A Scalable Low Voltage Analog Gaussian Radial Basis Circuit

Luke Theogarajan and L. A. Akers

***Abstract*—Gaussian basis function (GBF) networks are powerful systems for learning and approximating complex input-output mappings. Networks composed of these localized receptive field units trained with efficient learning algorithms have been simulated solving a variety of interesting problems. For real-time and portable applications however, direct hardware implementation is needed. We describe experimental results from the most compact, low voltage analog Gaussian basis circuit yet reported. We also extend our circuit to handle large fan-in with minimal additional hardware. Our design is hierarchical and the number of transistors scales almost linearly with the input dimension making it amenable to VLSI implementation.**

***Index Terms*—Analog VLSI, Gaussian basis function, neural networks.**

I. Introduction

Neurons with response characteristics that are locally tuned to a particular range of the input variable have been found in many parts of the central nervous system [1]. Examples include cells in the somatosensory cortex that respond selectively to stimulation from localized regions of the body surface, and orientation selective cells in the visual cortex that respond selectively to stimulation which are both local in retinal position and local in angle of object orientation [2]. Populations of these locally-tuned cells have been found organized in cortical maps where the input variable varies in an approximate linear fashion with position in the map [1]. These maps are also observed to have overlapping receptive fields. Overlapping receptive fields offer the capability for improving signal to noise ratios and for providing fault tolerance.

Computer simulations of GBF networks are sufficient for many applications. However, hardware implementation of these systems are mandatory for many real-time, or low power, portable applications such as vision and speech recognition, robotics, and numerous other interactive control and signal processing applications. Hierarchical networks of GBF's are used as elementary feature detectors, then these features are combined in multiple ways to build complex feature detectors. In the past few years there have been a number hardware implementations of the Gaussian basis function in analog [3]–[6], and pulse forms [7]. The speed advantage of the hardware approach will continue to expand as the systems are scaled to the number of processing elements found in biological systems. For these reasons, we have developed a compact, low voltage, analog circuit implementation of the GBF network.

II. The Multidimensional Gaussian Basis Circuit

The Gaussian function implementation in most of the circuits referenced above are direct mathematical implementations. However, most neural models are often simplified for analytical tractability, and are not intended to be an accurate representation of its biological counterpart. Therefore we believe that rather than designing a circuit to give an exact Gaussian, a circuit that has the essential properties of the Gaussian will suffice. This is a peak at the desired center and a nonlinear continuous drop on either side as the input moves away from the center. Therefore, we designed a circuit which has a general Gaussian or "Bump" shape.

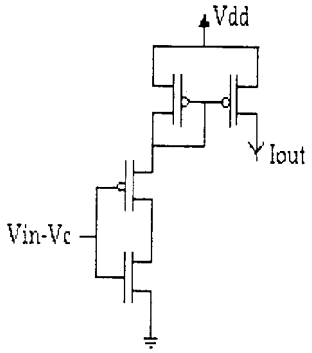

Fig. 1. Gaussian basis circuit.

When two transistors are connected in series, there occurs a self correlation of currents, and if the currents have a differential or complementary nature a bump output results [3]. One way of implementing this differential or complementary nature is to use a differential amplifier. An alternate method is to use a device which has a complementary characteristic to the same input voltage. PMOS and NMOS devices have such complementary characteristics. By using this inherently complementary nature we have been able to design a circuit which approximates a Gaussian surface. The circuit is shown in Fig. 1. The input to the circuit is $V_{in} - V_c$, where V_c is the center. We can to store this center using the floating gate transistor synapse as presented in [8]. In order to use the exponential relationship between the input voltage and output current, we operate our circuit mostly in the subthreshold region of operation. One way to achieve this is to lower the supply voltage such that both devices operate in the subthreshold region. However, this results in a very small current for all the input voltage swing. We have developed a method for the circuit to be in the subthreshold voltage region of operation for the tails of the output current, and be in the saturated above threshold voltage region for the peak of the output current. This gives an excellent peak-to-valley ratio, and good current drive. We do this with a nonlinear resistor, in our case by a drain connected PMOS transistor. The PMOS load also facilitates the mirroring of the current to the output transistor. The load transistor drops the voltage to the source of the correlating PMOS transistor as a function of the current through the circuit. Thus the voltage seen by the source of the PMOS transistor when it dominates is lower than the supply and after a few tenths of volts forces the circuit into the subthreshold region of operation. The advantage in using this method is that the dynamic bias variation to the source of the PMOS transistor allows the circuit to operate to a large extent in the subthreshold regime but helps to keep the supply voltage relatively high. The point where the built-in center occurs is not at $V_{dd}/2$ as in the case of a the current through a simple inverter, but at a smaller voltage slightly above threshold. This above threshold operation at the peak is beneficial, since the peak of a Gaussian resembles a quadratic and increases the current drive.

The supply voltage has an effect on the symmetry of the Gaussian shape of the output. This occurs due to the antisymmetrical loading on the PMOS side which leads to an extended range of operation for the PMOS transistor as the supply is increased. In order for the circuit

Manuscript received October 3, 1995; revised April 2, 1996. This paper was recommended by Associate Editor C. G. Lau.

The authors are with the Center for Solid State Electronics Research, Arizona State University, Tempe, AZ 85287 USA.

Publisher Item Identifier S 1057-7130(97)07652-0.

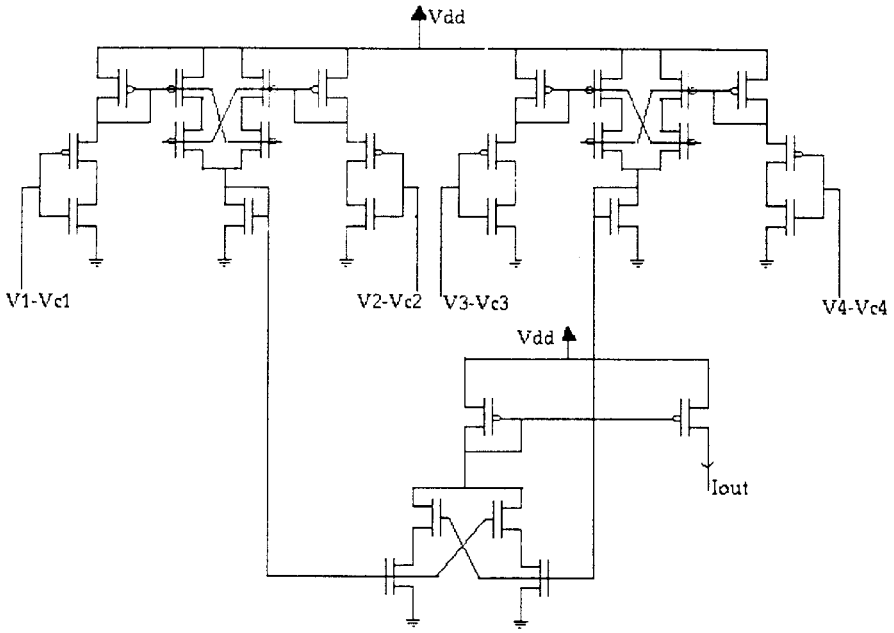

Fig. 2. The multidimensional Gaussian circuit. The circuit is shown for 4 inputs. The same principle can be used to extend this to n inputs.

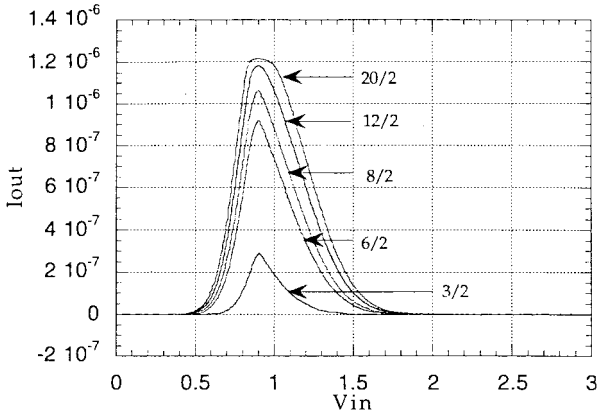

Fig. 3. The measured output curves of the circuit for equal PMOS and NMOS transistor sizing of 20/2, 3/2 for the largest to the smallest current peaks. The input to the circuit was with V_c equal to zero.

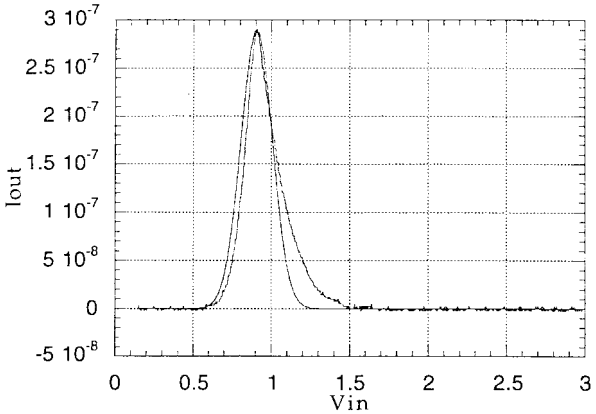

Fig. 4. Comparison of the measured circuit output(curve with longer tail) to a best fitting Gaussian.

to operate approximately symmetric and for low power dissipation, the supply is fixed at 3 V. Decreasing the supply below 3 V will force the entire circuit in subthreshold operation just lowering the current levels and circuit speed. However, operating the circuit above 3 V will increase the anti-symmetry. These effects are shown later with experimental data.

An advantage of the design is the circuit operates both in subthreshold and above threshold. Thus the physics of the device gives us an exponential and at the peak we obtain moderately high current levels and noise immunity. Most circuits are operated in the subthreshold region by subthreshold current sourcing. Our circuit is not limited by an external current source and has a dynamic subthreshold region of operation.

The extension of our circuit to the multidimensional case relies on the property of multiplication of Gaussians gives a Gaussian. By multidimensional we mean having multiinputs with one overall output. This circuit is shown in Fig. 2. When two transistors are connected in series the output current is a correlation of the inputs. Correlation can also be thought of as a normalized multiplication (for two inputs). If more than two transistors are connected in series then the output current is given by

$$\frac{1}{I_{\text{out}}} = \sum_{k=1}^{n} \frac{1}{I_k} \tag{1}$$

where n is the number of inputs.

III. Experimental Results

The chips were fabricated in the MOSIS 2-μm Nwell process. One chip had many one input Gaussian circuits with various transistor sizing, and another chip had 12 input circuits. The output was recorded using the Lab View data acquisition software. Since we did not provide any weighted current mirrors in the chip to give a high current output and Lab View requires a voltage input, we used a npn transistor configured in the emitter follower mode to give the required output voltage to be measured. The input to chip was a sawtooth from 0–3 V. We made all measurements with V_c equal to 0 V. The results from our chip for the one dimensional case for various sizings is shown in Fig. 3.

For constant W/L ratios of the PMOS to the NMOS, the built-in center occurs at the same point. While the width of the Gaussian is

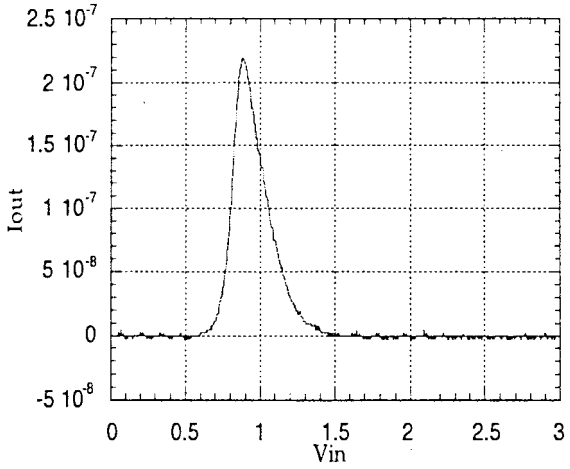

Fig. 5. The measured output of the 12 input multidimensional Gaussian circuit with all the inputs tied together.

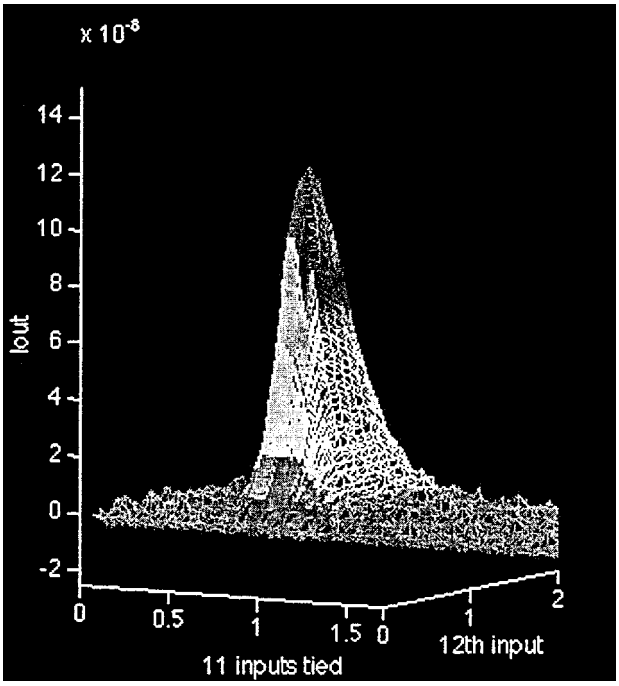

Fig. 6. The measured output of the multidimensional Gaussian circuit. 11 inputs are tied together. The 11 inputs are fed a sawtooth and the twelfth input is manually stepped using a resistor string.

set by the width to length ratio fixed during layout, the peak of the current can be varied with the supply voltage.

A comparison of the measured circuit output to a Gaussian is shown in Fig. 4. The circuit is the curve with the longer tail on the negative exponential. This is due to anti-symmetrical loading on the PMOS side and can be adjusted by varying the device sizing. The difference in area from an ideal Gaussian and our circuit implementation is 11%.

The 12 input multidimensional Gaussian circuit was tested as follows. First all the twelve inputs were tied together and the output is shown in Fig. 5. Next, eleven inputs were tied together and the twelfth input was manually stepped through a voltage range from 0–2 V in steps of 0.1 V using a resistor string. These results are shown in Fig. 6.

IV. Conclusion

Gaussian basis functions are universal approximators. We have built, fabricated and tested a very compact electronic implementation of such a function. We have also implemented a multidimensional extension of our circuit. This type of multidimensional cell may be used to identify complex features, like lines and bars, as opposed to simple ones, like points, for the one dimensional case.

References

[1] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, pp. 281–294, 1989.
[2] ____, "Learning with localized receptive fields," in *Proc. Connectionist Models Summer School*, D. Touretzky, G. Hinton, and T. Sejnowski, Eds. San Mateo, CA: Morgan-Kaufmann, 1988.
[3] T. Delbruck, "Bump circuits," *Caltech Int. Docu.*, CNS Memo 26, 1993.
[4] J. Choi, B. J. Sheu, and J. C.-F. Chang, "A Gaussian synapse circuit for analog VLSI neural networks," *IEEE Trans. VLSI Syst.*, vol. 2, pp. 129–133, Mar. 1994.
[5] J. Anderson, J. C. Platt, and D. B. Kirk, "An analog VLSI chip for radial basis functions," in S. J. Hanson, J. D. Cowan, and C. L. Giles, Eds., *Advances in Neural Information Processing Systems*. San Mateo, CA; Morgan Kaufmann, 1993, vol. 5, pp. 765–772.
[6] S. S. Watkins and P. M. Chau, "A radial basis function neurocomputer implemented with analog VLSI circuits," in *Proc. IEEE/INNS Int. Joint Conf. Neural Net.*, vol. II, Baltimore, MD, 1992, pp. 607–612.
[7] S. Churcher, A. F. Murray, and H. M. Reekie, "Programmable analogue VLSI for radial basis function networks," *Electron. Lett.*, vol. 29, no. 18. pp. 1603–1605, Sept. 1993.
[8] P. Hasler, C. Diorio, B. Minch, and C. Mead, "Single transistor learning synapse with long term storage," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1995, vol. 3, Seattle, WA.

Improved Model Reduction Procedure for 2-D Separable Denominator Digital Systems

Chengshan Xiao

***Abstract*— An improved procedure is presented for the model reduction of 2-D separable denominator digital systems $H(z_1, z_2)$. The proposed procedure is based on modifying the factorization $H(z_1, z_2) = H_2(z_2)H_1(z_1)$ and minimizing an existing frequency error bound developed by Zhou, Li, and Lee in 1994. In terms of small actual frequency errors and tight frequency error bounds, this procedure is superior to the existing model reduction methods.**

***Index Terms*—Model reduction, two-dimensional systems.**

I. Introduction

The *balanced* model reduction has now become a powerful technique for the approximation of a system. This model reduction method was initially studied by Moore [1] and Kung [2] for

Manuscript received December 19, 1995; revised June 12, 1996. This work was supported by the NSFC, and Australian OPRS and UPRA of the University of Sydney. This paper was recommended by Associate Editor S. Goto.

The author was with the Department of Electrical Engineering, University of Sydney, Sydney, NSW 2006 Australia. He is now with Nortel, Ottawa, ONt. K1Y 4H7, Canada (e-mail: cxiao@nortel.ca).

Publisher Item Identifier S 1057-7130(97)07651-9.