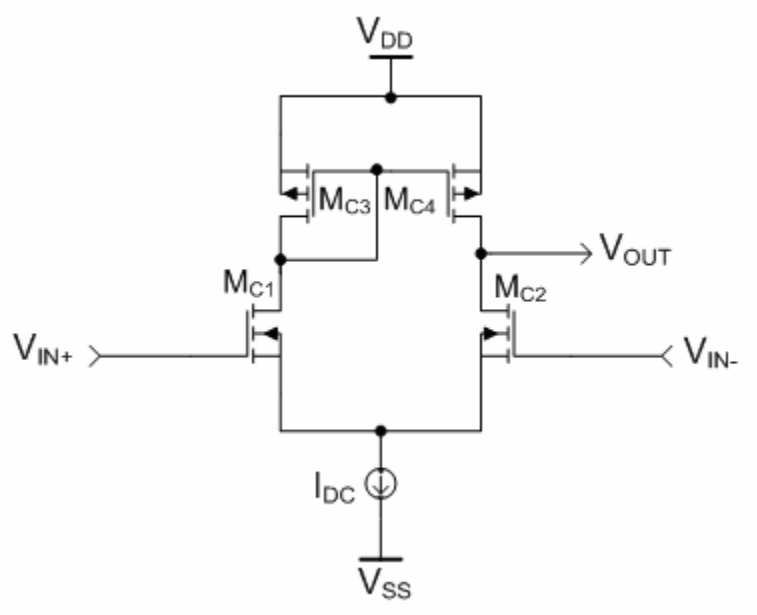


Figure 2-14 Voltage comparator implementation.

2. 4. 2. Simulation Results and Discussion

The circuit of figure 2-12 is simulated. V_{SET2} is set to V_{SS} to enable controlling switches S_1 and S_2 (to remove the falling edge from the implemented function). V_{SET1} is kept at zero potential (for enabling the peaking current sources) to implement the Sigmoid function. D_3 is set to zero potential, and the rest of D_n are set to V_{SS} (M_{13} and M_{14} are activated). Table 2-3 shows dimensions of the additional transistors.

Table 2-3 Transistors dimensions.

Transistors	Dimensions
M_{S1}	40/1
M_{S2}	15/1
M_{N1} M_{N2} M_{C1} M_{C2}	4/1
M_{N3} M_{N4} M_{C3} M_{C4}	3/1

Figure 2-15 shows a comparison between a simulated and ideal Sigmoid functions.

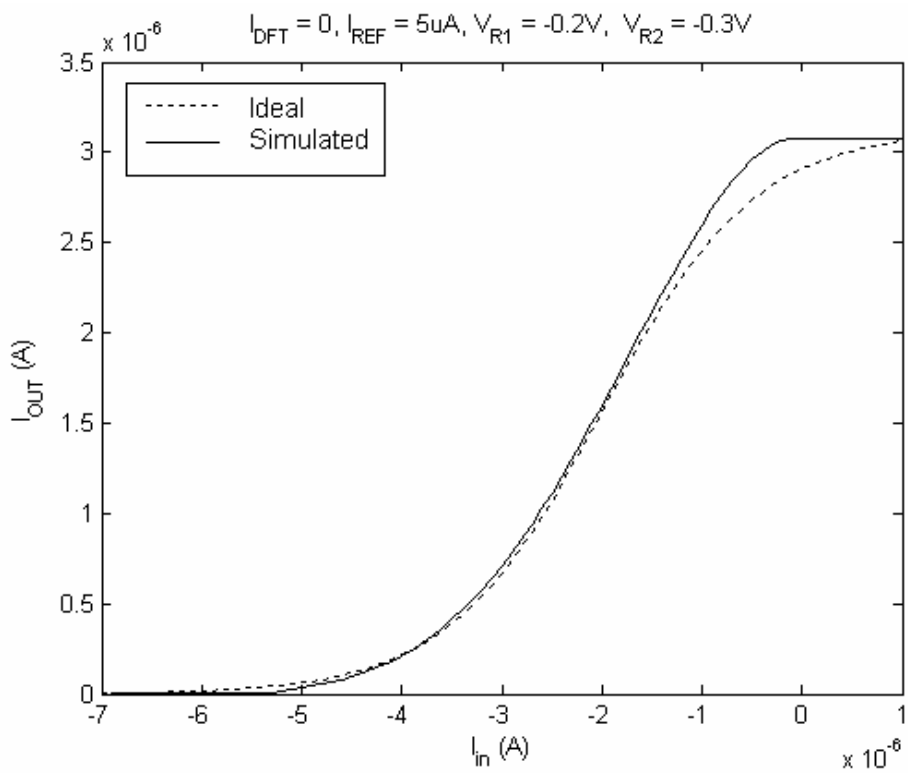


Figure 2-15 Sigmoid function ($V_{SET2} = V_{SS}$, $V_{SET1} = 0$).

From this figure it is obvious that the error is much higher around the peak. This results from the fact that, for the nonlinear blocks of Fig. 2-8, focus was on the implementation of the Gaussian function which has smooth flat characteristics on both sides. Nonetheless, we can tune this circuit to have more accurate Sigmoid function representation by adjusting V_{R1} , V_{R2}. Fig. 2-16 shows that it is possible to make the function representation more symmetric by tuning V_{R1} .

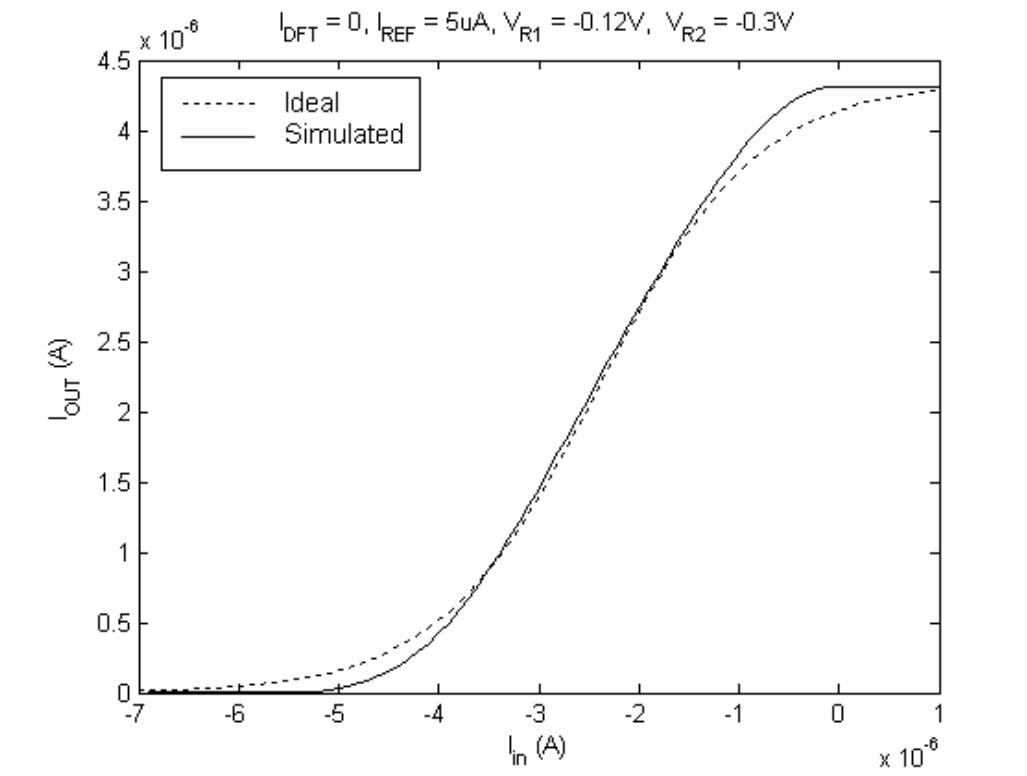


Figure 2-16 Symmetric Sigmoid-like representation.

Although we could achieve more symmetrical representation the error is still almost the same. We can improve the accuracy further by tuning V_{R2} also (as shown in Fig. 2-17). The least RRMS error achieved for the sigmoid function is around 3% which is not better than the reported results (see Table 1-2). Threshold, peak and slope can also be programmable. Threshold can be controlled or shifted by means of the current source I_{DFT}. I_{REF} controls the peak (Fig. 2-18). Peak can also be fine tuned by tuning the voltage V_{R1} (Fig. 2-19). Slope tuning is possible by changing V_{R2} (Fig. 2-20).

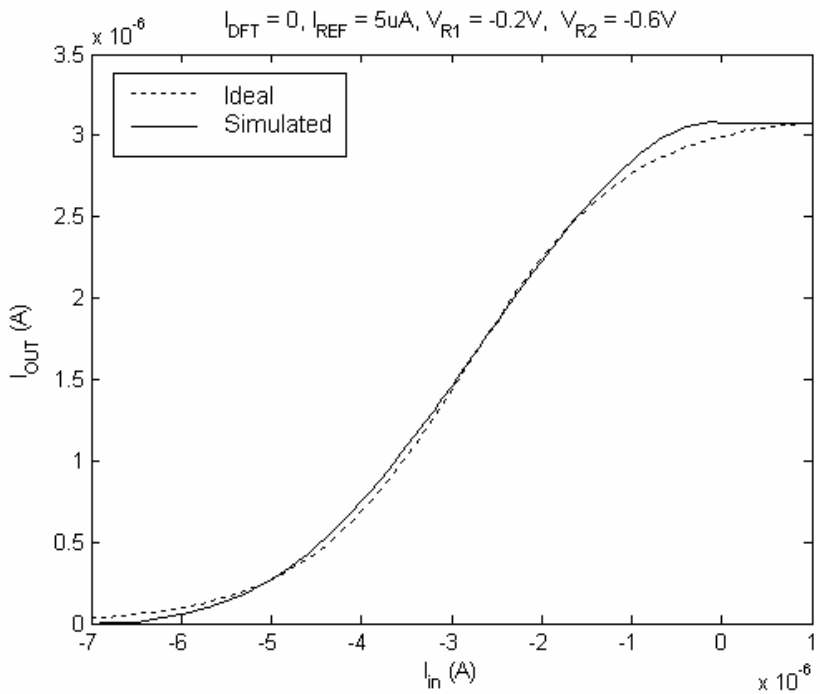


Figure 2-17 Improving the accuracy by tuning V_{R2}.

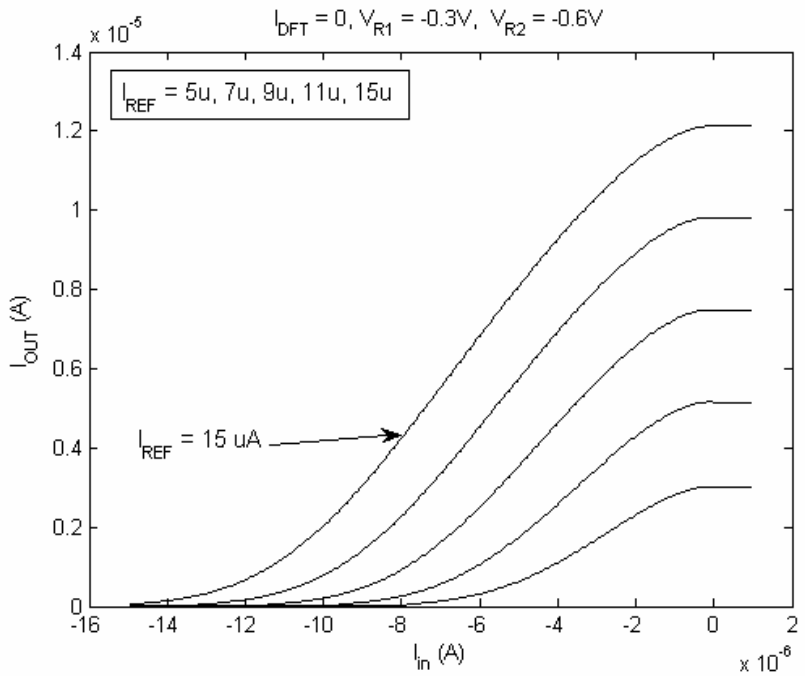


Figure 2-18 Peak programmability by changing I_{REF}.

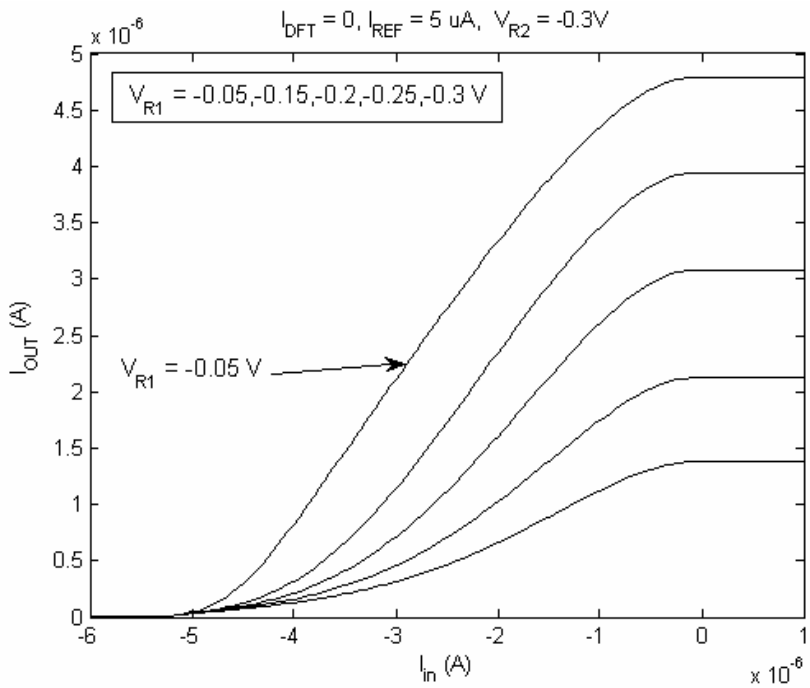


Figure 2-19 Peak's fine tuning.

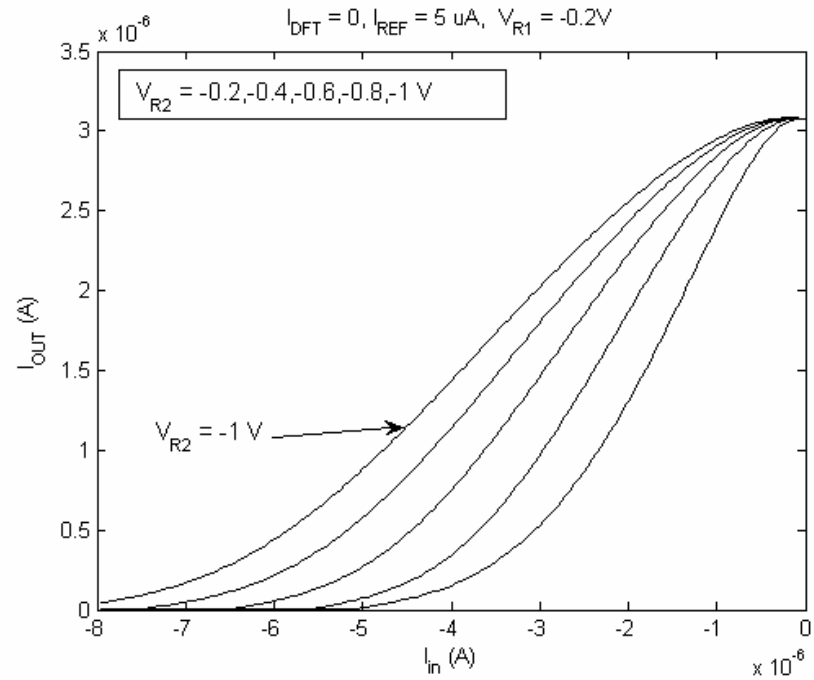


Figure 2-20 Slope programmability by changing V_{R1}.

2. 5. Sigmoid Function Realization Improvement

In the previous section, the minimum RRMS error achieved for Sigmoid function realization was 3%. Achieving a better accuracy was not possible due to the fact that the two peaks (peak 1 and peak 2 of Fig. 2-5) were not symmetric (asymmetry between the two peaks was important for the RBF realization), and it was not possible to make them very close to each other by only tuning V_{R1} and V_{R2}. If we refer to Fig. 2-15, it can be seen that the error of the implemented function was maximum at the peak. And it was explained in section 2.3 (see Fig. 2-5) that the RBF curves close to the peak were characterized mainly by peak 1 which was implemented using a peaking current-source formed of transistors M_{30} – M_{32} . In this section a better accuracy is achieved by adding another peaking current-source, with characteristics similar to the characteristics of peak 2 realized by the peaking current-source formed of transistors M_{25} – M_{27}. The added peaking current-source will operate in a similar way to the operation of the peaking current-source formed of transistors M_{30} – M_{32} but with a little bit different characteristics.

Figure 2-21 shows the modified circuit. A peaking current-source formed of transistors M_{35} – M_{37} is added to the circuit. This peaking current-source will be controlled by the voltage source V_{SET1_3} . For Sigmoid realization V_{SET1_1} and V_{SET1_3} should be set to zero (switching on), while V_{SET1_2} should be set to V_{SS} (switching off). By trial and error it was found that the best accuracy can be achieved by setting V_{R3} to -0.15 V. Transistors M_{35} – M_{37} have the same dimensions as M_{25} – M_{27}.

Figure 2-22a shows the realized function.

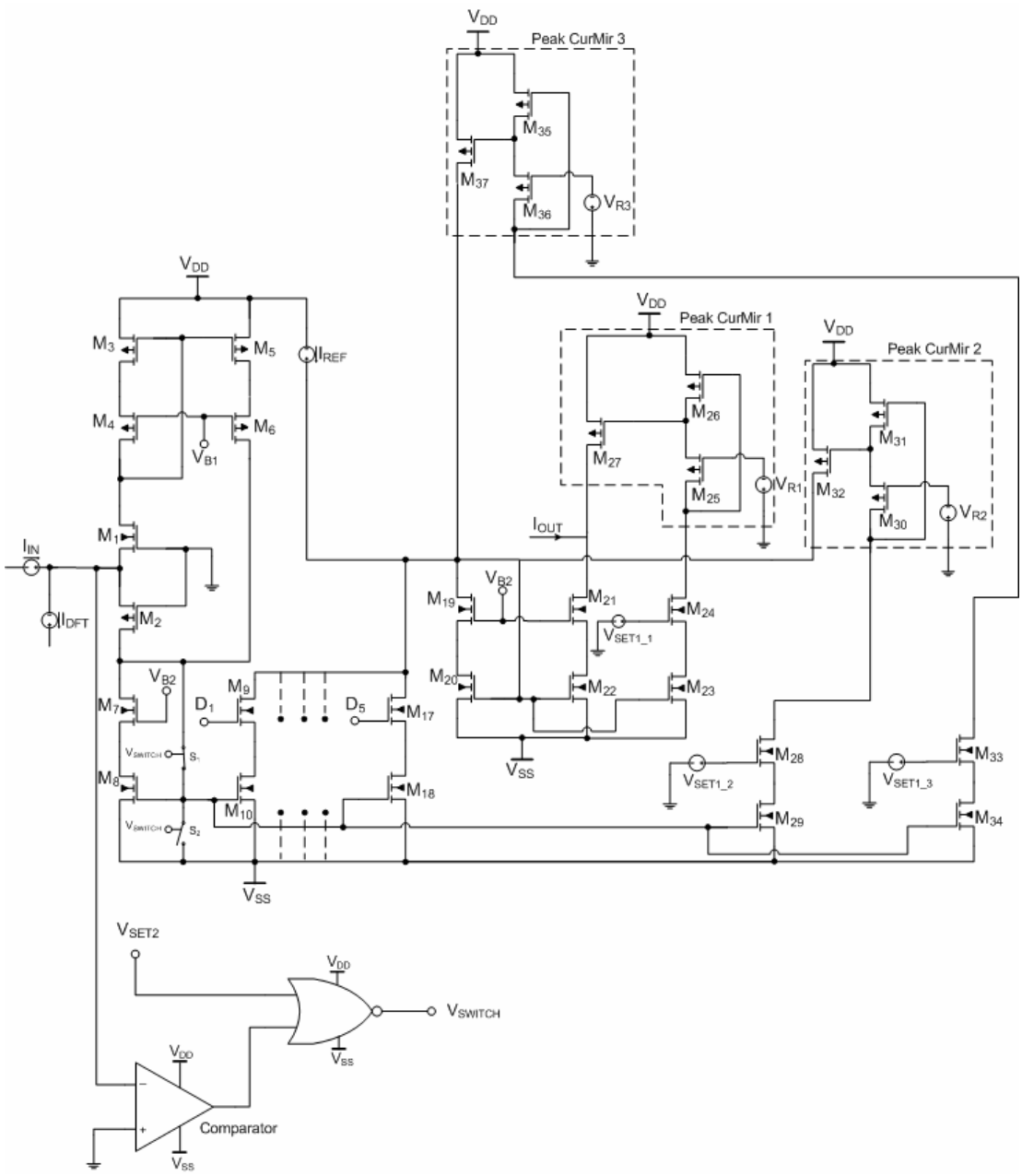
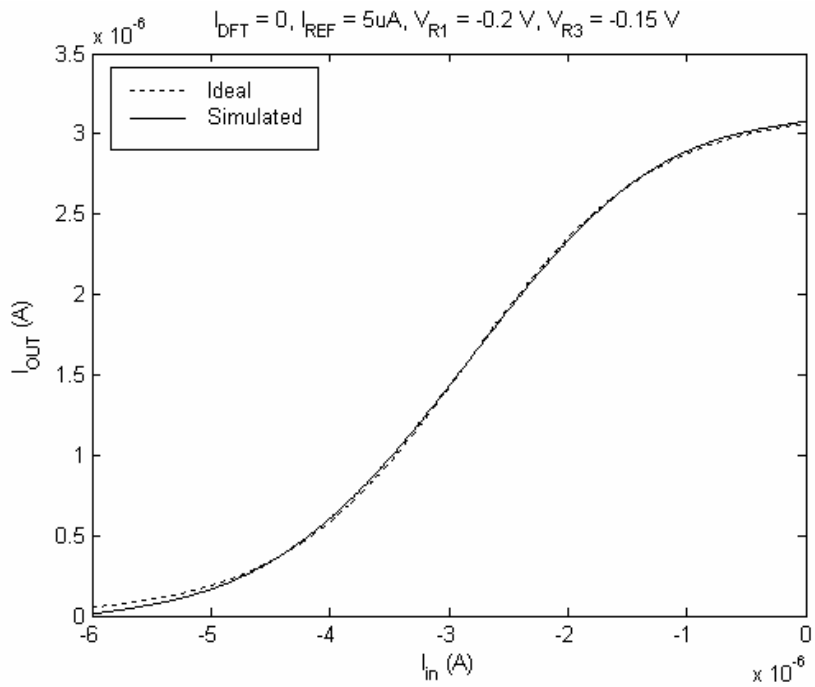
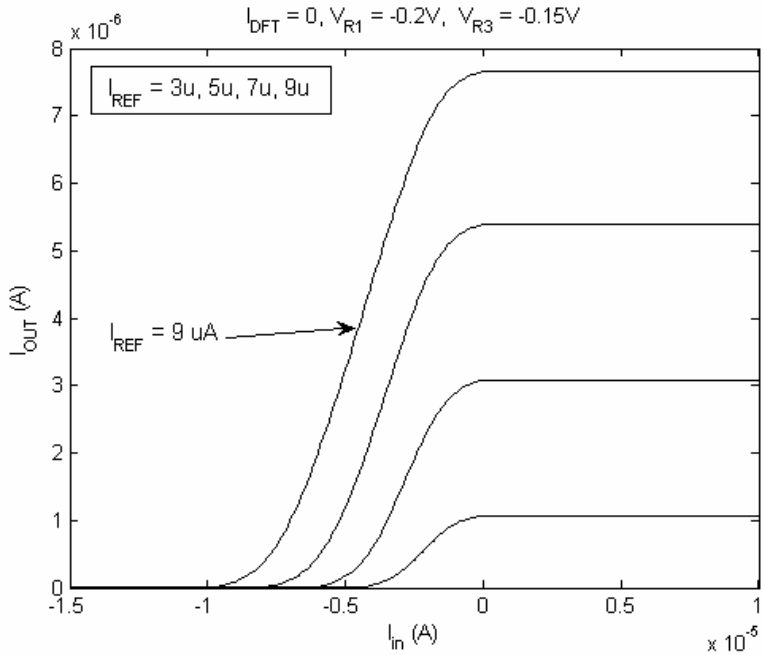


Figure 2-21 A modified circuit for more accurate Sigmoid function realization.

The accuracy of the realized Sigmoid function is now much better. RRMS error is now around 1%. Figure 22b shows the peak programmability using I_{REF}.



(a)



(b)

Figure 2-22 (a) The realized Sigmoid function (b) Peak programmability

Threshold programmability is also possible by using different values for I_{DFT} as shown in Fig. 2-23.

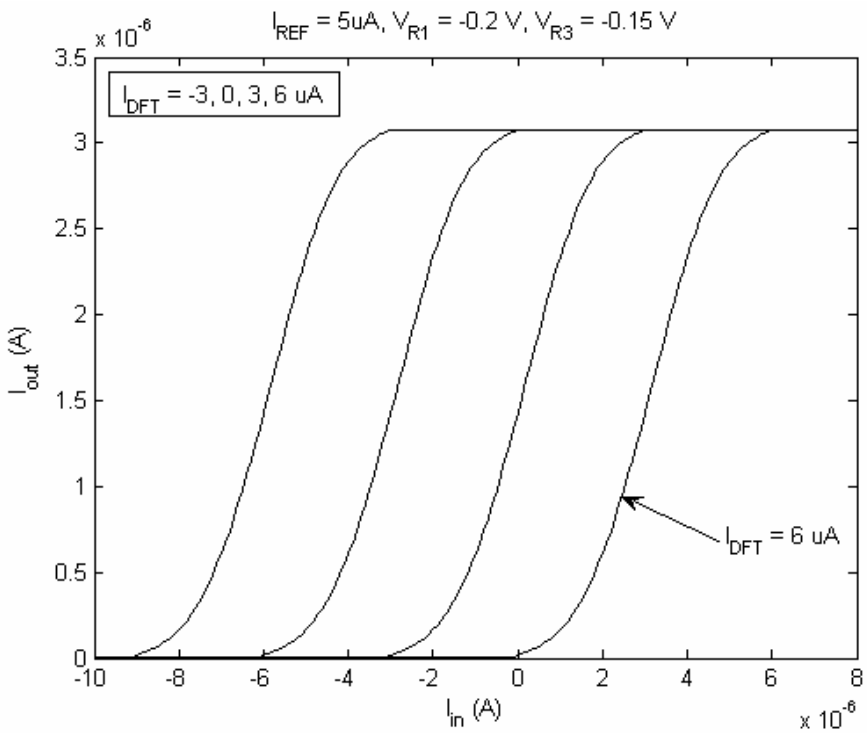


Figure 2-23 Threshold programmability of the Sigmoid function.

The slope programmability was achieved by changing the voltage source V_{R2} (Fig. 2-20), significantly affecting the accuracy and the symmetry of the function. Another effective way is proposed here for achieving more accurate slope programmability.

In section 2.2, a digitally controlled combination of parallel connected current mirrors, with different dimension ratios, has been used to control the slope of the TBF lines (M_9 – M_{17}). The digital word composed of five digits D_1 – D_5 were deciding the slope of the function curves. This way of control can also be used to control the slope of the sigmoid function.

But the problem is that the peaking current-source formed of transistors $M_{35} - M_{37}$ (Fig. 2-21) will not be affected by changing the gain of the current mirror formed of transistors ($M_9 - M_{18}$). This will result in a distortion in the realized function because only one peaking current-source will be affected ($M_{25} - M_{27}$). To solve this problem, the circuit is modified to apply the change to both peaking current-sources. Transistors M_{33}, M_{34} of figure 2-21 were replaced in Fig. 2-24 by transistors $M_{33} - M_{42}$ to form with transistors M7, M8 a digitally programmable current mirror. Transistors dimensions have been chosen to achieve the same current mirroring ratios for $M_9 - M_{18}$ and for $M_{33} - M_{42}$. Now the third peaking current-source is formed of transistors $M_{43} - M_{45}$ (Fig. 2-24). The same digital word D1 – D2 will be used to control the gain of both programmable current mirrors. Now we can control the slope of the Sigmoid function while preserving the symmetry and the accuracy of the implemented function (see Fig. 2-25).

The value of the slope (gain) can be chosen between 0.25 and 7.75, corresponding to digital words 00000 and 11111, respectively. Voltage-controlled switches S_3, S_4 were used to switch off the third peaking current-source formed of transistors $M_{43} - M_{45}$ when peace-wise linear function or RBF realization is of interest. S_3 and S_4 are implemented using PMOS and NMOS transistors, respectively. For Sigmoid function realization V_{SWITCH_S} should be set to V_{SS} to enable the third peaking current-source. Otherwise, V_{SWITCH_S} should be set to V_{DD} (for disabling the third peaking current-source).

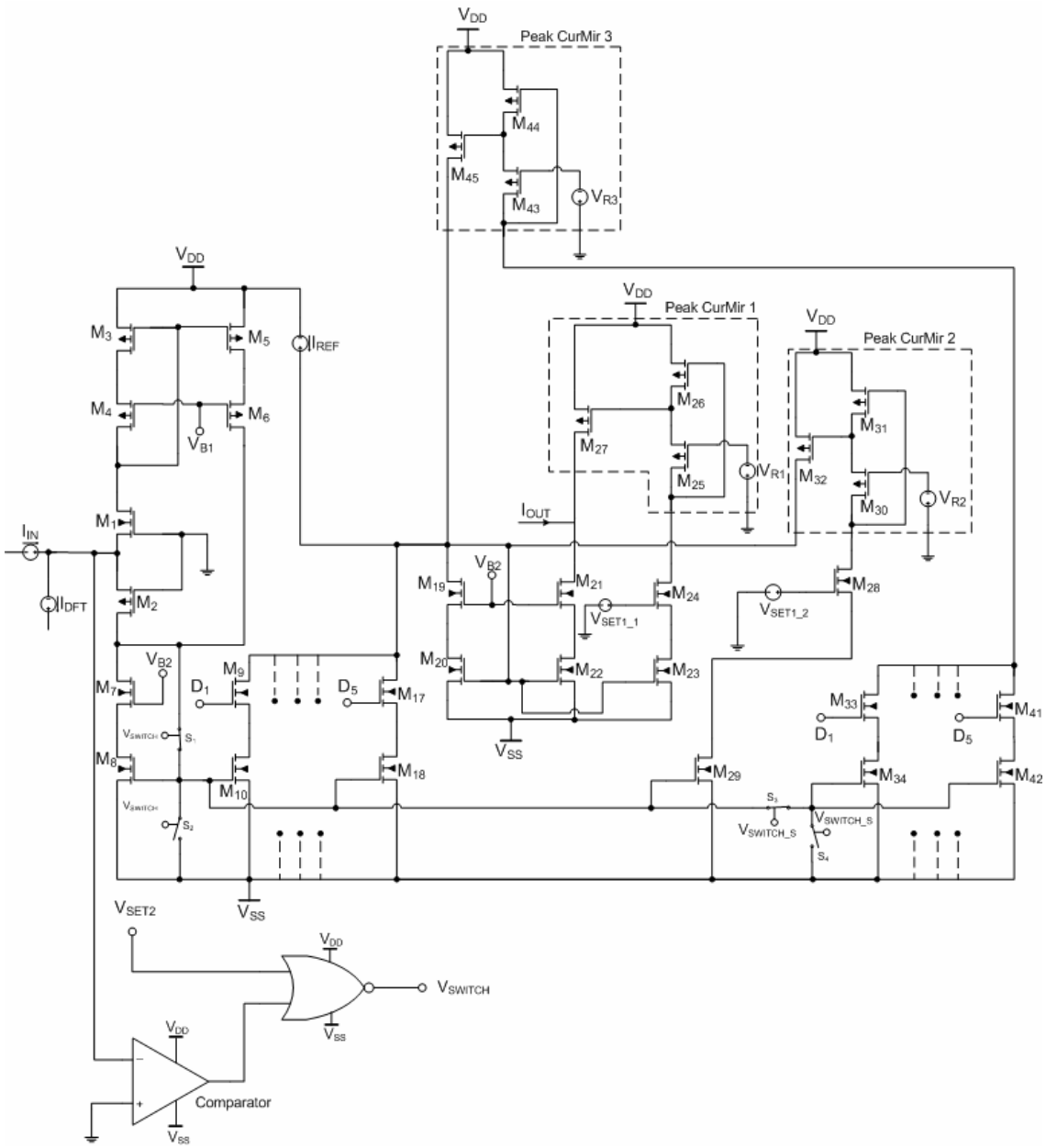


Figure 2-24 The modified circuit for accurate slope programmability (recall that we are always using V_{DD} = +1V, V_{SS} = -1V and V_{B1} = V_{B2} = 0).

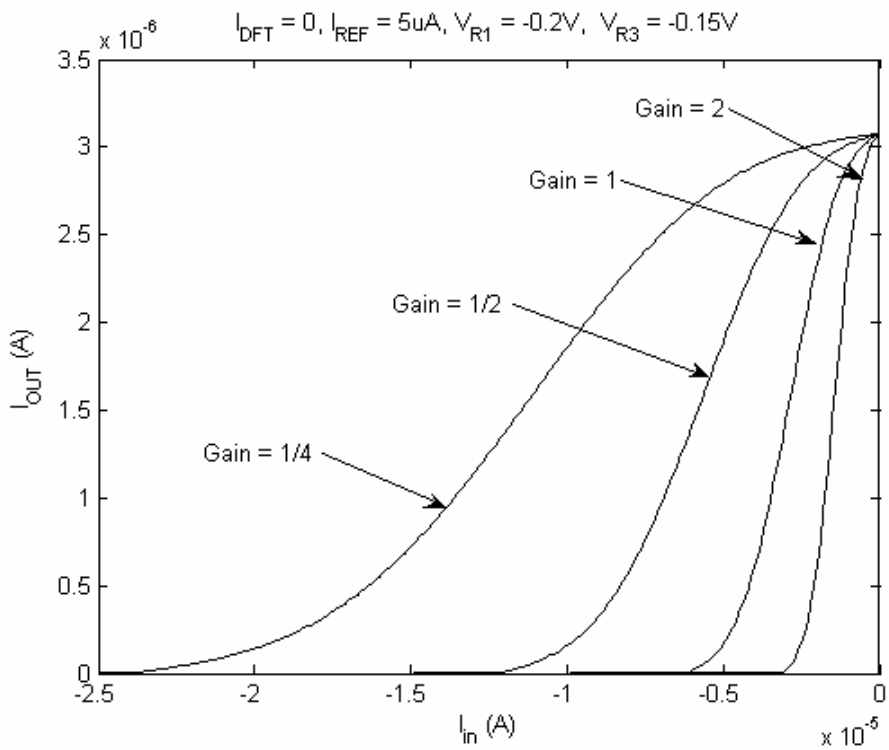


Figure 2-25 Slope programmability of the Sigmoid function.

Table 2-4 includes the dimensions of the additional transistors:

Table 2-4 Additional transistors dimensions.

Transistors	Dimensions
M_{33}	3.5/1
M_{34}	1.625/1
M_{35}	7/1
M_{36}	3.25/1
M_{37}	14/1
M_{38}	6.5/2
M_{39}	28/5
M_{40}	13/1
M_{41}	56/1
M_{42}	26/1
M_{43}	20/1
M_{44}	6/2
M_{45}	5/1

2. 6. SatLin Function Realization

To switch from the Sigmoid to the SatLin function we need to switch off the nonlinear peak current-sources. This can be done simply by setting V_{SET1} to V_{SS}.

2. 6. 1. Simulation Results and Discussion

The circuit of figure 2-24 was simulated. For a piecewise-linear input-output characteristic, V_{SET1} is set to V_{SS} and V_{SWITCH_S} is set to V_{DD} (to switch off the peaking current sources). Figure 2-26 shows the realized SatLin function.

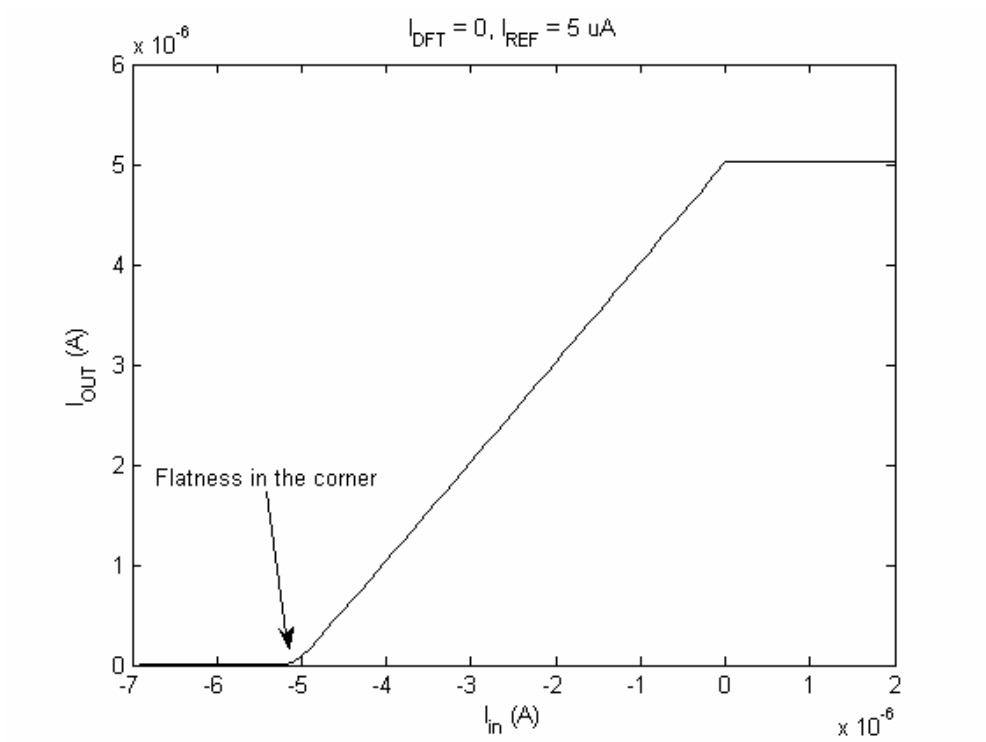


Figure 2-26 SatLin function realization ($V_{SET2} = V_{SS}$, $V_{SET1} = V_{SS}$).

Although we can observe a small flatness in the first corner still the error is very small (less than 1%). Figure 2-27 demonstrates the programmability of the circuit. Threshold (the position of the break points), as usual, is controlled by I_{DFT} .

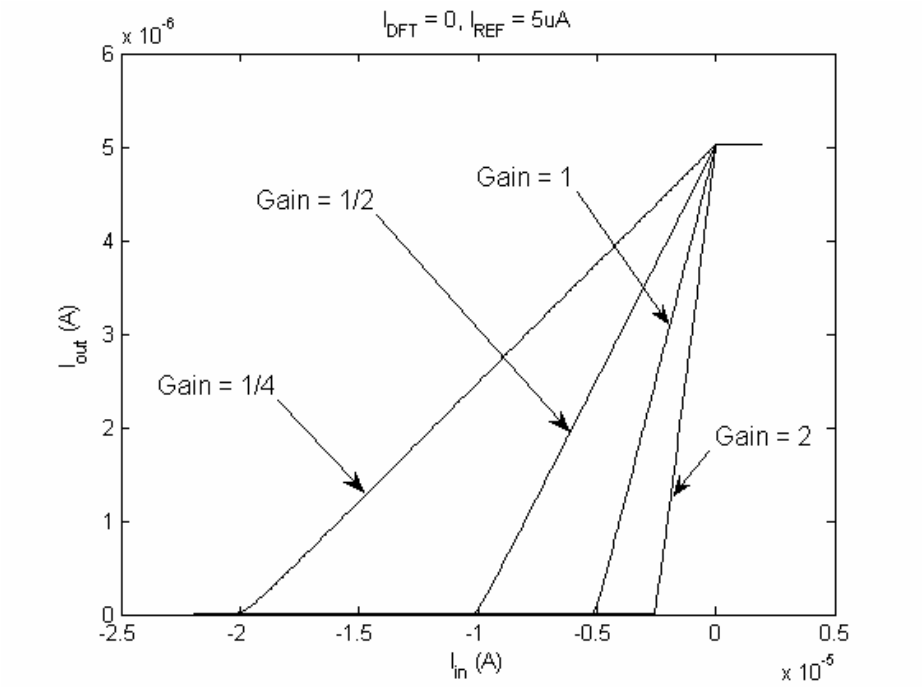


Figure 2-27 (a) SatLin's gain programmability .

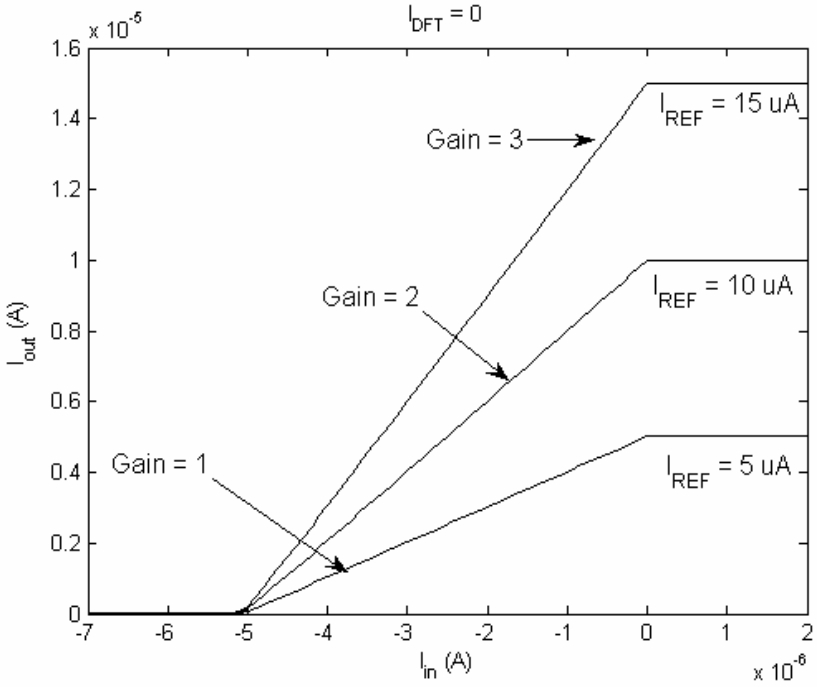


Figure 2-27 (b) SatLin's gain-peak programmability.

CHAPTER 3

Characteristics and Evaluation of the Circuit

3. 1. Evaluation of the Implemented Piecewise-Linear Functions

It was mentioned earlier that the RRMS error of the implemented piecewise-linear functions is very small (around 1%). It is preferable for these functions to have sharp transitions at the break points. In the previous chapter a little undesired flatness was observed at some of these points (Fig. 3-1). The sharpness of these break points of both functions will be evaluated using mathematical criteria [35].

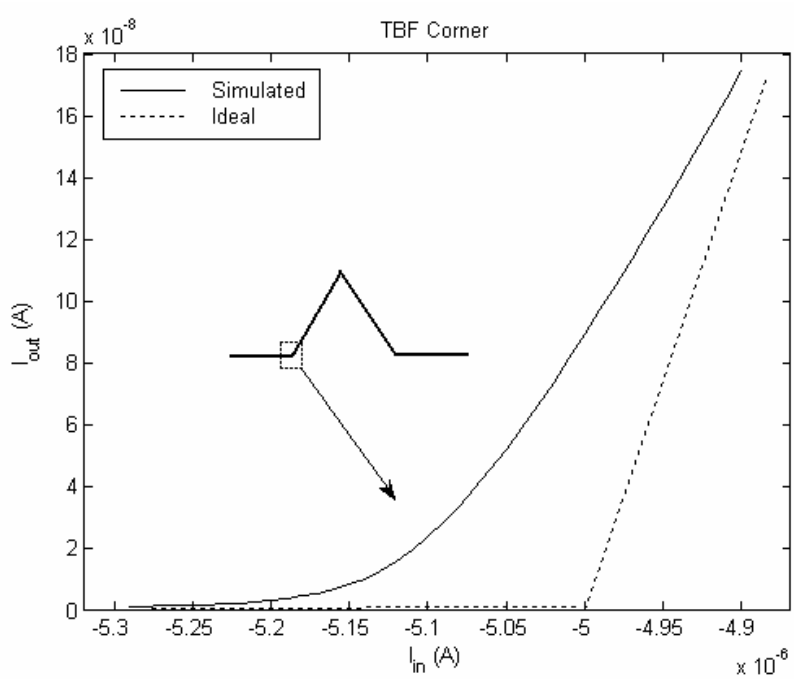


Figure 3-1 Comparison between simulated and ideal TBF corners.

In this criteria the derivative of the realized peace-wise linear function, with respect to the input current, will indicate the sharpness of the transition points. Ideally, at the transition points (break points), the line of the derivative should have a perpendicular characteristic. Perpendicular transitions of the derivative means ideal sharp transitions between the different curves of the realized functions.

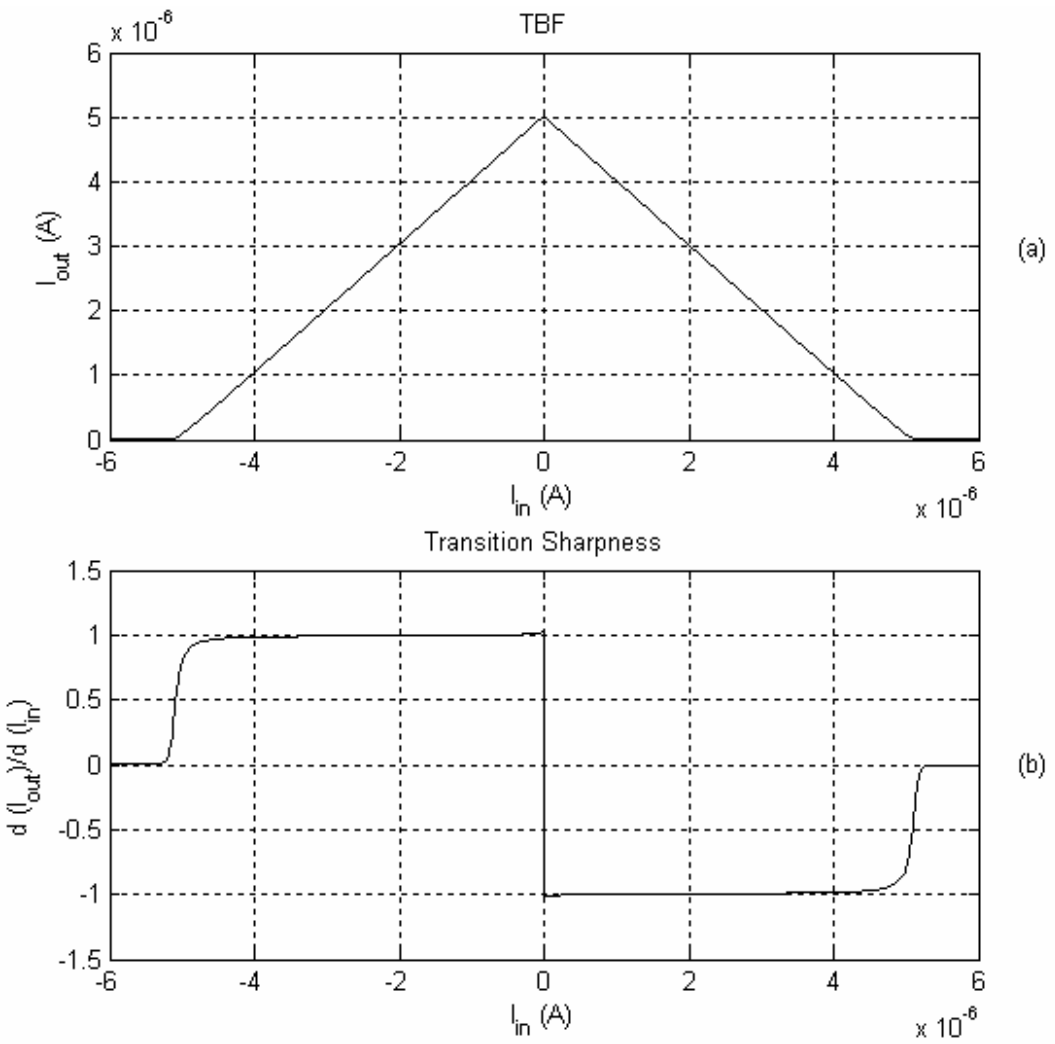


Figure 3-2 The TBF (a) and its derivative with respect to the input (b).

Figure 3-2 shows the current transfer function of the TBF and its derivative. For an ideal TBF, the derivative should have a rectangular shape along every straight line. As can be inferred from Fig. 3-2, a sharp and almost perfect transition is achieved for the peak transition point. Transition states of the side points are still satisfactory and are comparable with the results achieved in [35].

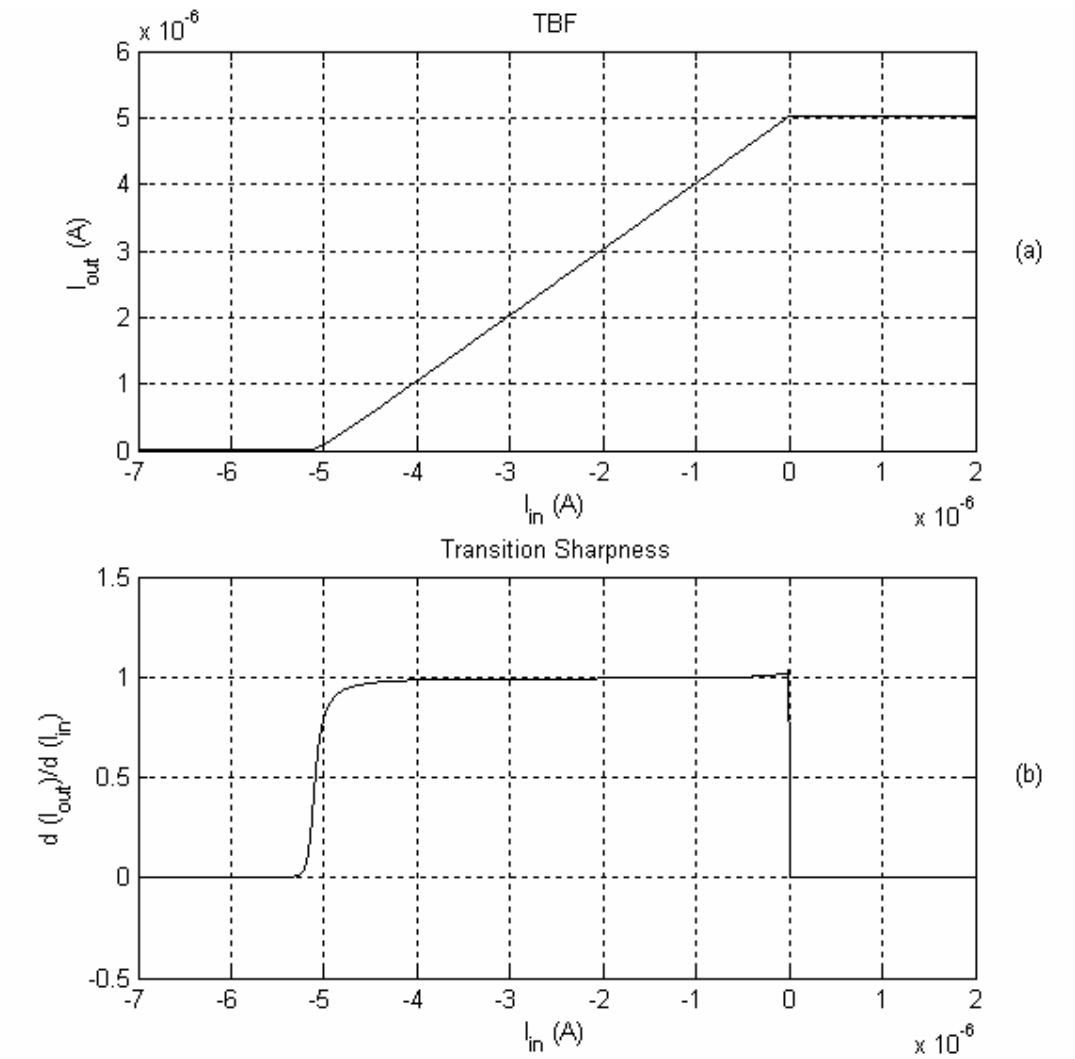


Figure 3-3 SatLin function (a) and its derivative (b).

Numerical comparison between the derivatives of the two side transition points (the side transition points of Fig. 3-2b and the equivalent figure of the circuit proposed in [35]) can be made in relation to the dynamic range. Figs. 3-2b and 3-3b reveal that transition between zero and 80% of the derivative's maximum value consumes approximately 0.3uA change in the input current. The ratio between this value and the dynamic range of the proposed circuit (which is around 85uA as will be seen later) is around 0.0035 (0.35%). While for [35] the transition between zero and 80% of the derivative's maximum value consumes approximately 0.16V change in the input voltage. The ratio between this value and the dynamic range of the reported circuit (which is around 3V) is around 0.053 (5.3%).

Figure 3-3 shows the current transfer function of the SatLin function and its derivative. From fig. 3-3 it can be seen that although a switching technique was used for the peak's break point, almost the same performance achieved for the TBF is achieved also for the SatLin function.

3. 2. Circuit Parameters

3. 2. 1. Bandwidth

In order to estimate the bandwidth of the proposed circuits, the circuit of Fig. 2-8 will be simulated with I_{DFT} = 0 and I_{REF}= 5uA. An AC current source will be used with AC current magnitude: ACMAG = 2.5 uA and Offset DC current: I_{DC} = 2.5 uA. The DC offset current is used to insure the operation in the rising or the falling edges of the functions. Two different configurations will be analyzed (TBF and RBF). Fig. 3-4 shows the frequency response of the TBF circuit (V_{SET} should be set to V_{SS}). Fig. 3-5 shows the frequency response of the RBF circuit (voltage sources V_{R1}, V_{R2}, and V_{SET} should be set to -0.2V, -0.3V and 0 voltages, correspondingly).

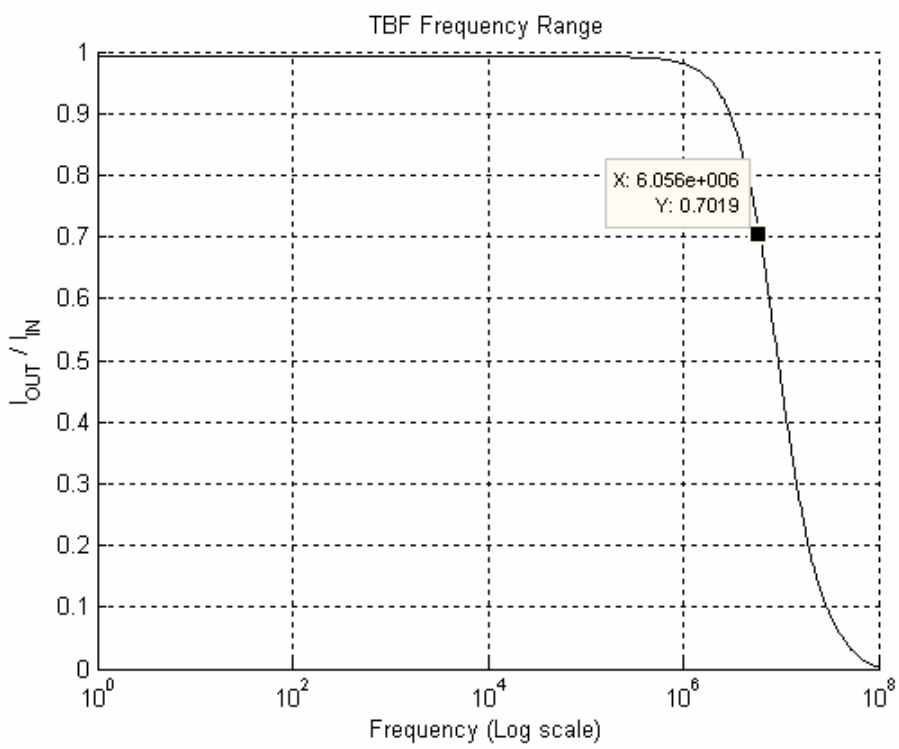


Figure 3-4 The TBF frequency response.

From Figure 3-4, the bandwidth is almost 6 MHz for the TBF implementation.

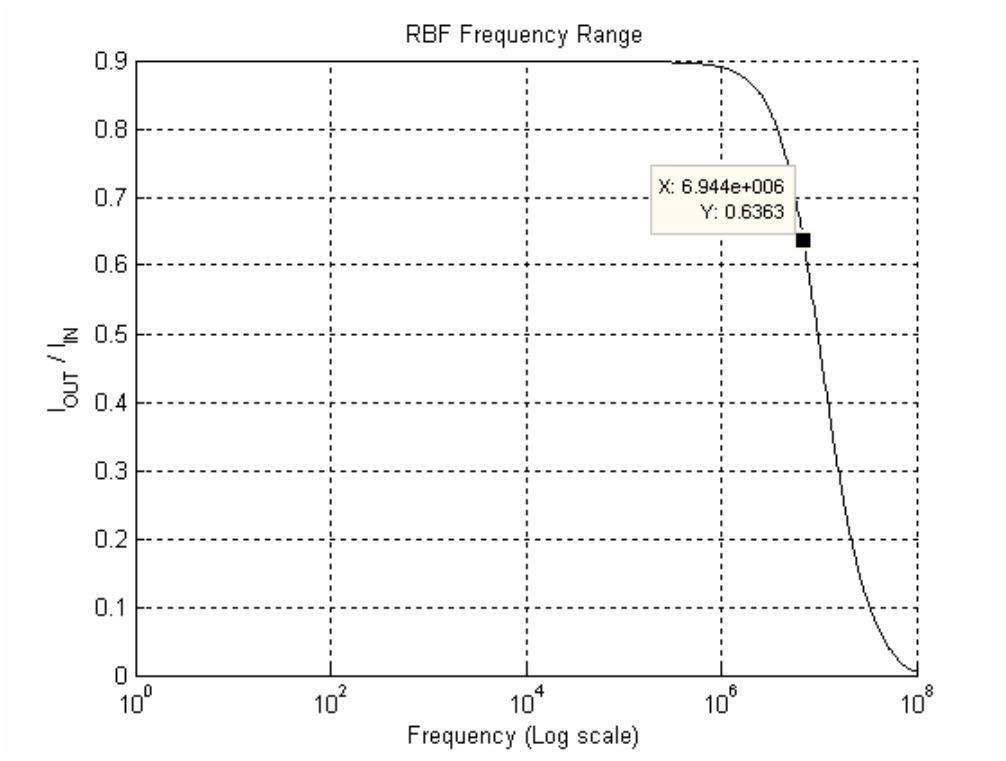


Figure 3-5 The RBF frequency response.

The corner frequency of the RBF function (Fig. 3-5) is increased to almost 7 MHz. Figure 3-6, which includes both functions' frequency responses, reveals that the increase in the frequency range of the RBF configuration compared to the TBF, is mainly due to the reduction of its magnitude (the magnitude of the RBF). This reduction in the magnitude is due to the DC currents of the peaking current sources.

Introducing MOSFET switches to the proposed circuit in Fig. 2-24 (S_1, S_2, S_3 and S_4), significantly reduces the frequency range of the implemented functions (around 100 kHz). Switches S_1 and S_3 are the most affecting between these switches (S_1 and S_3 were implemented using PMOS transistors).

Replacing the single PMOS transistors, realizing switches S_1 and S_3, by CMOS switches (Transmission Gates), avoids the dynamic range limitations associated with a single MOS switch. This results in increasing the frequency range of the implemented functions to almost 4 MHz for the TBF and 4.3 MHz for the RBF.

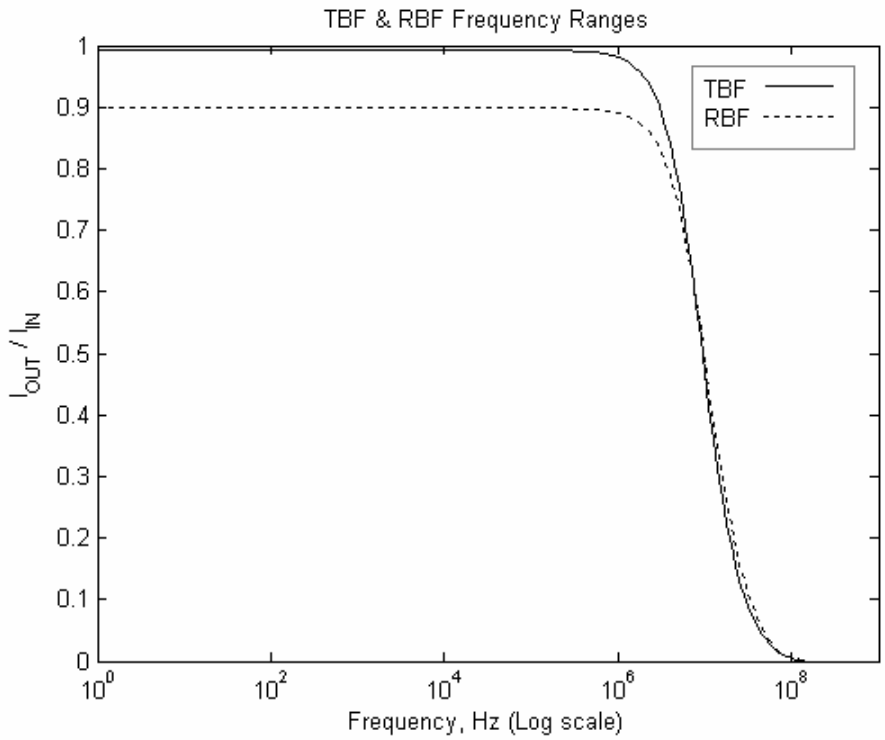


Figure 3-6 TBF and RBF frequency responses.

3. 2. 2. Dynamic Range

For normal operation of the circuit, most of the transistors should operate in the saturation region. Different factors can affect the transistors' operation region like the power supply, the transistors' dimension ratios and the magnitude of the current flowing through them. In order to estimate the dynamic range of the proposed circuit, we need to observe the operation region of the circuit's transistors for different values

of the input current. Different positive and negative values of I_{in} will be tried (I_{DFT} = 0). The case, when negative input current is applied, is more critical for the dynamic range. This is due to the fact that the negative input current passes from V_{DD} to V_{SS} through four different transistors (M_5, M_6, M_7, M_8), while the positive current passes mostly through three transistors (see Fig. 2-24). This additional (V_{DS}) voltage in the pathway of the negative input current reduces the maximum allowed current which can flow through this path without disturbing the operation region of the transistors. It was found from the simulation results that, for normal operation of the circuit, I_{in} should not exceed -93uA in the absolute value. Exceeding this value (I_{in} < -93uA), causes M_8 to operate in the triode region. As a result of this, the current delivering between the input and the output parts of the circuit starts deviating from the ideal case. When exceeding the value of -105uA, M_2 inters also the triode region, which affects more the current delivering between the different parts of the circuit. From the above mentioned observations it was concluded that, for normal and linear operation of the circuit, the input current should fall in the range between -93uA and +93uA (-93uA < I_{in} < +93uA, or, -93uA < I_{in}-I_{DFT} < +93uA, when I_{DFT} does not equal zero). The value of I_{REF} should also be restricted between these two values. To prove these results, the proposed circuit of Figure 2-24, configured as a TBF circuit, was simulated using these biasing parameters: I_{DFT} = 0, I_{REF} = 250 uA, V_{SET1} = V_{SS}, V_{SET2} = V_{DD} and V_{SWITCH_S} = V_{DD}. Figure 3-7 shows the simulation result of this configuration.

Symmetry and curves linearity will reveal the dynamic range of the proposed circuit.

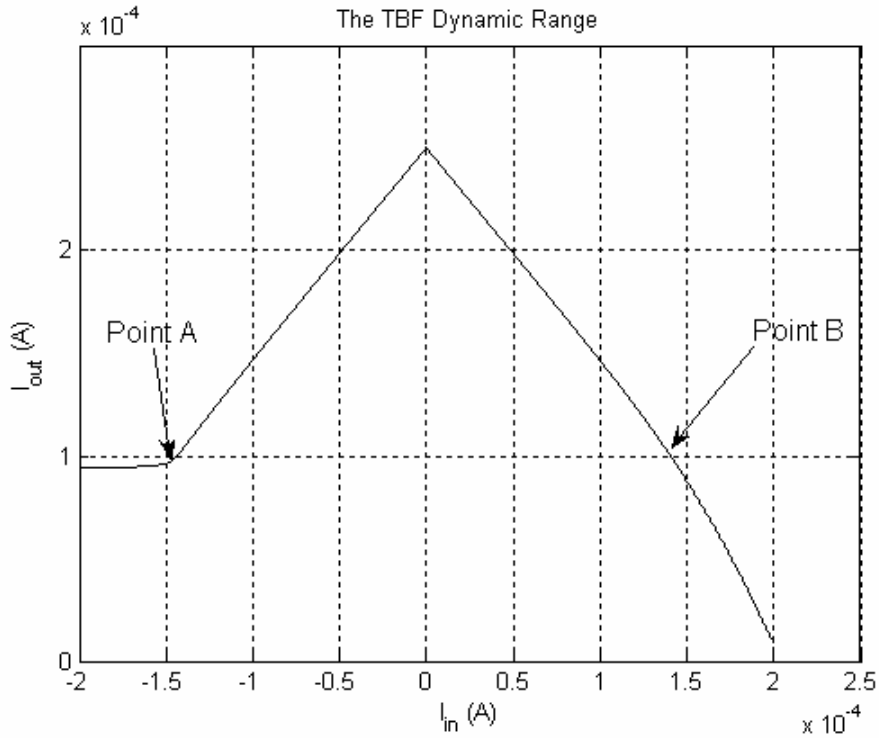


Figure 3-7 TBF realization for a wide range.

Fig. 3-7 reveals that there is a break point in the rising edge of the implemented function (Point A of Fig. 3-7). This means that the maximum useful value of the negative input current is -143 uA. Point B marks the region where the distortion or the deviation from linearity of the falling edge of the implemented function will start. In other words and after accurately observing the linear range of the implemented function, we can conclude that the dynamic range of the proposed circuit is almost $\pm$ 93uA (very little was mentioned about the dynamic range of the reported current-mode circuits, for example in [20] the dynamic range is 8uA).

A better dynamic range was stated in [27] (around $\pm$ 100uA), but it is clear that the proposed circuit of Fig. 2-24 is much better in terms of accuracy and programmability.

3. 2. 3. Power Consumption

Table 3-1 summarizes the circuit power consumption for different biasing conditions.

Table 3-1 Circuit power consumption

Realized Function	Power consumption, in uWatts, for different I_{DFT} values (I_{REF} = 5 uA, V_{DD} = +1V V_{SS} = -1V).				
	I_{DFT}=0 uA	I_{DFT}=5 uA	I_{DFT}=-5 uA	I_{DFT}=10 uA	I_{DFT}=-10 uA
TBF	35	45	35.1	60	40
RBF	43.7	56.4	46.4	80	60
Sigmoid	81.4	56.4	87.2	80	92.1
SatLin	72.7	45.1	78.5	60	83.3

From Table 3-1 it is obvious that when the circuit is used to implement Sigmoid or SatLin functions the power consumption increases. This excess results mainly from activating the Nor-Gate.

3. 2. 4. Input and Output Resistances

In this section, input and output resistances of the proposed circuit will be evaluated. Figure 3-8 shows the input and the output routes of the current. Fig. 3-8a assumes a positive input current (the upper transistors will be off).

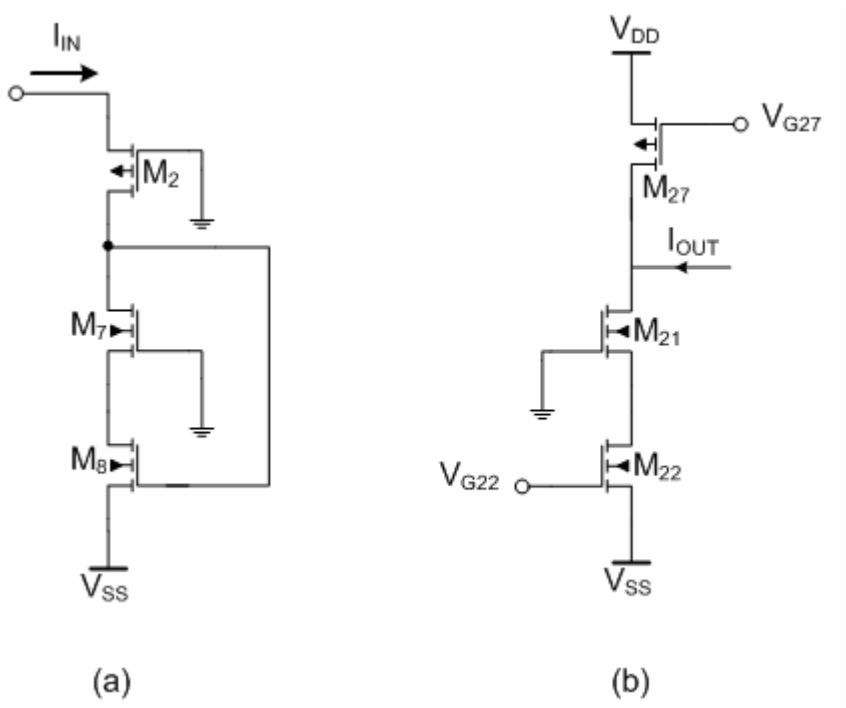


Figure 3-8 **a)** The input current route and **b)** The output current route.

Using Fig. 3-8b, an approximate expression of the output resistance is found for a case when a piecewise-linear function is implemented (M_{27} is off), and for a case when a nonlinear function is implemented (M_{27} is on), respectively:

$$r_{out} = (g_{m21} r_{ds21}) r_{ds22} \tag{3-1}$$

and

$$r_{out} = (g_{m21} r_{ds21}) r_{ds22} \parallel r_{ds27}, \tag{3-2}$$

where g_m is the transconductance, and r_{ds} is the channel resistance of the transistor. gm and gds (the small signal channel conductance of the transistor, $g_{ds} = 1/r_{ds}$) of all transistors are given by the simulation program: g_{m21} = 232 uA/V, g_{ds21} = 1.1 uA/V, g_{ds22} = 1.03 uA/V, g_{ds27} = 0.0636 uA/V .

Substituting these results in equations 3-1 and 3-2 gives:

$r_{out} = (g_{m21} r_{ds21}) r_{ds22} = 204.7 \ M\Omega$, (M_{27} is off).

$r_{out} = (g_{m21} r_{ds21}) r_{ds22} \| r_{ds27} = 14.58 \ M\Omega$, (M_{27} is on).

These results were found using $I_D(M_{21}) = I_D(M_{22}) = 10$ uA and $I_D(M_{27}) = 1.25$ uA.

The output resistances, calculated by the simulation program, are given next:

$r_{out} = (g_{m21} r_{ds21}) r_{ds22} = 256.3 \ M\Omega$, (M_{27} is off).

$r_{out} = (g_{m21} r_{ds21}) r_{ds22} \| r_{ds27} = 14.8 \ M\Omega$, (M_{27} is on).

The simulated results are very close to the approximated counterparts. This proves the validity of our approach. It is also clear from these results that, activating the peaking current sources (which activates M_{27}) results in significant degradation in the output resistance value. This problem can be solved by cascading M_{27} with another transistor (r_{out} can be made greater than 100 $M\Omega$).

Using Fig. 3-8a, an approximate expression for the input resistance can be found (refer to appendix B for the derivation and the detailed expression of r_{in}):

$$r_{in} \cong \frac{g_{m7} r_{ds7} r_{ds8} (g_{m8} r_{ds2})}{g_{m8} r_{ds8} (g_{m2} r_{ds2})(g_{m7} r_{ds7})} = \frac{1}{g_{m2}} \qquad (3\text{-}3)$$

Using the parameters, calculated by the simulation program (see Table B-1), and substituting it in equation 3-3 results in:

$r_{in} \cong 5.8 \ k\Omega$.

The input resistance, calculated by the simulation program, is: $r_{in} = 4.6\ k\Omega$. The simulated result is close to the approximated one, which proves the validity of our approximation. This value of the input resistance is considered a relatively high resistance. An optimization process of the circuit revealed that increasing the dimension ratio (W/L) of transistor M_2 can reduce a little the input resistance of the circuit. But this in turn, will increase the area of the chip.

3. 3. Temperature Variation effect

In this section, the proposed circuit is simulated to observe the effect of temperature variations on the accuracy of the implemented function. The temperature was varied between $-70\ ^{\circ}C$ and $+50\ ^{\circ}C$. I_{REF} = 5uA and I_{DFT} = 0. Simulation results reveal that temperature variations have a clear effect on the circuit performance when it is configured to implement a nonlinear function (RBF or Sigmoid functions). This is mainly because the fact that at least ten more transistors are activated when the proposed circuit is configured to implement a nonlinear function. As the Sigmoid function is simply a part of the RBF, we will consider temperature variations effects only on the realized RBF function (see Table 4-2 for biasing details). Figure 3-9 shows the simulation results of the RBF for different temperature values.

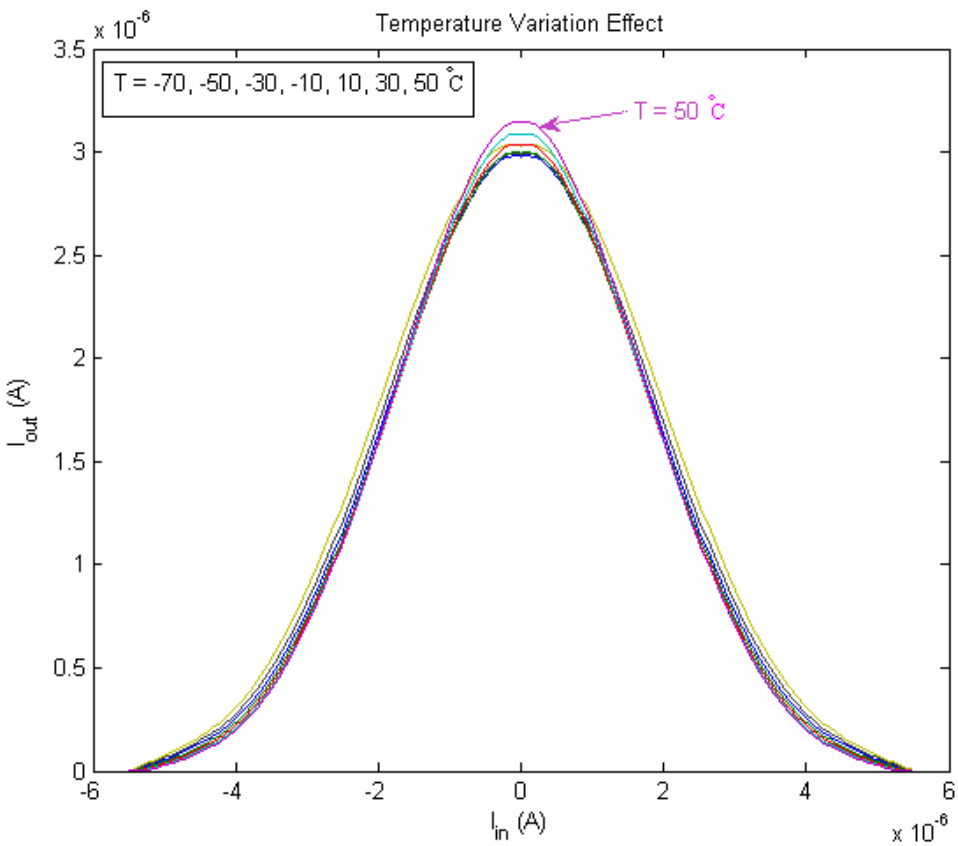


Figure 3-9 Temperature variation effect on the realized RBF.

By comparing the Peak values of RBF functions, realized at different temperatures, one can see that the peak is increasing by 2.17% when increasing the temperature from the nominal value $(T = 27\ ^{\circ}C)$ to a temperature $(T = 50\ ^{\circ}C)$. When decreasing the temperature from the nominal value $(T = 27\ ^{\circ}C)$ to the value $(T = -70\ ^{\circ}C)$, the peak is decreasing by only 1.33% of the nominal peak value. These variations in temperature result in a change in the peak currents equal to 67 nA and 41 nA, respectively (3.5% change in the peak for the whole temperature range).

3. 4. Effects of Technology Parameters Variation

A good design should not be affected significantly by small variations in the design technology parameters (W/L, K_P and V_{th}). The effect of 1% change in technology parameters on the accuracy of the circuit's realized functions will be analyzed.

Two parameters are explicitly defined in the MOSFET model used for simulation (see Appendix A for model details). These two parameters are VTH0 and K1. TBF and RBF circuits are simulated (using I_{DFT} = 0 and I_{REF} = 5uA) with 1% deviation in these two model parameters. Monte Carlo analysis is the most suitable analyzing tool for such problems.

3. 4. 1. TBF

Fig. 3-10 shows the Monte Carlo analysis of the TBF circuit (see Table 4-2 for the biasing details), for 1% VTH0 deviation. Simulation results show a deviation in the peak value between 4.58 uA and 5.2 uA, which corresponds to a 12% change in the nominal value (5 uA).

Fig. 3-11 shows the Monte Carlo analysis of the TBF with 1% deviation in the parameter K1.

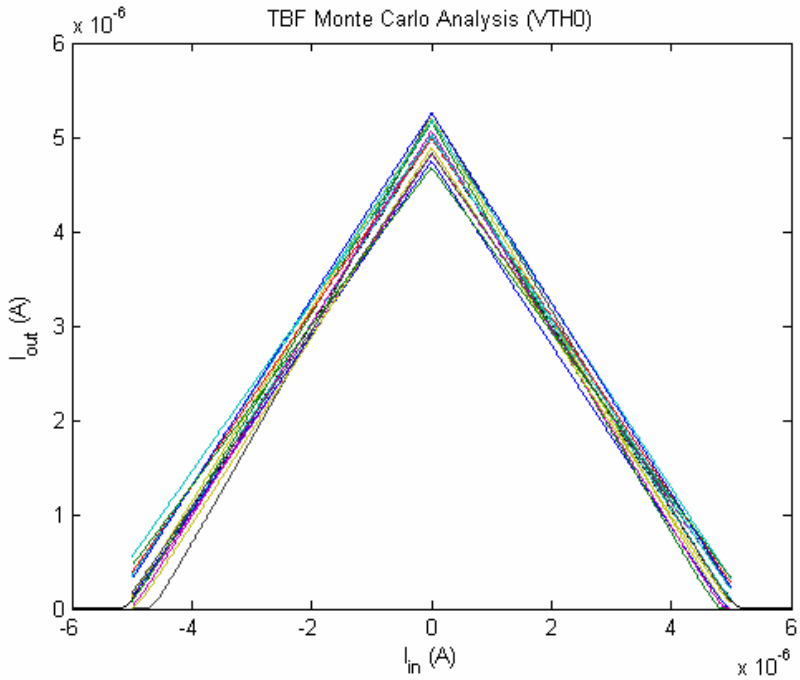


Figure 3-10 Monte Carlo analysis of the TBF with 1% VTH0 deviation.

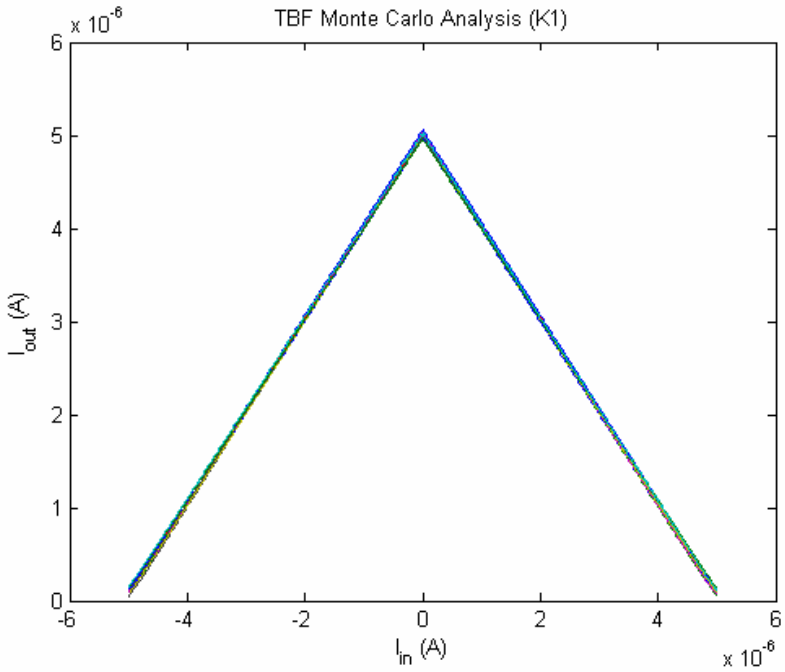


Figure 3-11 Monte Carlo analysis of the TBF with 1% K1 deviation.

From Fig. 3-11 it appears that the deviation in the peak value is 4.9 and 5.01 uA which corresponds to a change in 2% of the nominal value. It is clear that VHT0 deviation has more considerable effect on the accuracy of the implemented function than K1 deviation. However the 12% peak deviation of the TBF (which results from 1% deviation in VTH0), does not mean a large deviation from the TBF function. As can be observed from figure 3-10, mainly, the magnitude of the function is affected (almost negligible distortion in the linearity or the symmetry of the function's curves is observed).

3. 4. 2. RBF

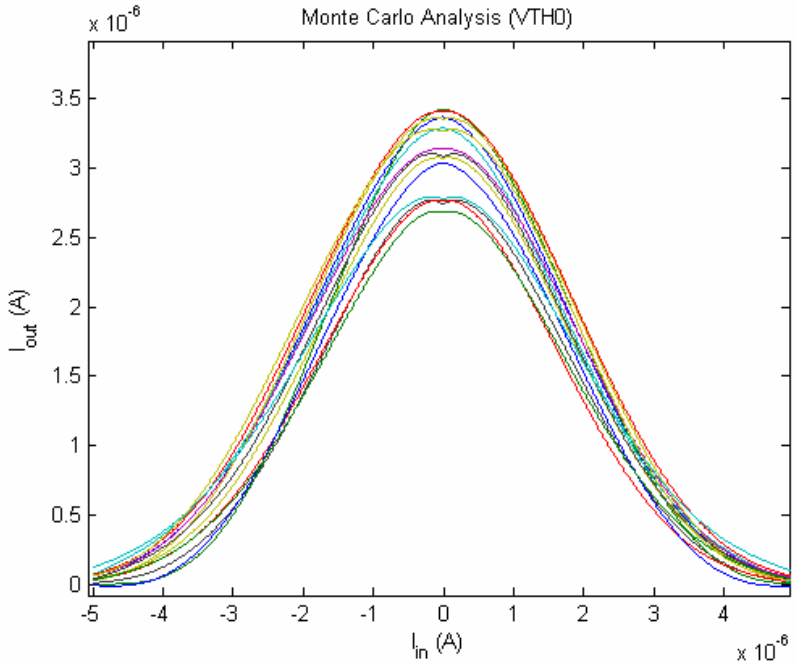


Figure 3-12 Monte Carlo analysis of the RBF with 1% VTH0 deviation.

The circuit of Fig. 2-24, configured as a RBF circuit (refer to Table 4-2), was simulated to investigate the effect of variations in K1 & VTH0. Fig. 3-12 shows the Monte Carlo analysis of the RBF with 1% VTH0 deviation. From Fig. 3-12 it appears that the deviation in the peak value is 2.68 and 3.41 uA which corresponds to a change in 22% of the nominal value (3 uA). Here also we observe that the main effect of this deviation is on the peak of the function. The function is still symmetrical. Standard deviation is changing by almost the same percent as the peak. So it is expected that the 1% change in VTH0 will result in a small amount of deviation from the Gaussian approximation. Fig. 3-13 shows the Monte Carlo analysis of the RBF, with 1% deviation of K1. 3.5% change of the nominal value is observed.

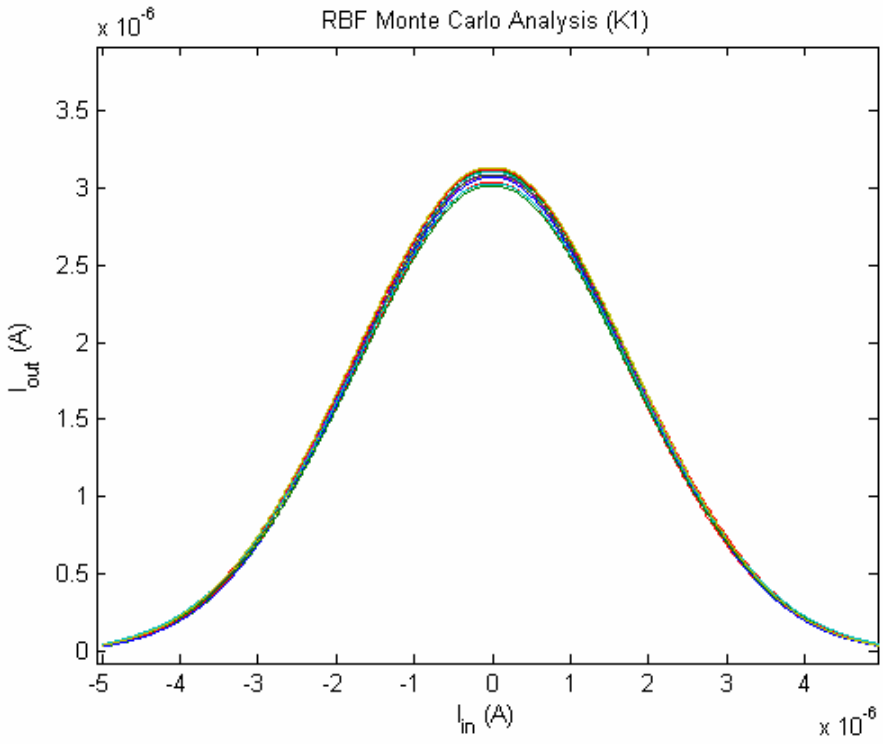


Figure 3-13 Monte Carlo analysis of the RBF with 1% K1 deviation.

So far, Orcade Cadence Pspice 9.2 program has been used for simulations. But this program does not offer the possibility to change W or L parameters of the transistors using Monte Carlo analysis. As an alternative, Hspice program does offer this possibility and has been used to simulate the effect of 1% deviation of W and L parameters. Figures 3-14 and 3-15 show Monte Carlo analyses for the TBF and the RBF, respectively. They were simulated using the same biasing conditions as in the previous simulations. The peak of the TBF function is varying between 5.1 uA and 4.9 uA which means a possible 2% deviation from the nominal value. The peak of the RBF function is varying between 2.94 uA and 3.17 uA which means a possible 4% deviation from the nominal value (which is around 3.05).

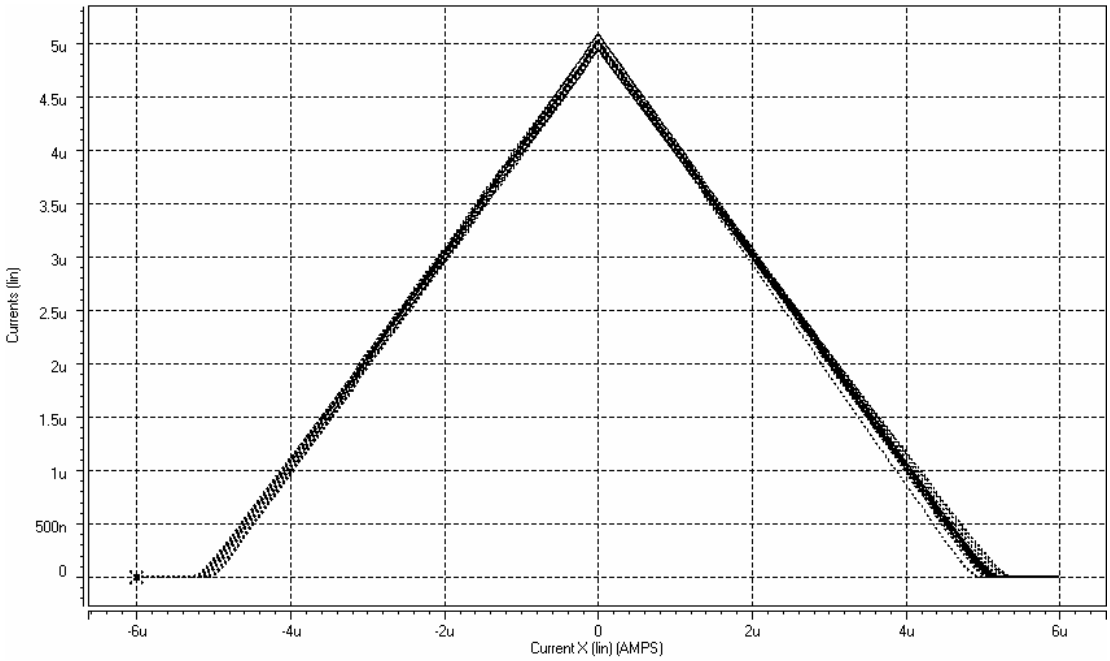


Figure 3-14 Monte Carlo analysis of the TBF.

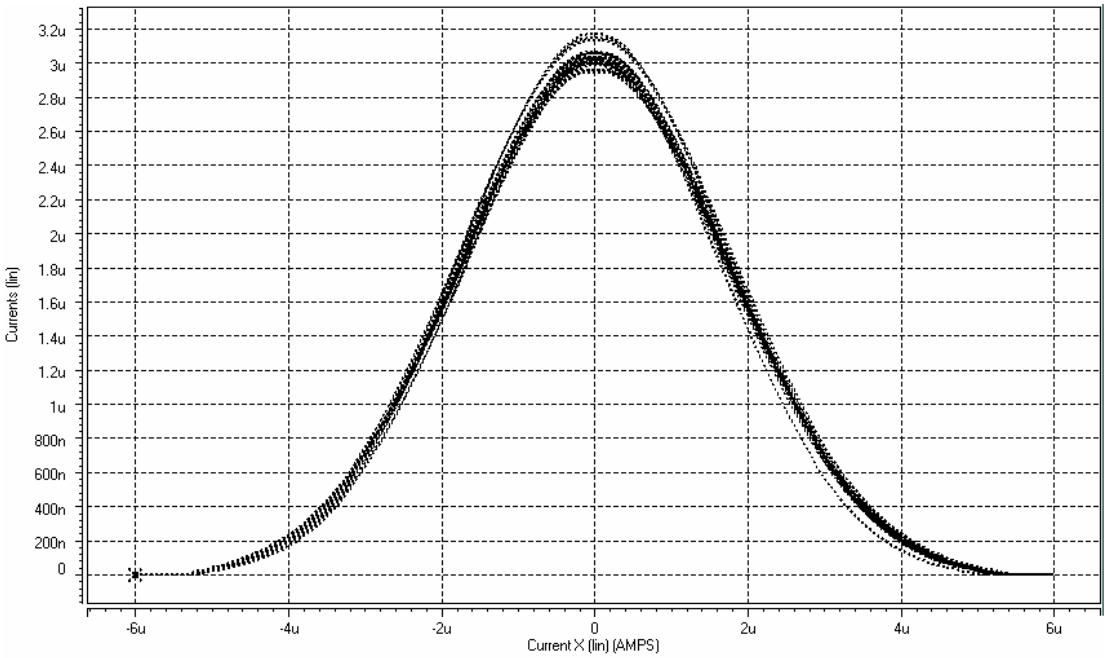


Figure 3-15 Monte Carlo analysis of the RBF.

CHAPTER 4

Conclusion and Future Work

4. 1. Conclusion

In this thesis, a reconfigurable CMOS generic circuit for Neural Networks applications was proposed. Its general specifications are summarized in Table 4-1.

Table 4-1 Circuit's general specifications

Technology of implementation	0.18 u CMOS (TSMC)
Realized Functions	TBF, RBF, Sigmoid and SatLin
Input-Output Signal Representation	Current – Current
CMOS Operation Mode	Strong Inversion
Interpolation Data (Current polarity)	Positive

The circuit is configurable and can realize any of the four functions, using the same input-output terminals, by setting the voltage sources V_{SET1} , V_{SET2} and V_{SWITCH_S} as given in table 4-2.

Table 4-2 Functions configuration settings

Realized Function	V_{SET1_1}	V_{SET1_2}	V_{SWITCH_S}	V_{SET2}
TBF	V_{SS}	V_{SS}	V_{DD}	V_{DD}
RBF	0 V	0 V	V_{DD}	V_{DD}
Sigmoid	0 V	V_{SS}	V_{SS}	V_{SS}
SatLin	V_{SS}	V_{SS}	V_{DD}	V_{SS}

A special concern was paid in this work on the achievement of good programmability of the functions parameters. Programmability is achieved for all parameters (peak, threshold, slope and standard deviation), using some voltage and current sources.

Table 4-3 summarizes the main characteristics of different function implementations.

Table 4-3 Realized functions circuit parameters

Realized function	Accuracy of the implemented function (typical)	Temperature variation effects on the peak amplitude $T^{\circ} = -70 \div 50^{o}C$	Power Consumption (uWatts) $I_{REF} = 5\mu A$ $I_{DFT} = -10 \div 10\mu A$	Bandwidth
TBF	99.3%	Less than 1%	$35 \div 60$	6 MHz
RBF	99.2%	3.5%	$43.7 \div 80$	6.9 MHz
Sigmoid	98.8%	3.5%	$56.4 \div 92.1$	4 MHz
SatLin	99.4%	Less than 1%	$45.1 \div 83.3$	4 MHz

TBF and RBF functions are implemented with high accuracy (the best reported accuracy was 98% [43]), relatively low power consumption and good frequency range. These advantages make this circuit suitable for VLSI design. Sigmoid and SatLin realizations have some disadvantages, mainly the smaller bandwidth due to the use of transistor switches, and the higher power consumption compared to the other functions implementations.

The effects of temperature and technology parameters variations were also considered and analyzed in this thesis. It was found that the proposed circuit was significantly affected by the technology parameters variations, especially by the variation of the threshold voltage, VTH0. The effects were mainly on the magnitude of the functions (the distortion of the implemented functions was insignificant).

The main disadvantages of the proposed circuit are believed to be:

- Relatively high input resistance (around $5k\Omega$ for $I_{in} = 10\mu A$).
- The use of transistor switches, which results in degrading the frequency range.
- One direction of the output current (I_{OUT} and I_{REF} are always positive).
- Relatively large number of transistors, exploited in the circuit.

4. 2. Future Work

Future work, mainly, should be concentrated on improving the circuit in different aspects like frequency range of the unsymmetrical functions (SatLin and Sigmoid), bipolarity of the output current and finally lowering the input resistance of the circuit.

Different suggestions are proposed for future work:

- The output resistance of the circuit is sufficiently large (several $M\Omega$). But it is a little bit degraded when the circuit is configured to implement a nonlinear function. Cascading the output transistors of the peaking current-sources will minimize this effect.
- Different techniques are proposed in the literature for lowering the input resistance of current mode circuits. Some of them might be applicable and useful for the proposed circuit, like using class AB current mirrors with negative feedback (chapter 5 of [54]), or using a negative current-parralel feedback [55] to reduce the input resistance and increase the output resistance.
- Probably a different and more effective approach may be used for Sigmoid and SatLin functions realization. A different approach using less number of circuits and switches is preferable.
- In the future design, stacking transistors should be avoided in order to reduce the area and also reduce the power supply.

Appendix A

MOSFET BSIM3V3 Transistor Model Parameters

RUN: T16X (MM_NON-EPI) VENDOR: TSMC

TECHNOLOGY: SCN018 FEATURE SIZE: 0.18 microns

T16X SPICE BSIM3 VERSION 3.1 PARAMETERS

ORCADE 9.2 Level 7, Star-HSPICE Level 49

Table A-1 Transistor Model Parameters

Parameter's Name	Value	Parameter's Name	Value
TNOM	27	TOX	4.1E-9
XJ	1E-7	NCH	2.3549E17
VTH0	0.3581698	K1	0.574024
K2	2.751715E-3	K3	1.959368E-3
K3B	2.2040222	W0	7.371562E-7
NLX	1.768395E-7	DVT0W	0
DVT1W	0	DVT2W	0
DVT0	1.4705192	DVT1	0.4151006
DVT2	0.0343357	U0	296.2894586
UA	-6.93439E-10	UB	1.32165E-18
UC	-1.35144E-11	VSAT	9.146457E4
A0	1.7920403	AGS	0.3415021

Parameter's Name	Value	Parameter's Name	Value
B0	-1.763016E-8	B1	-1E-7
KETA	6.109641E-3	A1	2.006795E-4
A2	0.9923701	RDSW	127.7755888
PRWG	0.5	PRWB	-0.2
WR	1	WINT	0
LINT	9.512723E-9	XL	-2E-8
XW	-1E-8	DWG	-1.647691E-9
DWB	-7.387757E-9	VOFF	-0.0720847
NFACTOR	2.4126738	CIT	0
CDSC	2.4E-4	CDSCD	0
CDSCB	0	ETA0	0.0603469
ETAB	-0.0640255	DSUB	1
PCLM	0.8413441	PDIBLC1	0.0737257
PDIBLC2	0.01	PDIBLCB	-0.0946392
DROUT	0.5318923	PSCBE1	7.990582E10
PSCBE2	2.575736E-8	PVAG	4.297626E-3
DELTA	0.01	RSH	6.5
MOBMOD	1	PRT	0
UTE	-1.5	KT1	-0.11
KT1L	0	KT2	0.022
UA1	4.31E-9	UB1	-7.61E-18

Parameter's Name	Value	Parameter's Name	Value
UC1	-5.6E-11	AT	3.3E4
WL	0	WLN	1
WW	0	WWN	1
WWL	0	LL	0
LLN	1	LW	0
LWN	1	LWL	0
CAPMOD	2	XPART	0.5
CGDO	7.23E-10	CGSO	7.23E-10
CGBO	1E-12	CJ	9.92536E-4
PB	0.7270294	MJ	0.3574892
CJSW	2.47496E-10	PBSW	0.5750347
MJSW	0.1322155	CJSWG	3.3E-10
PBSWG	0.5750347	MJSWG	0.1322155
CF	0	PVTH0	-3.36027E-4
PRDSW	-5	PK2	-9.513629E-4
WKETA	2.169006E-3	LKETA	-9.246664E-3
PU0	22.0242664	PUA	8.96812E-11
PUB	1.210283E-24	PVSAT	1.648121E3
PETA0	1E-4	PKETA	2.23841E-3

Appendix B

Derivation of the Circuit's Input Resistance

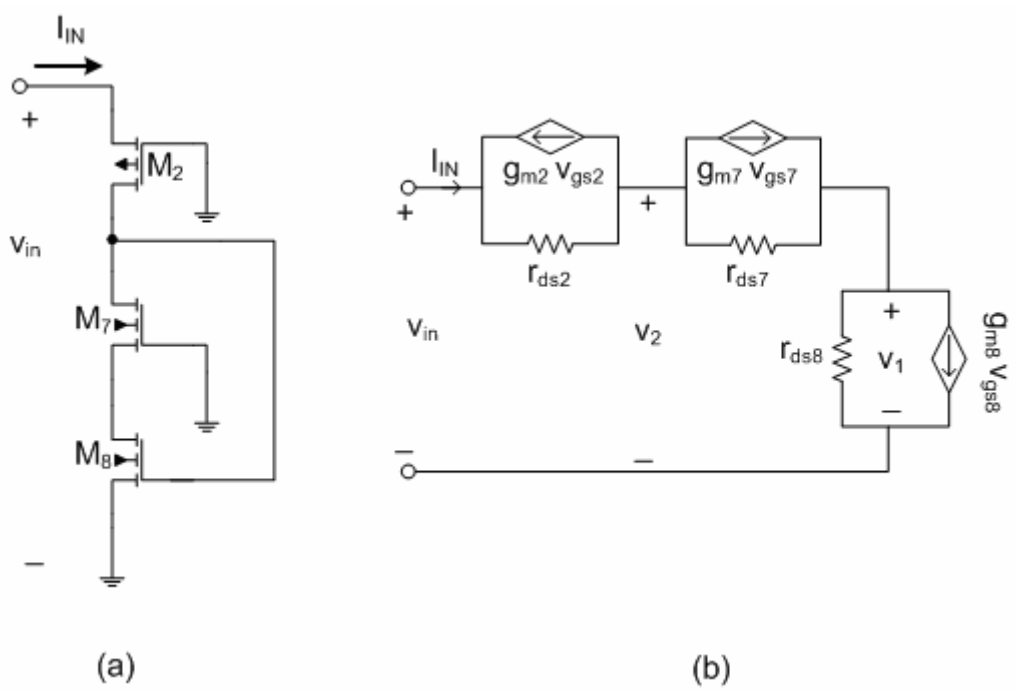


Figure B-1 a) the input circuit's route of the positive input current (after replacing the DC voltages by Ground terminals), b) The AC equivalent circuit.

To find an expression for the input resistance, we need to find V_{in}/I_{in} in terms of channel resistances and transconductances of the transistors, realizing the circuit given in Fig. B-1. We will start by finding an expression for V_{in} using Fig. B-1b:

$$v_{in} = r_{ds2}(i_{in} + g_{m2}v_{gs2}) + r_{ds7}(i_{in} - g_{m7}v_{gs7}) + r_{ds8}(i_{in} - g_{m8}v_{gs8}) \qquad \text{B-1}$$

Now, using figures B-1a and B1-b, expressions for the Gate-Source voltages will be found and will be substituted in Equation B-1 for the final expression of V_{in}.

$$v_{gs2} = -v_{in} \qquad \text{B-2}$$

$$v_{gs8} = v_2 = v_{in} - r_{ds2}(i_{in} + g_{m2}v_{gs2}) \qquad \text{B-3}$$

Substituting B-2 in B-3 results in:

$$v_{gs8} = v_{in} - r_{ds2}(i_{in} - g_{m2}v_{in}) = v_{in}(1 + g_{m2}r_{ds2}) - i_{in}r_{ds2} \qquad \text{B-4}$$

$$v_{gs7} = 0 - r_{ds8}(i_{in} - g_{m8}v_{gs8}) = r_{ds8}g_{m8}v_{gs8} - r_{ds8}i_{in} \qquad \text{B-5}$$

Substituting B-4 in B-5 results in:

$$v_{gs7} = g_{m8}r_{ds8}v_{in}(1 + g_{m2}r_{ds2}) - i_{in}r_{ds8}(g_{m8}r_{ds2} + 1) \qquad \text{B-6}$$

After substituting B-2, B-4 and B-6 in B-1, we will end up by a relation including input current and voltage, and transistors' transconductances and channel resistances. After rearranging the different terms of this relation, we will end up by a relation between the input voltage and the input current (which represents the input resistance of the circuit):

$$\frac{v_{in}}{i_{in}} = R_{in} = \frac{r_{ds2} + r_{ds7} + r_{ds8} + g_{m7}r_{ds7}r_{ds8}(1 + g_{m8}r_{ds2}) + g_{m8}r_{ds8}r_{ds2}}{1 + g_{m8}r_{ds8}(1 + g_{m2}r_{ds2})(1 + g_{m7}r_{ds7}) + g_{m2}r_{ds2}} \qquad \text{B-7}$$

Using the simulation values given in Table B-1 (these values were found by the simulation program for I_{in} = 10uA), and using the relation $r_{ds}=1/g_{ds}$, we can calculate R_{in}: $R_{in} \cong 5.4\ k\Omega$.

Next, to find an approximated expression for B-7, the small terms will be neglected in both the numerator and the denominator of this relation:

$$R_{in} \cong \frac{g_{m7} r_{ds7} r_{ds8} (g_{m8} r_{ds2})}{g_{m8} r_{ds8} (g_{m2} r_{ds2})(g_{m7} r_{ds7})} = \frac{1}{g_{m2}} = 5.8 \ k\Omega \qquad \text{B-8}$$

Table B-1 Transistors' small-signal parameters calculated by the simulation program.

Transistor Number	Transconductance, g_m (uA/V)	Channel Conductance, g_{ds} (uA/V)
M_2	170	0.458
M_7	205	15.7
M_8	181	0.952

The input resistance, calculated by the simulation program, is: $r_{in} = 4.6 \ k\Omega$. The simulated result is close to the approximated one.

Appendix C

Layout Design and Spice File Extraction

Magic program has been used for drawing the layout of one of the proposed circuits. A technology file - SCN6M_SUBM.10.TSMC – corresponding to TSMC18 process has been used for the layout design. The lambda of this process is 0.1um. Figure C-1 shows the resulting layout (interdigitized drawing) of the circuit given in Fig. 2-8.

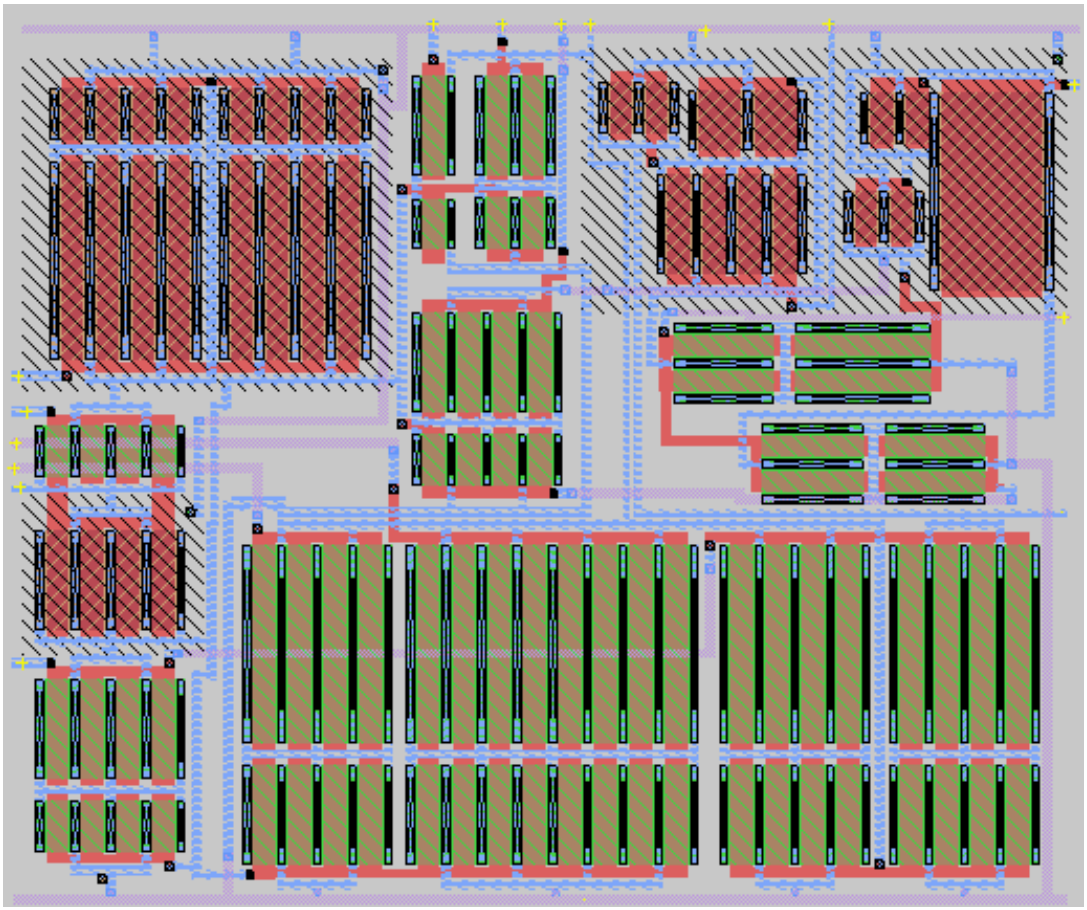


Figure C-1 The layout of the circuit depicted in Fig. 2-8.

Interdigitized way, for drawing transistors, were used to minimize the dimensions of the resulting layout. The resulting dimensions of this layout are approximately 55um for the width and 45um for the height. Spice file, with 107 transistors and 31 parasitic capacitances, was extracted from this layout and used for simulations. All extracted transistors dimensions were identical to the desired ones, except for transistor M23 (the desired W dimension was 13.5um, and the resulting one from the layout was 13.6um). This change in M_{23} dimension did not affect the previously simulated TBF, because of the fact that this transistor is activated only when RBF is realized. The effect of this change on RBF realization is depicted in Figure C-2.

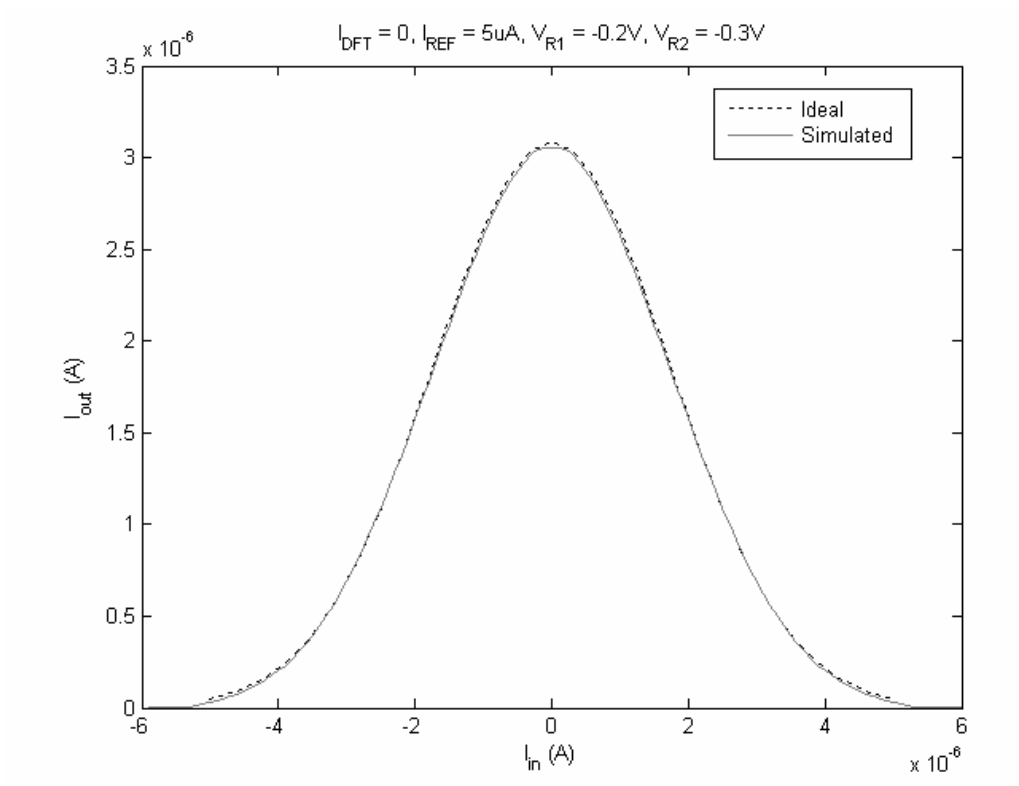


Figure C-2 The realized RBF using the parameters extracted from the layout.

Figure C-2 reveals that the error of the simulated function is increased a little compared to the error of the simulated function in Figure 2-9a. But the RRMS error still small and around 1%. This increase in the error results from the small reduction of the maximum amplitude of the realized function (this reduction results from increasing the width (W) of the transistor M_{23}).

The extracted Spice file is also used to analyze the frequency response of the circuit. Figure C-3 shows the frequency response depicted previously in Fig. 3-5 plus the frequency response of the implemented function after using the parameters extracted from the layout. Because of the parasitic capacitances, the frequency range of the circuit is decreased a little (around 6.85 MHz).

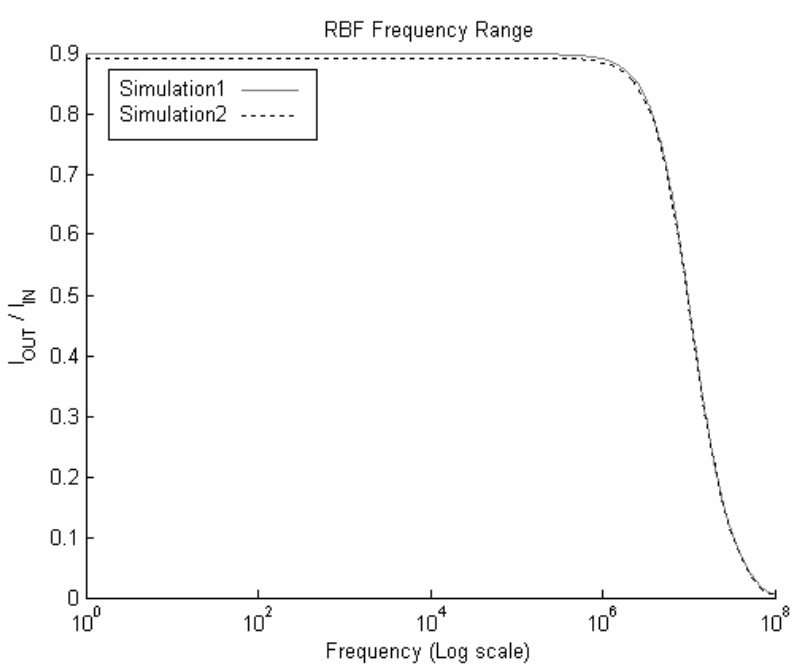


Figure C-3 The frequency range of the realized RBF function.

The extracted Spice file does not have a clear effect on the frequency range of the TBF function (for the TBF realization, V_{SET} of Fig. 2-8 is set to V_{SS}). Figure C-4 shows the frequency response of the TBF, depicted previously in Fig. 3-4, plus the frequency response of the TBF, simulated using the parameters extracted from the circuit layout (with dashed line).

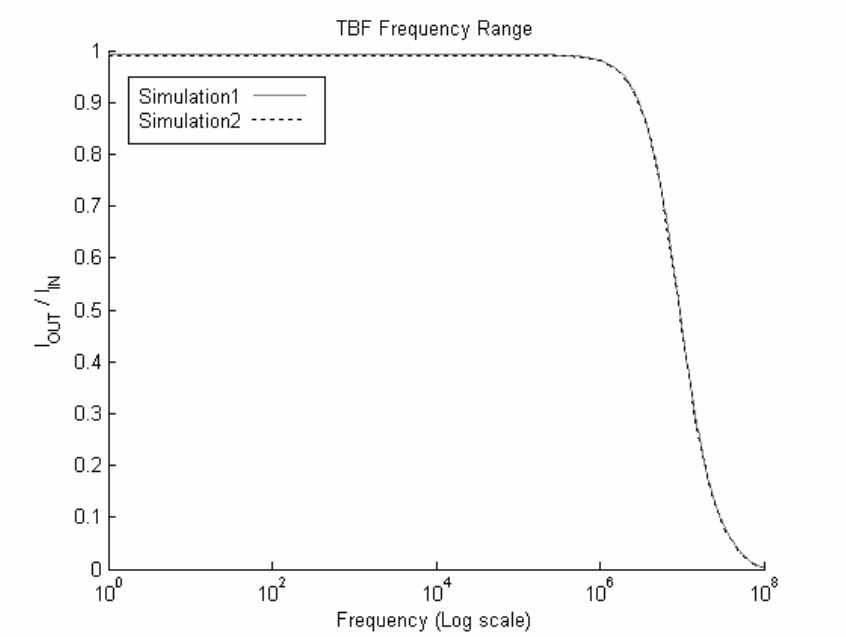


Figure C-4 The frequency range of the realized TBF function.

The frequency range of the circuit is still around 6 MHz. At the end, and from the simulation results, performed using the extracted parameters from the layout, it can be concluded that only a small effect has been observed on the accuracy of the realized RBF function. It can be also concluded that almost all transistors' dimensions were accurately extracted, thanks to the technology used in the layout design (TSMC 0.18um, and lambda=0.1um).

References

[1] Simon Haykin, *Neural Networks: a Comprehensive Foundation*, New Jersy: Prentice Hall, 1999.

[2] P. D. Wasserman, *Advanced Methods in Neural Computing*, New York: Van Nostrand Reinhold, 1993.

[3] J. Choi, Bing J. Sheu, "VLSI design of compact and high-precision analog neural network processors" *Proc. IEEE/INNS Inter. Joint Conf. Neural Networks*, vol. 2. pp. 637-641, July 1992.

[4] Sudhir M. Gowda, Bing J. Sheu, "Design and Characterization of Analog VLSI Neural Network Modules" *IEEE J. of Solid-State Circuits*, vol. 28, no. 3, pp. 301-312, March 1993.

[5] Bing J. Sheu, Joongho Choi, *Neural Information Processing and VLSI*, Kluwer Acadimic Puplishers, 1994.

[6] M. Masmoudi, M. Samet, F. Taktak, "A Hardware Implementation of Neural Network for the Recognition of printed Numerals" *International Conference on Microelectronics*, 1999. ICM '99, pp. 113 - 116.

[7] C. Mead, *Analog VLSI and Neural Systems*, Reading, PA: Addison-Wesley, 1989.

[8] A. F. Murray, "Silicon implementations of neural networks," *IEE Proceeding of Radar and Signal Processing*, vol. 138, no. 1, pp. 3-12, Feb. 1991.

[9] S. Stayanarayana, Y. P. Tsivids, H. P. Graf, "A Reconfigurable VLSI Neural Network" *IEEE Journal of Solid-State Circuits*, vol. 27, pp. 67-81, Jan 1992.

[10] A. F. Murray, Robin Woodburn, "The Prospects for Analogue Neural VLSI," *International Journal of Neural Systems*, vol. 8, nos. 5 & 6, pp. 559-579, 1997.

[11] V. F. Koosh and R. M. Goodman, "Analog VLSI neural network with digital perturbative learning," *IEEE Transactions on Circuits and Systems II : Analog and Digital Signal Processing*, vol. 49, no. 5, pp. 359-368, May 2002.

[12] Bernabe Linares-Barranco, Edgar Sanchez-Sinencio, "A Modular T-Mode Design Approach for Analog Neural Network Hardware implementation," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 5, May 1992.

[13] E. Sanchez-Sinencio, J. Ramirez-Angulo, B. Linares-Barranco, A. Rodriguez-Vazquez, "Operational transconductance amplifier-based nonlinear function synthesis," *IEEE Journal Of Solid-State Circuits*, vol. 24, pp. 1576-1586, Dec. 1989.

[14] M. Forti, L. pancioni, S. Rocchi, V. Vignoli, "Accurate CMOS implementation of PWL CNN neuron activations,"*IEEE International Symposium on Circuits & Systems*, vol. 1, pp. I-221- I-224, May 2002.

[15] J. Ramirez-Angulo, E. Sanchez-Sinencio, A. Rodriguez-Vazquez, "A Piecewise-linear function approximation using current mode circuits," *IEEE International Symposium on Circuits & Systems,* vol. 4, pp. 2025-2028, May 1992.

[16] Thorsten Kettner, Christian Heite, "Analog CMOS Realization of Fuzzy Logic Membership Functions," *IEEE Journal Of Solid-State Circuits,* vol. 28, no. 7, pp. 857-861, July 1993.

[17] T. Serrano-Gotarredona, B. Lineares-Barranco, "Current-mode fully-programmable piece-wise-linear block for neuro-fuzzy applications", *Electronics Letters*, vol. 38, no. 20, pp. 1165-1166, 2002.

[18] Marco Balsi, "Current-mode programmable piecewise-linear neural synapses", *Inernational Journal of Circuits Theory and Applications,*Vol. 31, pp. 265-275, 2003.

[19] K. Wawryn, B. Strzeszewski, "Low power VLSI neuron cells for artificial neural networks" *IEEE International Symposium on Circuits and Systems, vol. 3*, pp. 372-375, May 1996.

[20] K. Wawryn, A. Mazurek, "Current Mode circuits for Programmable Neural Networks" *IEEE International Symposium on Circuits & Systems*, pp. 698-681, May 2000.

[21] A. Rodriguez-Vazquez, M. Delgado-Restituto, "CMOS design of chaotic oscillators using state variables: a monolithic Chua's circuit," *IEEE International Symposium on Circuits & Systems*, vol. 40, no. 10, pp. 596-613, Oct. 1993.

[22] N. Manaresi, R. Rovatti, E. Franchi, G. Baccarani, " A current-mode Piecewise-linear function approximation circuit based on Fuzzy-logic," *IEEE International Symposium on Circuits & Systems*, vol. 3, pp. 127-130, June 1998.

[23] N. Manaresi, R. Rovatti, E. Franchi, G. Baccarani,, "A Micropower Current-Mode CMOS Function Approximation Circuit," *International Symposium on VLSI Technology, Systems and Applications*, pp. 290-293, June 1999.

[24] R. C. Frye, E. A. Rietman, C. C. Wong, "Back-Propagation Learning and Nonidealities in Analog Neural Network Hardware", *IEEE Trans. On Neural Networks*, Vol. 2, No. 1, pp. 110-117, January 1991.

[25] J. Choi, S. H. Bang, B. J. Sheu, "A Programmable Analog VLSI Neural Network processor for Communication Receivers" *IEEE Trans. On Neural Networks*, vol. 4, no. 3, pp. 484-495, May 1993.

[26] D. Lim, G. S. Moschytz, "A Programmable, Modular CNN Cell" *IEEE Conference Cellular Neural Networks and their Applications*, pp. 79-84, December 1994.

[27] H. Djahanshahi, M. Ahmadi, G. A. Jullien "Design and VLSI Implementation of a Unified Synapse-Neuron Architecture", 1996, *Proceedings Sixth Great Lakes Symposium on VLSI*, pp. 228 – 233, 22-23 March 1996.

[28] D. Coue, G. Wilson, "A Four-Quadrant Subthreshold Mode Multiplier for Analog Neural-Network Applications" *IEEE Trans. on Neural Networks*, Vol. 7, Issue 5, pp. 1212 – 1219, Sept. 1996.

[29] K.M. Al-Ruwaihi "CMOS analogue neuron circuit with programmable activation functions utilizing MOS transistors with optimized process/device parameters" *IEE Proc.-Circuits Devices and Systems*, vol. 144, no. 6, pp. 318-322, Dec. 1997.

[30] K. T. Lau, "A Low Power Synapse/Neuron cell for artificial neural networks", *Microelectronics Journal* 30, pp. 1261-1264, 1999.

[31] T. Shima, T. Kimura, Y. Kamatani, "Neuron Chips with On-Chip Back-Probagation and/or Hebbian Learning" *IEEE Journal of Solid-State Circuits*, vol. 27, no. 12, pp. 1868-1875, Dec 1992.

[32] G. Bogason "Generation of a neuron transfer function and its derivatives" *Electronics Letters*, vol. 29, no. 21, pp. 1867-1869, 14th Oct. 1993.

[33] J. Ngolediage "Hardware Implementation of Nonlinear Network Cells", *International Conference on System Engineering,* Coventry, U.K. , pp. 158-165, Sep 1991.

[34] J. E. Ngolediage, R. N. G. Naguib, S. S. Dlay, "A New Class of Analog CMOS Neural Network Circuits", *IEE Colloquium on Hardware Implementation of Neural Networks and Fuzzy Logic*, pp. 7/1 - 7/3, 9 Mar 1994.

[35] A. J. Annema "Hardware realization of a neuron transfer function and its derivative" *Electronics Letters*, vol. 30, no. 7, 31st March 1994.

[36] C. Lu, B. Shi, L. Chen, "A programmable on-chip BP learning neural network with enhanced neuron characteristics", *International Symposium on Circuits & Systems*, pp. 573 - 576 vol. 2, 2001.

[37] Chun Lu, Bingxue Shi, Lu Chen, "An expandable on-chip BP learning neural network chip" *International Journal of Electronics*, Vol. 90, No. 5, pp. 331-340, 2003.

[38] T. Delbu¨ck, " 'Bump' circuits for computing similarity and dissimilarity of analog voltage," *International Joint Conference on Neural Network Society*, Seattle, WA, 1991.

[39] S. S. Watkins and P. M. Chau, "A radial basis function neurocomputer implemented with analog VLSI circuits," in *International Joint Conference on Neural Networks*, vol. 2, pp. 607–612, 1992.

[40] E. Gatt, J. Micallef, E. Chilton, "Hardware Radial Basis Function Neural Networks for Phoneme Recognition", *ICECS 2001. The 8th IEEE International Conference on Electronic Circuits and Systems*, vol. 2, pp. 627 – 630, Sept. 2001.

[41] R. Kuppuswamy, L. Theogarajan, L. A. Akers, "A programmable Gaussian node", *IEEE International Conference on Neural Networks*, Vol. 2, pp. 880 – 885, June 1996.

[42] S. Churcher, A. F. Murray, and H. M. Reekie, "Programmable analogue VLSI for radial basis function networks," *Electronics Letters*, vol. 29, no. 18, pp. 1603–1605, Sept. 1994.

[43] J. Choi, B. J. Sheu, and J. C.-F. Chang, "A Gaussian synapse circuit for analog neural networks," *IEEE Transactions on VLSI Systems*, vol. 2, pp. 129–133, Mar. 1994.

[44] S. T. Lee, K. T. Lau, "An Analog Gaussian Synapse For Artificial Neural Networks," *Proceedings of the 38th Midwest Symposium on Circuits and Systems,* vol. 1. pp. 77 – 80, Aug. 1995.

[45] R.-J. Huang and T.-D. Chiueh, "Circuit implementation of the multivalued exponential recurrent associative memory," in *World Congress Neural Networks*, San Diego, CA, 1994, pp. 618–623.

[46] Shang-Yi Lin, Ren-Jiun Huang, Tzi-Dar Chiueh "A Tunable Gaussian/Square Function Computation Circuit for Analog Neural Networks" *IEEE Transactions on Circuits and Systems-II*, vol. 45, no. 3, pp. 441-446, Mar. 1998 .

[47] D. S. Masmoudi, A. T. Dieng, M. Masmoudi, "A Subthreshold Mode Programmable Implementation of the Gaussian Function for RBF Neural Networks Applications", *Proceedings of the 2002 IEEE Internatinal Symposium on Intelligent Control*, pp. 454-459, Oct. 2002 .

[48] Munir Samet, "A Beta basis function neural network in CMOS subthreshold mode", *Internatinal Journal of Electronics*, vol. 88, No. 6, pp. 645-657, 2001 .

[49] Mutlu Avci, Tulay Yildirim, "Generation of Tangent Hyperbolic Sigmoid function for microcontroller based digital implementations of Neural Networks," *Turkish Symposium on Artificial Intelligence and Neural Networks*.

[50] K. Basterretxea, J. M. Terela, "Approximation of sigmoid function and the derivative for hardware implementation of artificial neurons," *IEE Proceedings - Circuits Devices and Systems*, vol. 151, no. 1, pp. 18- , February 2004.

[51] P. R. Gray, P. J. Hurst, S. H. Lewis, R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. , John wiley & Sons, New York 2001.

[52] Behzad Razavi, *Design of Analog CMOS Integrated Circuits*, McGraw-Hill, New York, 2001.

[53] R. Jacob Baker, Harry W. Li, David E. Boyce, *CMOS Circuit Design, Layout, and Simulation*, John Wiley, New York, 1998.

[54] Phillip E. Allen, Douglas R. Holberg, *CMOS Analog Circuit Design*, Oxford University Press, New York, 2002.

[55] Bendong Sun, Fei Yuan, "New Low-Voltage Fully-Balanced Wide-Band Differential CMOS Current amplifier," *The 45th Midwest Symposium on Circuits and Systems*, Vol. 2, pp. 57-60, 2002.

VITA

- My name is Abdullah Bakri Shwehneh.
- I was born in Aleppo, Syria, in 1973.
- I got my first degree in 1998 from Sumy State University (Sumy-Ukraine) having specialized in Industrial Electronics. I was qualified as an Electronic Techniques Engineer.
- In 1999 I joined Aleppo University (Aleppo-Syria) for pursuing a post-graduate diploma degree in Automatic Control.
- In 2001, I started Working in "Electronic Brain Company for Computer and Electronics" as an Electronic & Computer Technique Engineer.
- Since 2002, I started working in KFUPM as a Research Assistant and pursuing the Master Degree in Electronics (EE department).
- My main interests are: VLSI analog design and Neural Networks hardware implementation.