# A Programmable Gaussian Node

Ravishankar Kuppuswamy, Luke Theogarajan and L.A. Akers
Center for Solid State Electronics Research
Arizona State University, Tempe, AZ 85287-5706
email: lakers@asu.edu

*ABSTRACT*

Analog Gaussian basis circuit integrated with a non-volatile storage memory cell is described. Hardware implementations of the Gaussian basis circuit with on-chip learning is needed for real time and portable applications. Each Gaussian basis cell is symbiotically inter- linked with its own long-term storage memory cell forming a highly localized architecture. Experimental results of both the Gaussian basis circuit and the memory element is presented. We show simulated results of an application of our cell in a MERAM system.

## I. Introduction

The human brain is comprised of cells with response characteristics that are locally tuned to a particular range of input variables. These cells access information required from memory bodies located in the vicinity. The Gaussian basis function network is an implementation of a powerful system for learning and approximating complex input-output mappings. However, most of these networks are computer simulations. A hardware implementation is needed for high speed and portable systems. We have designed a Gaussian cell[9] which is extremely compact and is integrated with a non-volatile memory element in a hierarchically connected architecture. Each Gaussian cell has a local memory element thereby emulating the brain structure to a large extent. The integrated architecture is highly expandable and scales up to a large number of processing elements. We present both simulations and experimental results of these circuits.

## II. Gaussian Basis Circuit

The Gaussian function implementations in most of the circuits reported[1, 2, 4, 5, 10] are direct mathematical implementations. 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 smooth 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.

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[5]. 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 Figure 1. The input to the circuit is $V_{in}$-$V_c$, where $V_c$ is the center.

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

0-7803-3210-5/96 $4.00©1996 IEEE

in a very small current for the entire 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 non-linear 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 the above 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 Vdd/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 output of the circuit is shown in Figure 4. We have also extended this circuit to handle many inputs using a novel current correlator multiplication method[9].

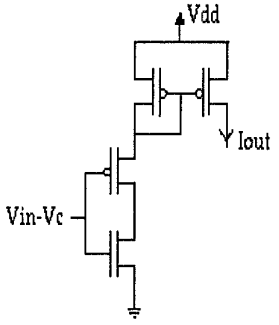


Figure 1. Gaussian basis circuit

## III. Analog Memories

Analog memories for the storage of synaptic weights are generally classified by the duration of charge storage. Short-term storage cells are usually modeled as capacitors[11] at the gate of the transistor and the weight is transferred to the gate using a pass transistor operating as a switch. The leakage current through the junction of this pass transistor limits the storage time. Medium-term storage cells are primarily short-term storage cells with additional refresh circuitry to rejuvenate the charge at periodic intervals of time. Long-term storage elements[6, 7, 11] are usually electrically programmable non-volatile memories where the charge is stored on a floating electrode. The floating electrode is insulated by $SiO_2$, thereby making charge leakage from the gate highly negligible. Our Gaussian basis cell targets applications such as feature recognition and detection. These applications require the weights to be stored for long periods of time and not change on an continuous basis. Our choice of memory naturally targeted the non-volatile storage option. As the cells use hot electron injection and Fowler-Nordheim Tunneling mechanisms for their write and erase procedures, they usually have to be fabricated using special processing techniques. We have hence chosen a storage cell[6, 7] motivated by the Caltech group and fabricated using the 2μ standard MOSIS process to be used for weight storage in the Gaussian basis circuit.

## IV. The Memory Cell

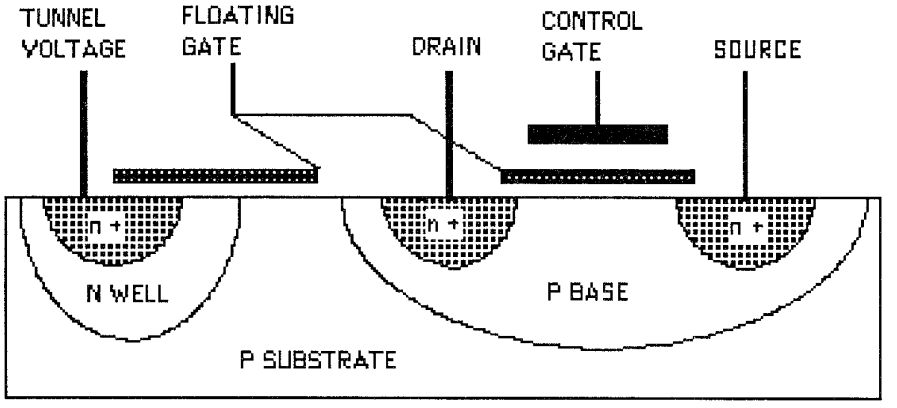


Figure 2. Crossection of non-volatile memory cell. The cell contains p-base layer encompassing the source and drain regions. The four terminals of the cell are source, drain, control gate and tunneling gate

The memory cell[7,8] is basically made of a single NMOS field effect transistor. This transistor is innovatively modified to make it a non-volatile memory storage element fabricated using a standard process. The cross-section of the memory cell is shown in Figure 2. The four basic terminals are the source, drain, control gate and tunneling gate. The source and drain regions are n+ regions lying in a p-substrate, totally enclosed by the moderately doped p-base region. The p-base region plays the pivotal role in the operation of the cell. The p-base layer fulfills two basic functions. Firstly it increases the threshold voltage, thereby creating a region where the potential drops greater than 3.1V within 0.2μm to provide the hot electrons enough energy to cross the oxide barrier. Secondly it also raises the electric field intensity of the channel thereby increasing the rate of hot electron injection. Both these properties act in unison to provide a range of operating drain voltages in the subthreshold region. This cell has a double poly structure with poly1 acting as the floating gate and poly2 as the control gate. The charge is stored on the poly1 floating gate, which is insulated from the rest of the cell by high quality $SiO_2$ making the cell a non-volatile storage element.

The charge on the floating gate is modified using both the tunneling and control gates. The control gate terminal is formed directly on the poly2 layer using a capacitance contact. The poly1 layer runs over a well region and the tunneling gate terminal is formed using a well contact. The charge on the floating gate is modified using both tunneling and hot electron injection mechanisms. A high positive voltage is applied on the well region, attracting the electrons on the floating gate into the well tunneling through the oxide barrier. This tunneling process decreases the electrons on the gate, thereby increasing the voltage on the floating node. Now a large drain voltage combined with a positive voltage on the control gate attracts hot electrons on to the floating gate. This decreases the voltage on the floating electrode.

## V. The $V_{in}$ - $V_c$ Circuit

Most analog VLSI circuits for neural networks employ the differential pair as the heart of the circuit design. This differential pair allows the computation of V1-V2. This design philosophy has its advantages and disadvantages. Our design does not use a differential pair and hence we need a separate circuit to do this computation and is the subject of this section. We have used a long-term memory cell[7] in a novel fashion to store $V_c$. One method of achieving a difference in voltages is to put them in series. This method is however not a very feasible one since most voltages are referenced

to ground. In order to make our circuit compatible to all methods of storage we present a novel two transistor circuit to do the difference computation. The circuit is shown in Figure 3. The drain connected NMOS transistor acts as a resistive load. The voltage at the output node of this circuit is given by

$$\mathbf{V_o = V_{in} - I_d R_d}$$

where $I_d$ is the current due to a voltage $V_c$ on the gate of the lower transistor and $R_d$ is the resistance offered by the drain connected NMOS transistor. If the upper and lower transistor are identical then the value of $I_d R_d$ is equal to $V_c$. This achieves the desired computation. For the transistors to be identical the upper transistor should not have any body effect, therefore it should be placed in a separate well. The SPICE output of the Gaussian cell for different $V_c$ voltages using the $V_{in}$-$V_c$ circuit is shown in Figure 5.

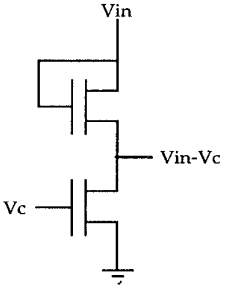


Figure 3. Vin-Vc circuit

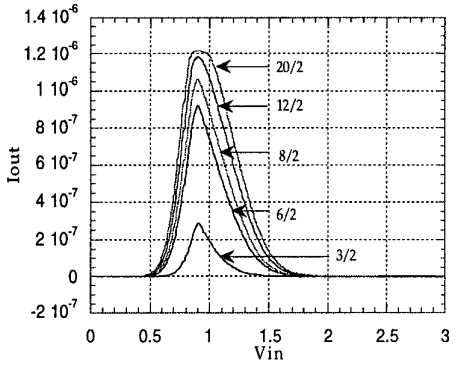


Figure 4. The measured output curves of the for equal PMOS and NMOS transistor sizing of 20/2, 3/2 for the largest to the smallest current peaks. The input to the circuit was with $V_c$ equal to zero.

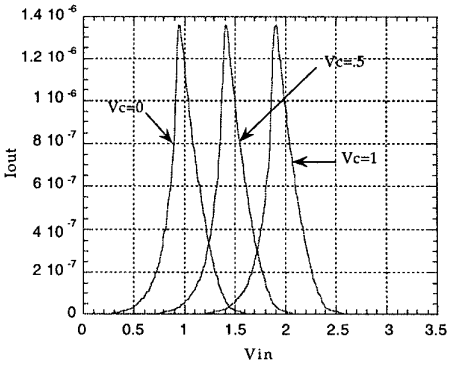


Figure 5. The SPICE simulated output of the circuit Gaussian circuit with the $V_{in}$ - $V_c$ circuit for different $V_c$ voltages

## VI. Memory Testing and Integrated Architecture

The charge on the floating gate represents the weight stored in the memory element. The voltage stored on the floating gate was directly measured by connecting the floating gate to a negative feed-back operational amplifier buffer circuit. A circuit[1] comprising of a NMOS and PMOS transistor with their drains connected to a feedback capacitance was also used. The floating gate was connected to the gate of a PMOS transistor and the capacitance for linear voltage storage control on the other plate of the capacitance. This node can be programmed to store any rail to rail voltage.

The tunneling operation was conducted by applying a high positive voltage on the well (32-35V). This voltage stripped the floating electrode of its electrons and increased the voltage on the gate to about 5-6V. To cease the tunneling operation, the voltage on the tunneling terminal was brought down to 25V and the floating gate voltage was maintained. Accordingly the voltage measured at the capacitance node reads zero after the tunneling operation. Hence the cell is now completely erased. The electrons are now provided with both the energy and direction to cross the oxide barrier. The control voltage is maintained at 8V and the drain voltage is brought up to about 4-5V. This attracts electrons to the floating gate and decreases the voltage on the gate. The capacitance node voltage now increases in a linear fashion. Experimental results indicating voltage changes at the output capacitance node due to both tunneling and injection processes are shown in Figure 6.

The memory cell was combined with the Gaussian basis circuit and different sizes of NxN architectures were designed. The basic idea of the integration process was to efficiently reduce the number of I/O lines. The total area occupied by each memory cell and its read/write circuitry is approximately 70 $\mu m^2$. The memory cell has four basic lines. All the cells of the architecture have common source and control terminals. The drain terminal of each cell was connected to its write circuitry. The tunneling line of all the cells in the architecture were presently independent of each other. On designing larger architectures, a separate addressing scheme would help in optimizing on the number of tunneling lines. The source terminal of all the cells in the architecture was connected to ground. The output weight $V_c$ stored in the memory cell was fed as an input to the $V_{in}$ - $V_c$ computational circuit.

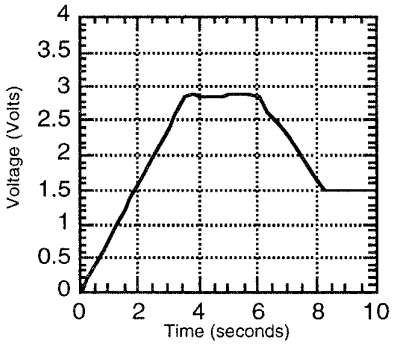


Figure 6. Output voltage of the capacitance during injection and tunneling operations

## VII. Application

The Gaussian basis circuit can be used in Gaussian radial basis classifiers. We show a SPICE simulation of a circuit implementing a multi-valued exponential recurrent associative memory (MERAM) [3]. This application employs a multi-dimensional Gaussian basis circuit as a similarity computing element. Our implementation is a variation of the implementation presented in [8]. A three input, three output, MERAM was simulated. Three patterns of three components were stored. For the first simulation the stored patterns were

$$\begin{pmatrix} 1 & 1 & 1 \\ 1.5 & 1.5 & 1.5 \\ 2 & 2 & 2 \end{pmatrix}$$

The pattern applied to the input of the network was $(1.5 \quad 1.7 \quad 1.2)$. The network settled into the stored pattern $(1.5 \quad 1.5 \quad 1.5)$ as shown in Figure 7.

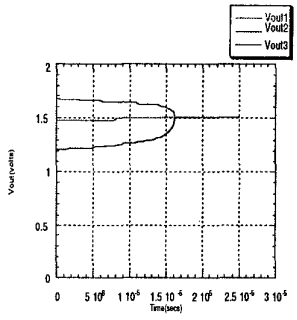


Figure 7. Output of MERAM network node as a function of time

## VIII. Conclusions

We have described a very compact analog Gaussian basis cell and memory element in a highly localized architecture. Both simulations and experimental results demonstrating the operation of these circuits have been presented.

## IX. Acknowledgments

We would like to thank Paul Hasler at California Institute of Technology for helpful discussions on the non-volatile memory design and testing procedures. This work was supported by the Office of Naval Research.

## X. References

[1]. 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,*Advances in Neural Information Processing Systems,*, Vol. 5, San Maetro, CA; Morgan Kaufmann Publishers Inc., 1993, pp. 765-772.

[2]. S. Churcher, A. F. Murray and H. M. Reekie, "Programmable Analogue VLSI for Radial Basis Function Networks", *Electronics Letters*, 2nd September 1993, Vol. 29, No. 18. pp. 1603-1605.

[3]. T. D. Chieuh and H. K. Tsai, "Multivalued Associative Memories Based on Recurrent Networks", *IEEE Trans. on Neural Networks*, vol4, no. 2, pp. 364-366, Mar 1993.

[4]. Joongho Choi, Bing J. Sheu, and Josephine C.-F. Chang, " A Gaussian Synapse Circuit for Analog VLSI Neural Networks," *IEEE Trans. on VLSI Systems*, Vol.2, March 1994, pp. 129-133.

[5]. T.Delbruck, "Bump Circuits," *Caltech Internal Document*, CNS Memo 26,1993.

[6]. Chris Diorio, Sunit Mahajan, Paul Hasler, Bradley A. Minch and Carver Mead, " A High-Resolution Non-Volatile Analog Memory Cell," *Caltech Internal Document.*

[7]. Paul Hasler, Chris Diorio, Bradley A. Minch and Carver Mead, "Single Transistor Learning Synapses," *Caltech Internal Document.*

[8]. R.J. Huang and T. D. Chieuh, "Circuit Implementation of the Multivalued Exponential Recurrent Associative Memory", *World Congress on Neural Networks*, vol II, 1994, pp. 618-623.

[9]. Luke Theogarajan and L.A. Akers, " A Multi-Dimensional Analog Gaussian Radial Basis Circuit," *Proc. of International Symposium of Circuits and Systems* , 1996.

[10]. S. S. Watkins and P. M. Chau, "A Radial Basis Function Neurocomputer Implemented with Analog VLSI Circuits", *Proc. IEEE/INNS Int. Joint Conf. Neural Net.*, vol. II, pp. 607-612, Baltimore, MD, 1992.

[11]. Vittoz etal, " Analog Storage Of Adjustable Weights," in *VLSI Design of Neural Networks.* Boston: Kluwer, 1991, pp. 344-363.