A Four-Quadrant CMOS Analog Multiplier for Analog Neural Networks

Navin Saxena and James J. Clark

Abstract— **A four-quadrant CMOS analog multiplier is presented. The multiplier uses the square-law characteristic of an MOS transistor in saturation. Its major advantage over other four-quadrant multipliers is its combination of small area and low power consumption. In addition, unlike almost all other designs of four-quadrant multipliers, this design has single ended inputs so that the inputs do not need to be pre-processed before being fed to the multiplier, thus saving additional area. These properties make the multiplier very suitable for use in the implementation of artificial neural networks. The design was fabricated through MOSIS using the standard 2 μm CMOS process. Experimental results obtained from it are presented.**

I. Introduction

Recently, analog multipliers have found use in the area of artificial neural networks [3]–[6], [14]. These massively parallel analog systems have demonstrated potential for solving a wide range of difficult problems, and thus analog computing techniques have become more widespread. Along with this, the use of analog multipliers has also increased. However, while high linearity is the prime issue for the multipliers in conventional applications like modulation, demodulation, frequency translation, etc., neural systems require their multipliers to be small and modular, and have low power consumption.

In the past, several designs of analog CMOS multipliers have been proposed [1], [2], [7]–[13], [15]–[19]. However, most of these multipliers were designed with the conventional applications in mind. Of these, the multipliers in [1], [8], [12], [13], [15], [16], [18] use resistors, which are area-intensive in VLSI layout, and hence are unsuitable for neural network applications. Some multipliers use switch-capacitor techniques [7], [19], but the area of the capacitors cause the multipliers to become large. The multipliers in [2], [10], [17] have differential inputs, requiring the values being fed to the multiplier inputs to be generated from the actual values that have to be multiplied by shifting and/or inverting them. This makes the design non-modular because the user has to be aware of the biasing conditions and has to pre-process the values to be multiplied before feeding them to the multiplier. Only the multipliers in [11], [9] were designed specifically for neural systems. However, the one in [11] is limited to two quadrant operation only, while the multiplier in [9] has differential inputs and output, which requires additional hardware and consequently more silicon area. In contrast, the four-quadrant multiplier presented in this paper has single ended inputs. Thus, it can be more easily integrated as a module in a large system. Also, its area is much smaller than all of the four-quadrant multipliers mentioned above. These properties, along with its low power consumption, make it very suitable for use in the implementation of large systems like neural networks.

In this paper, the principle of operation and design of the multiplier is first described. In this section, the design of the multiplier, as well as how to arrive at the range of operation of the multiplier, is described. In the next section, the experimental results are presented. A summary of the achieved results is presented in conclusion.

Manuscript received October 13, 1993; revised February 22, 1994. This work was supported by the Joint Services Electronics program grant no. N00014-89-J-1023 and the NSF Center of Excellence in Systems Research, University of Maryland/Harvard University System Research Center grant no. CDR-85-00108.

The authors are with the Division of Applied Sciences, Harvard University, Cambridge, MA 02138.

IEEE Log Number 9402124.

II. Principle of Operation and Design

The basic principle of operation of the proposed multiplier is based on the well-known identity:

$$(V_1 + V_2)^2 - V_1^2 - V_2^2 = 2V_1V_2.$$

In our multiplier, the squaring is achieved using the square relationship between the gate-to-source voltage and the drain current on an MOS transistor in saturation. The following subsections describe the circuit design and the range of operation of the multiplier.

A. Design of the Four-Quadrant Multiplier

The circuit diagram of the multiplier is shown in Fig. 1. It consists of two current mirrors and five n-MOSFET's, M_{1-5}, which have the same aspect ratios. We assume that the five transistors are in the saturation region of operation. Also, their sources and substrates are tied together so the body-effect is not important. According to the square-law characteristic of an MOS transistor in saturation, the drain currents I_1 and I_2, flowing through transistors M_1 and M_2, respectively, are given by

$$I_1 = K(V_{\text{in1}} - V_b - V_T)^2 \quad (1)$$

$$I_2 = K(V_{\text{in2}} - V_b - V_T)^2 \quad (2)$$

where V_{in1} and V_{in2} are the inputs to the multiplier, V_b is a bias voltage (see Fig. 1), V_T is the threshold voltage of the transistors, and K is the transconductance parameter, which is the same for the transistors because their aspect ratios are same. Because of current mirror C_1, the current flowing through transistor M_5 is the same as that through transistor M_1, and since all other parameters are the same for the two transistors, their gate to source voltages must be equal. Hence,

$$V_{D5} = V_{\text{in1}} + V_{\text{in2}} - V_b = V_{G3} \quad (3)$$

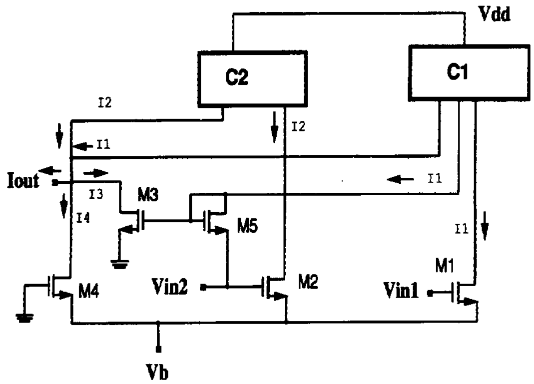

Fig. 1. Circuit diagram of four-quadrant multiplier.

where, V_{D5} is the drain voltage of M_5 and V_{G3} is the gate voltage of M_3. The currents I_3 (using (3)) and I_4, flowing through transistors M_3 and M_4 are given by:

$$I_3 = K(V_{\text{in}1} + V_{\text{in}2} - V_b - V_T)^2 \quad (4)$$

$$I_4 = K(-V_b - V_T)^2 \quad (5)$$

where we assume that $V_b < -V_T$. Current mirrors C_1 and C_2 feed currents I_1 and I_2 respectively onto the output node. Thus, the output current I_{out} is given by:

$$I_{\text{out}} = I_1 + I_2 - I_3 - I_4.$$

On substitution of I_1, I_2, I_3 and I_4 from (1), (2), (4), and (5) into the above equation we obtain

$$I_{\text{out}} = -2KV_{\text{in}1}V_{\text{in}2}. \quad (6)$$

Thus, (6) gives four-quadrant multiplication for single ended inputs.

In this design, the two inputs are not symmetrical because the source at $V_{\text{in}2}$ has to sink current while the source at $V_{\text{in}1}$ does not. If we have a current mirror that gets current I_1 from current mirror C_1 and sinks the same current from the node at $V_{\text{in}2}$, the source at $V_{\text{in}2}$ will not have to sink current since the current flowing in and out of the node will be the same. However, this will result in a slight increase in the area and power consumption of the multiplier.

B. Range of Operation

For an nMOS transistor to be in saturation, two conditions have to be satisfied:

1. V_{DS} should be greater than $V_{\text{GS}} - V_T$
2. $V_{\text{GS}} - V_T > 0$

where, V_{DS} and V_{GS} are the drain-to-source voltage and gate-to-source voltage, respectively, of the transistor. For transistors M_1 and M_2, the first condition can always be satisfied by design. For the second condition to be satisfied,

$$V_{\text{in}1}, V_{\text{in}2} > V_b + V_T. \quad (7)$$

For M_3 to be in saturation,

$$V_{\text{in}1} + V_{\text{in}2} - V_b - V_T > 0. \quad (8)$$

For M_4 to be in saturation,

$$V_b < -V_T. \quad (9)$$

since the first condition can always be satisfied by design. M_5 is always in saturation because its gate and drain are connected. However, the drain voltage of M_5 should be such as to keep the transistors in current mirror C_1 in saturation. If C_1 and C_2 are cascode current mirrors, the drain voltage of M_5, M_{D5}, has to satisfy the following condition,

$$V_{D5} < V_{\text{DD}} - V_T - 2\delta V, \quad (10)$$

where, δV is equal to $V_{\text{GS}} - V_T$ for the transistors in the current mirror. For a given current, δV depends on the aspect ratio of the transistors in the current mirrors. Putting (3), (8), and (10) together, we get:

$$V_b + V_T < V_{\text{in}1} + V_{\text{in}2} < V_{\text{DD}} - 2\delta V_{\max} + V_b - V_T \quad (11)$$

where $\delta V_{\max}$ is the maximum value of δV. Conditions described by (7) and (11) limit the range of operation of the multiplier.

An example would explain how the range of operation can be obtained from these biasing constraints. Let V_{DD} be 5 V, V_b be -2.4 V and $\delta V_{\max}$ be 0.1 V. From the fabrication parameters, we find V_T to be 0.8 V. Substituting these values in (11) we get $|V_{\text{in}1} + V_{\text{in}2}| < 1.6$ V. Thus, for symmetric four-quadrant operation, we have $|V_{\text{in}1}, V_{\text{in}2}| < 0.8$ V.

If one of the inputs is always positive (say 0.8 V $> V_{\text{in}1} >$ 0 V), then by changing V_b to -2.0 V, we can make the multiplier operate for $|V_{\text{in}2}| < 1.2$ V. If both the inputs are always positive, then by changing V_b to -0.8 V, we can operate the circuit for the input range 1.6 V $> V_{\text{in}1}, V_{\text{in}2} > 0$ V. Similarly, the range can be changed to 0 V $> V_{\text{in}1}, V_{\text{in}2} >$ -1.6 V, if both the inputs are negative. In this manner, by changing the bias voltage V_b we can operate the circuit at various input ranges. This fact can be used to increase the overall range of the multiplier by changing V_b depending on the sign of the input voltages.

III. Experimental Results

The circuit shown in Fig. 1 was fabricated with a 2 μm p-well standard CMOS fabrication process through MOSIS. Fig. 2 shows a photograph of the fabricated chip. The active chip area in 97 μm $\times$ 110 μm. The sources of all the n-MOS transistors are connected to their local substrate by putting them in individual p-wells wherever necessary. The W/L ratio for the five transistors is 4 μm/20 μm and cascode current mirrors were used.

For obtaining the dc-characteristics of the multiplier, the bias voltage V_b was chosen to be -2.4 V. V_T for each nMOSFET was approximately 0.8 V. The supply voltage, V_{DD} was 5 V. The dc transfer characteristics of the multiplier are shown in Fig. 3, where $V_{\text{in}1}$ is varied between -1 V and 1 V, for various values of $V_{\text{in}2}$, which was also varied between -1 V and 1 V. As is expected, the multiplier shows an approximately linear characteristic except when $|V_{\text{in}1}+V_{\text{in}2}| >$ 1.6 V, which is represented by the upper-left and upper-right part of the graph.

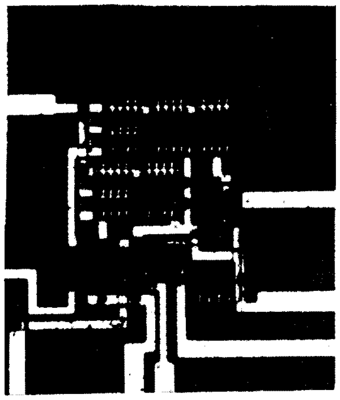

Fig. 2. Photograph of analog multiplier chip.

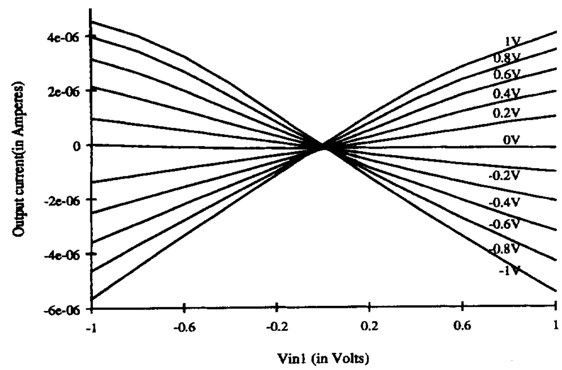

Fig. 3. Output characteristic of four-quadrant multiplier.

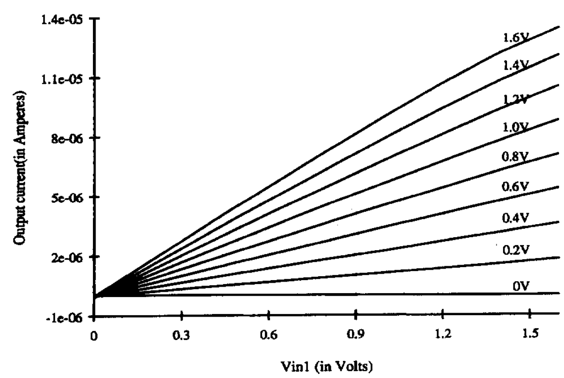

Fig. 4. Output characteristic of multiplier for positive inputs only.

To demonstrate how the input range can be adjusted, we make the bias voltage V_b equal 0.8 V and measure the dc characteristics of the circuit with both inputs varying from 0 to 1.6 V. The dc characteristics are shown in Fig. 4. Thus the input range of the multiplier has been shifted from $|V_{\text{in1}}, V_{\text{in2}}| < 0.8$ V to 0 V $< V_{\text{in1}}, V_{\text{in2}} < 1.6$ V.

The non-linearity of the multiplier is investigated by measuring the total harmonic distortion (THD) of the circuit. A sine wave of 2 kHz is applied to V_{in1} and the other input voltage is kept constant. Fig. 5 shows the THD as a function of the peak-to-peak voltage of V_{in1} for a constant V_{in2} of 0.8 V. As can be seen from the graph, the THD is less than 2% for the given range, except when the peak-to-peak voltage of V_{in1} exceeds 1.6 V. In that case, the linearity is decreased because the condition for saturation, $V_{\text{in1}} + V_{\text{in2}} < 1.6$ V, is not satisfied. Even when the p-p voltage of V_{in1} is 2 V, the THD is less than 3%. While this accuracy is modest, it is sufficient for most neural network applications.

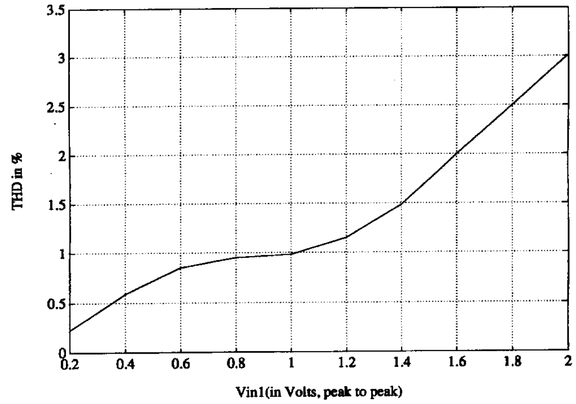

Fig. 5. Total Harmonic Distortion (THD) analysis of multiplier output.

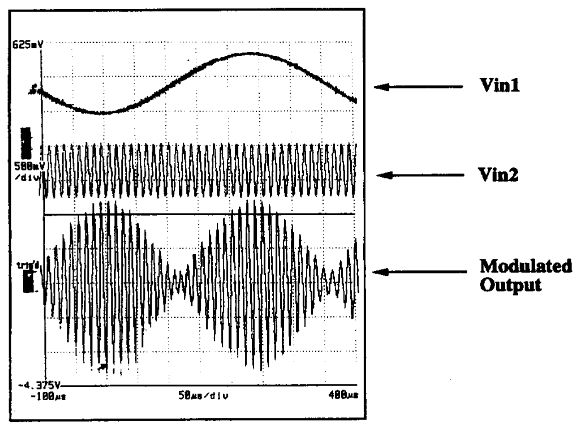

Fig. 6. Modulated output, one input 2 kHz and the other 85 kHz.

We investigate the behavior of the circuit as a modulator by feeding it a sinusoidal voltage of 2 kHz and peak-to-peak voltage of 0.8 V at one input, and another sinusoidal voltage of frequency 85 kHz and of peak-to-peak voltage 1 V at the other input. The current at the multiplier output is passed through a current-to-voltage converter and displayed on an oscilloscope. Fig. 6 shows the modulated output as seen on the oscilloscope.

The -3 dB bandwidth of the multiplier was measured to be 115 kHz. The simulated -3 dB bandwidth was 4.5 MHz. The anomaly is explained by the fact that during measurement, the multiplier was driving a pad, which slowed it down considerably.

The maximum power consumption for the multiplier within its operating range was approximately 1 mW, which is towards the lower end for the four-quadrant multipliers mentioned in this paper. The power consumption in the multiplier is proportional to the aspect ratio of the five transistors, M_{1-5}, all other parameters remaining constant. Hence, we can further reduce the power consumption of this circuit by decreasing the aspect ratio of the transistors. It is possible to reduce the aspect ratio, and hence the power consumption, by about 45% without affecting the major characteristics of the circuit including its area and range of operation. This fact was confirmed by SPICE simulations.

The input offsets were measured to be less than 15 mV. They can be attributed to random mismatches of the threshold voltages of the input transistors. The maximum output offset was less than 2% of the output range. It can be attributed to several factors. First are second-order effects such as mobility degradation and channel-length modulation, which we ignored in the calculation of the output current. If we include the second order effects in our calculations and use typical values, we can determine the output offset due to them. This offset can be reduced by changing the aspect ratio or gate voltage of transistor M_4. The other factor contributing to output offset is random mismatch between the transistors. Transistor mismatch can be reduced by increasing device sizes and using common-centroid geometry, but these will result in an increase in area of the multiplier.

IV. Conclusion

A new MOS four-quadrant multiplier with single ended inputs has been presented. The active area of the multiplier is 97 μm $\times$ 110 μm, which is smaller than the other four-quadrant multipliers mentioned in this paper by a factor of at least 8. This is achieved mostly because the number of transistors in the multiplier is small, and also because the dimensions of the transistors were chosen to be small and thus compromising some of the input-range and the linearity. Since the inputs are single ended, the pre-processing stage, required in multipliers with differential inputs, is eliminated, which saves additional area. The input range can be adjusted using a bias voltage. In the four-quadrant range of operation, both the inputs can be varied between -0.8 V and 0.8 V. The total harmonic distortion is less than 2% in the region of operation. The power consumption is less than 1 mW. The small area and low power consumption make it suitable for use in the implementation of artificial neural networks.

Acknowledgment

The authors would like to thank Prof. R. W. Brockett, Prof. W. Yang, Dr. D. J. Friedman and Dr. P. N. Belhumeur for their suggestions and support. They would also like to thank the reviewers for their helpful comments.

References

[1] J. N. Babanezhad and G. C. Temes, "A 20-V four-quadrant CMOS analog multiplier," *IEEE J. Solid-State Circuits*, vol. 20, no. 6, pp. 1158–1168, Dec. 1985.

[2] K. Bult and H. Wallinga, "A CMOS four-quadrant analog multiplier," *IEEE J. Solid-State Circuits*, vol. 21, no. 3, pp. 430–435, June 1986.

[3] J. J. Clark, "An analog CMOS implementation of a self organizing feedforward network," *Proc. IJCNN-WASH 90*, M. Caudill, Ed., vol. 1, pp. 118–121, 1990.

[4] S. M. Gowda *et al.*, "Design and characterization of analog VLSI neural network modules," *IEEE J. Solid-State Circuits*, vol. 28, no. 3, pp. 301–313, March 1993.

[5] M. Holler *et al.*, "An electrically trainable artificial neural network (ETANN) with 10240 'floating gate' synapses," *Proc. IEEE Int. Conf. Neural Networks*, vol. 2, Baltimore, MD, pp. 191–196, June 1989.

[6] P. W. Hollis and J. J. Paulos, "Artificial neural networks using MOS analog multipliers," *IEEE J. Solid-State Circuits*, vol. 25, no. 3, pp. 849–855, June 1990.

[7] Z. Hong and H. Melchior, "Four-quadrant CMOS analogue multiplier," *Electron. Lett.*, vol. 20, no. 24, pp. 1015–1016, Nov. 1984.

[8] Z. Hong and H. Melchior, "Analogue four quadrant CMOS multiplier with resistors," *Electron. Lett.*, vol. 21, no. 12, pp. 531–532, June 1985.

[9] N. I. Khachab and M. Ismael, "A nonlinear CMOS analog cell for VLSI signal and information processing," *IEEE J. Solid-State Circuits*, vol. 26, no. 11, pp. 1689–1699, Nov. 1991.

[10] Y. H. Kim and S. B. Park, "Four-quadrant CMOS analogue multiplier," *Electron. Lett.*, vol. 28, no. 7, pp. 649–650, March 1992.

[11] L. W. Massengill, "A dynamic CMOS multiplier for analog VLSI based on exponential pulse-decay modulation," *IEEE J. Solid-State Circuits*, vol. 26, no. 3, pp. 268–276, March 1991.

[12] J. S. Peña-Finol and J. A. Conelly, "A MOS four-quadrant analog multiplier using the quarter-square technique," *IEEE J. Solid-State Circuits*, vol. 22, no. 6, pp. 1064–1073, Dec. 1987.

[13] S. Qin and R. L. Geiger, "A ± 5 V CMOS analog multiplier," *IEEE J. Solid-State Circuits*, vol. 22, no. 6, pp. 1143–1146, Dec. 1987.

[14] D. B. Schwartz *et al.*, "A programmable analog neural network chip," *IEEE J. Solid-State Circuits*, vol. 24, no. 2, pp. 313–319, April 1989.

[15] H.-J. Song and C.-K. Kim, "An MOS four-quadrant analog multiplier using simple two-input squaring circuits with source followers," *IEEE J. Solid-State Circuits*, vol. 25, no. 3, pp. 841–848, June 1990.

[16] D. C. Soo and R. G. Meyer, "A Four-quadrant NMOS analog multiplier," *IEEE J. Solid-State Circuits*, vol. 17, no. 6, pp. 1174–1178, Dec. 1982.

[17] Z. Wang, "A CMOS four-quadrant analog multiplier with single-ended voltage output and improved temperature performance," *IEEE J. Solid-State Circuits*, vol. 26, no. 9, pp. 1293–1301, Sept. 1991.

[18] S. L. Wong et al, "Wide dynamic range four-quadrant CMOS analog multiplier using linearized transconductance stages," *IEEE J. Solid-State Circuits*, vol. 21, no. 6, pp. 1120–1122, Dec. 1986.

[19] M. Yasumoto and T. Enomoto, "Integrated MOS four-quadrant analogure multiplier using switched-capacitor techniques," *Electron. Lett.*, vol. 18, no. 18, pp. 769–771, Sept. 1982.