Neurocomputing 118 (2013) 329–333

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Letters

Area-efficient differential Gaussian circuit for dedicated hardware implementations of Gaussian function based machine learning algorithms

D. Vrtaric [a], V. Ceperic [a,b,*], A. Baric [a]

[a] University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia
[b] KU Leuven, Department of Electrotechnical Engineering, Kasteelpark Arenberg 10, Leuven, Belgium

ARTICLE INFO

Article history:
Received 12 September 2012
Received in revised form
23 February 2013
Accepted 24 February 2013
Communicated by R.W. Newcomb
Available online 14 March 2013

Keywords:
Gaussian function
Gaussian circuit
Machine learning hardware

ABSTRACT

A simple and area-efficient differential Gaussian circuit is presented for machine learning dedicated hardware implementations, where Gaussian functions are needed, e.g. for artificial neural networks (as transfer function), support vector machines (as kernel function) and fuzzy logic (as membership function). The proposed Gaussian circuit consists of only 4 transistors. Simulations in the 0.18-μm CMOS UMC technology show that the proposed circuit is more accurate, less susceptible to the process variations and requires less on-chip area when compared to state-of-the-art.

1. Introduction

Many artificial intelligence (AI) and machine learning methods, including modern classification, regression and reasoning algorithms, use the Gaussian function, e.g. artificial neural networks (as transfer function), support vector machines (as kernel function [1]) and fuzzy logic (as membership function) [2]. Even though these algorithms are commonly implemented in software, a dedicated hardware implementation offers several advantages such as real-time execution [3], implementation of very large systems consisting of millions of nodes [4,5], massively parallel execution [6], exceptionally fast response, etc. An overview of the advantages of the specialised hardware implementations can be found in [7]. In this paper, the design of a new Gaussian circuit suitable for implementation in very-large-scale integrated circuits is investigated.

The state-of-the-art implementation of the Gaussian circuit (named Gilbert Gaussian circuit) was recently proposed in [8] (Fig. 1(a)). They have shown that for AI and machine learning applications it is not needed to have the programmable sharpness γ of the Gaussian function. Instead, they control the sharpness by scaling the circuit input voltages. This enables the use of simpler, more accurate and robust Gaussian function circuits. This simplification is also used in this paper, i.e. the sharpness is presumed to be controlled by scaling the differential inputs to the Gaussian circuit.

Some Gaussian circuit implementations are not differential, e.g. [9–12]. Differential solutions are usually more applicable in low-power, low-voltage systems or in a very-large-scale integrated circuits because they are less susceptible to noise and external disturbances and therefore, in this paper we will focus on the differential version of the Gaussian circuit. The discussion about the importance of reduction of the influence of noise when evaluating Gaussian functions can be found in [13].

There are several important factors that must be considered when designing a Gaussian circuit:

- The circuit must be area-efficient because machine learning algorithms, such as e.g. support vector machines can consist of millions of kernel (Gaussian) functions. The decrease in the Gaussian circuit cell size can significantly reduce the overall chip area and therefore, reduce the overall cost and increase yield.
- The circuit must be accurate. Although some inaccuracies can be compensated by the training algorithm [8], if the circuit is accurate, the training procedure can be simplified and the overall accuracy is improved [14,15]. The importance of accurate evaluation of the Gaussian function is analysed in [13].
- The influence of process variations must be minimal because the training algorithms rely on the consistent shape of the Gaussian function within the chip.

2. Circuit implementation

The proposed circuit, shown in Fig. 1(b), consists of 4 transistors. The drain of the transistor M_1 is connected to the source of the

* Corresponding author at: University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia. Tel.: +385 91 1553885.
E-mail address: vladimir.ceperic@ieee.org (V. Ceperic).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.02.022

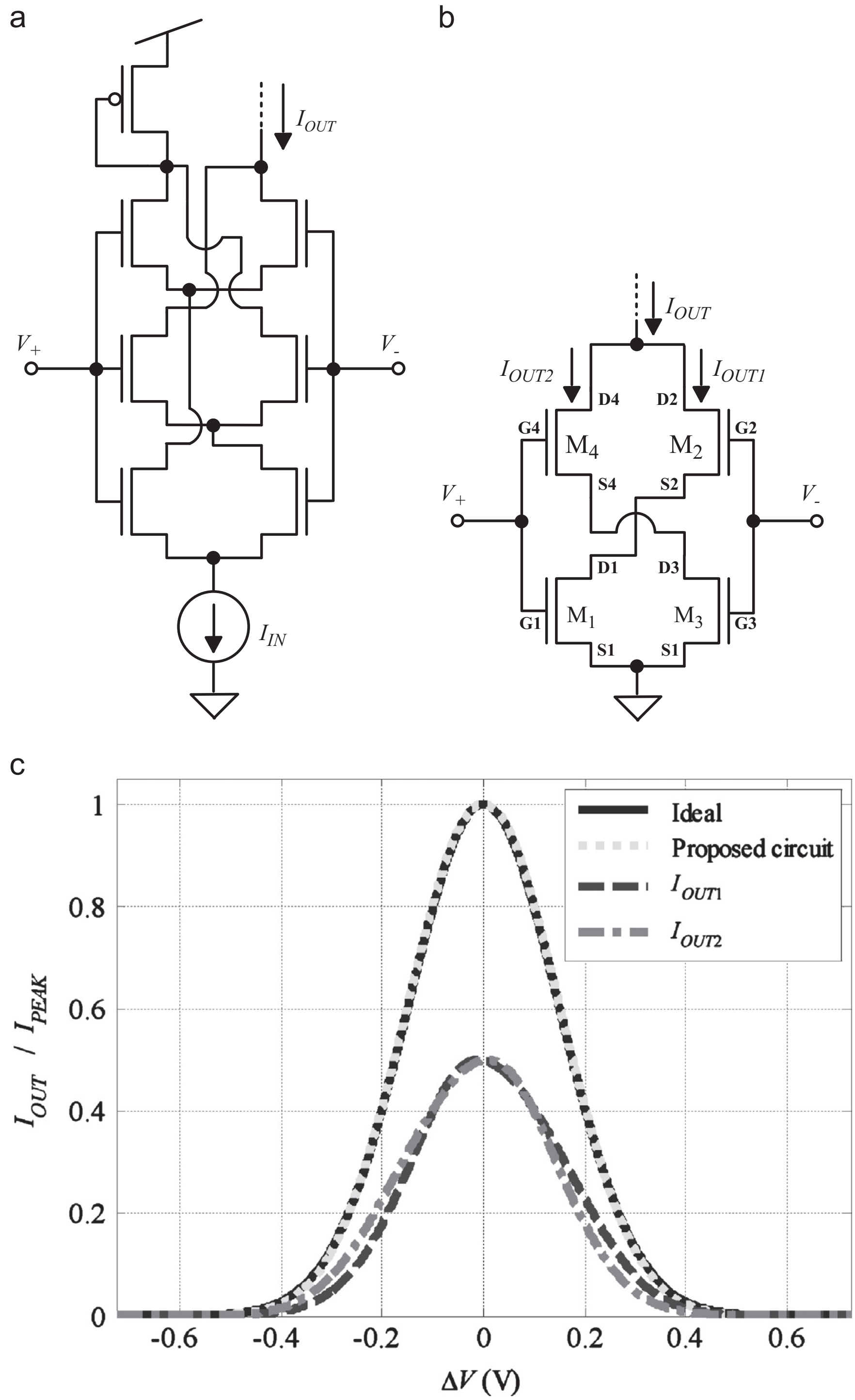

Fig. 1. (a) Gilbert Gaussian circuit [8]. (b) Proposed area-efficient differential Gaussian circuit. (c) Proposed circuit output characteristics and comparison with the ideal Gaussian function. Current source I_{IN} is usually implemented using a current mirror. A practical way to distribute I_{OUT} is to connect the output of the circuit (marked as the dashed line) to a current mirror.

transistor M_2, while the drain of the transistor M_3 is connected to the source of the transistor M_4.

There are four distinct situations that are related to the shape of the Gaussian function:

- When V_- is low (e.g. close to 0 V) and when V_+ is high (e.g. close to power supply voltage V_{DD}): the transistors M2 and M3 are in high impedance state and the current I_{OUT} is approximately equal to zero.
- When V_- is high (e.g. close to power supply voltage V_{DD}) and when V_+ is low (e.g. close to 0 V): the transistors M1 and M4 are in high impedance state and the current I_{OUT} is approximately equal to zero.
- When V_- and V_+ are equal (e.g. close to $V_{DD}/2$) the current I_{OUT} has the highest value. The amplitude of I_{OUT} can be varied e.g. by changing the width and length of the transistors. The circuit is designed to be used with a multiplier, as proposed in [8], to scale the resulting Gaussian function as needed and therefore, the most critical parameter is the accuracy of the