Implementation of a Charge-Based Neural Euclidean Classifier for a 3 Bit Flash Analog to Digital Converter

Bora M. Onat
Department of Electrical, Computer, and Systems Engineering,
Boston University, Boston, MA. 02215

Dr. John A. McNeill
Department of Electrical and Computer Engineering,
Worcester Polytechnic Institute, Worcester, MA. 01609

Dr. Uğur Çilingiroğlu
Department of Electrical Engineering,
Istanbul Technical University, Istanbul, Turkey

ABSTRACT

This paper describes the implementation of a Euclidean squared classifier with a charge based synaptic matrix and discriminator, based on a previously implemented Hamming classifier. The discriminator circuit is a generalized n-port version of the two-port differential charge-sensing amplifier that is conventionally used in DRAM's for bitline sensing. Both the quantifier and discriminator are implemented by charge based techniques, granting the simultaneous availability of high integration density, low power consumption and high speed. The analog-to-digital (A/D) implementation was chosen to illustrate the network's classification characteristics, since A/D conversion can be interpreted as classifying an input in terms of A/D quantization levels. A detailed analysis of the classifier configuration is presented. Design issues are addressed at both the system and circuit levels, and some limitations are identified. Simulation results of the the circuit confirming its theoretical performance are presented, as well as measurements of the implemented chip. The circuit occupies an area of $500\mu m \times 250\mu m$, operates with a single 5V power supply, and consumes less than 1mW of static power.

I. INTRODUCTION

The Neural Euclidean classifier, like the Hamming classifier [1], is one of the least complex artificial neural networks. Nevertheless, it achieves distinctive classification tasks with a very high connection efficiency. The network consists of a purely capacitive synaptic matrix which is preprogrammed during fabrication. For any arbitrary combination of exemplar count, convergence to any pattern other than that stored in its memory is impossible. The synaptic matrix measures the Euclidean squared distance between an m-dimensional input vector $(x_1, x_2, ..., x_m)$ and all of the previously fixed neural weights $(T_{i1}, T_{i2}, ..., T_{im})$. The "similarity scores" $(S_1, S_2, ..., S_n)$ thus generated for the exemplars are described by:

$$S_i = A' - A\sum_{j=1}^{m}(x_j - T_{ij})^2 \qquad (1)$$

0-7803-2615-6/95/$4.00 © 1995 IEEE

where A' is an arbitrary constant and A is an arbitrary positive constant. Therefore the similarity score obtains a maximum value when the input vector exactly matches the exemplar vector for the specific row i. After eliminating common mode terms that have no influence on the competitive decision process carried out in the discriminator subnet, equation (1) simplifies to

$$S_i = 2A\sum_{j=1}^{m}(x_j T_{ij}) - A\sum_{j=1}^{m} T_{ij}^2 \qquad (2)$$

So, for a Euclidean classifier, the capacitive quantifier subnet is the implementation of (2). The discriminator network identifies the best-matching exemplar (the row with maximum S_i) and determines it as the winner. The next section describes how this function is obtained via circuit variables.

II. Circuit Analysis

A. Quantifier

Figure 1 shows the complete schematic diagram of the circuit implementation. The quantifier subnet comprises 1) an $n \times m$ matrix of exemplar proportional synaptic capacitors $(C_{11}, ..., C_{nm})$, 2) the neural function invoking capacitors $(C_{T1}, ..., C_{Tn})$, 3) a normalization matrix containing dummy capacitors, 4) row and column precharge transistors clocked to Φ_1, and 5) input transmission gates clocked to Φ_2. Each row of synaptic capacitors is mask-programmed to represent the neural weight of the exemplar while the C_{Ti} capacitors assist generation of the similarity scores. The dc-driven dummy capacitors are also mask-programmed to equalize the total capacitance in each row.

The discriminator subnet (the right side of the circuit in Figure 1) is kept in a high impedance state until Φ_3 is activated, to ensure proper charging of capacitors during the first two clock stages. On Φ_2, input voltages are transferred to quantifier columns, perturbing row voltages through the synaptic capacitor matrix. These voltages convey the previously defined similarity scores. By writing the steady state total charge during the two clock phases Φ_1 and Φ_2, and equating them to satisfy charge conservation, we obtain

$$V_{Ri} = V_{ref} + \frac{V_{DD}\,K}{2\,C_{TOT}}\left(2\sum_{j=1}^{m}(x_j T_{ij}) - \sum_{j=1}^{m} T_{ij}^2\right) \qquad (3)$$

V_{Ri} is the voltage of row i and C_{TOT} is the total normalized capacitance of each row. K is a constant that involves sizing of the synaptic capacitors. Equation (3) has been obtained by using the following definitions

$$x_j = \frac{V_j}{V_{DD}} \quad T_{ij} = \frac{C_{ij}}{K} \quad \sum_{j=1}^{m} T_{ij} - \sum_{j=1}^{m} T_{ij}^2 = \frac{C_{Ti}}{K} \qquad (4)$$

$$C_{TOT} = C_{pi} + C_{Ti} + \sum_{j=1}^{m} C_{ij} \qquad (5)$$

V_j is the j th input voltage of the neural classifier. C_{pi} is the sum of all non-synaptic capacitances along row i. Note that the first term on the right-hand side of equation (3) is a common-mode term to be ignored, and the second term is identical to the similarity score defined previously. *The capacitive quantifier subnet generates the similarity scores in the form of row voltages at the end of the leading transition of* Φ_2.

B. Discriminator

The discrimination process begins as the sources of all NMOS transistors in the feedback matrix are pulled down on Φ_3. These transistors turn on in saturation and start discharging the rows. The rate of discharge is the lowest for the row of highest voltage because the lower voltages of other rows minimize its pull-down current. Therefore, the voltages of all non-best-matching rows eventually fall below V_T, the NMOS threshold voltage, while that of the best-matching row still remains above V_T. The moment this happens, the best-matching row returns to a high impedance state while other rows continue to be discharged until all are grounded. This completes the first phase of discrimination. When row pull up transistors are turned on with Φ_4, all row voltages start an upward swing. Only the voltage of the best-matching row, having an initial advantage of at least one V_T and the highest rate of pull-up, reaches V_{DD}. Other row voltages are kept below V_T by the pull down transistor M1 together with those feedback transistors

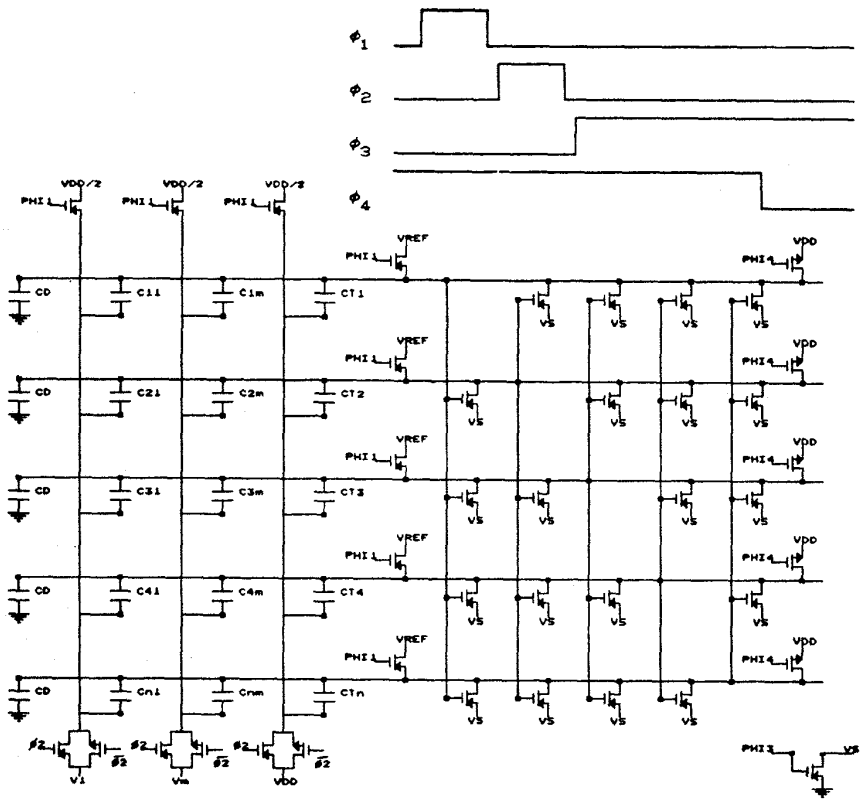

Figure 1: Circuit Diagram of the Neural Euclidean Classifier. The clock sequences are displayed in the upper right-hand corner.

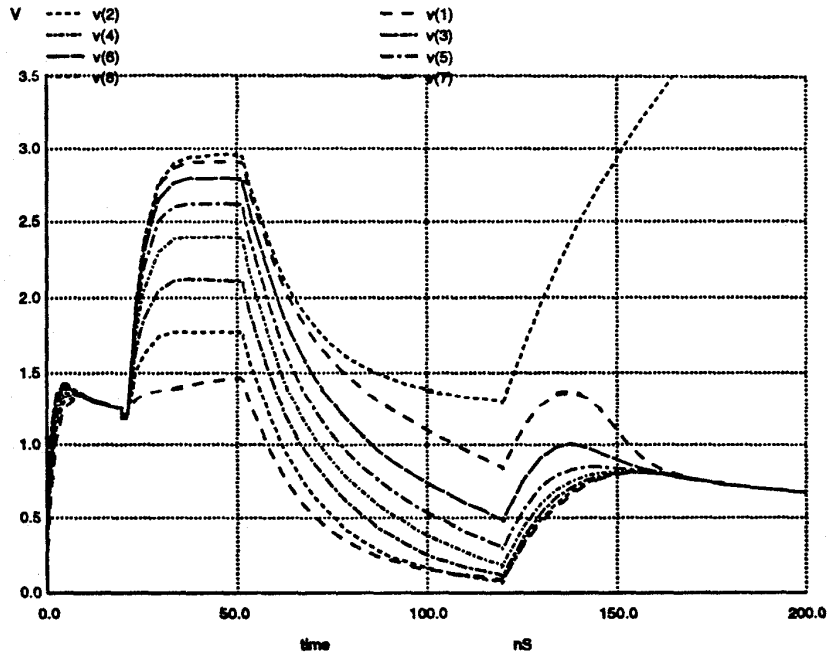

Figure 2: Spice simulation of the Neural Euclidean Classifier.

controlled by the best matching row. Output buffers digitize these signals to V_{DD} or $0V$, and transfer them to output channels during Φ_5.

By utilizing a Neural Euclidean Classifier, we have implemented a CMOS circuit with 8 rows and one input as an analog to digital converter. The SPICE simulation results are presented in Fig. 2 for the first four clock sequences. The first 20ns (Φ_1) display the precharging of the rows. Next, during Φ_2, the row voltages are perturbed according to the similarity scores. At 50ns (Φ_3), the discrimination process begins as the rows are discharged. The pull up transistors of each row pick the winner as they suppress the rest of the row voltages, once they are activated at 120ns (Φ_4).

III. Design Considerations

A. Offset Voltage

For the network to select the 'winner' properly, the network offset V_{off} originating from the quantifier and discriminator must be smaller than the minimum voltage difference between the highest two row voltages. The "quantifier offset" is defined as the amount by which a row voltage may differ from its pre-programmed value, due to randomly mismatching capacitive components as a result of fabrication. The "discriminator offset" describes the minimum voltage difference needed between the two highest row voltages for the winner selection. This nonzero voltage is necessary so that randomly mismatching row components and noise will not affect favoring the correct 'winner' of the two rows. The detailed formulation of the network offset has been presented in [2], from which total offset is calculated to be $30mV$ for $V_{DD} = 5\ V$ and a network with 8 rows and one input vector (this implies that $j = 1$ for all previous formulas).

The classifier, for the analog to digital circuit, is programmed so that the input of the converter is quantized to 8 equal levels. Note that the equal spacing is not a constraint: if desired, one may program smaller quanta spacings in some signal ranges to reduce quantization uncertainty. For this classifier to function properly, the minimum difference between the highest two row voltages, ΔV_R derived in [2], must be larger than the network offset voltage at any given input.

$$V_{off} \leq \frac{V_{DD}\ K}{2\ C_{TOT}} \times q^2 \tag{6}$$

where 'q' is the minimum quanta spacing between two programmed exemplars. The dummy capacitance of the row with the largest C_i and C_{Ti} is set to zero, in order to obtain a minimum C_{TOT}. Hence, an equation that gives a restriction of the sizing of the capacitors is obtained. Since K must be a positive constant, the denominator of the equation determines the minimum quanta spacing allowed.

$$K > \frac{2 \times V_{off} \times C_{row}}{V_{DD}q^2 - 2V_{off}(2T_i - T_i^2)_{max}} \tag{7}$$

$$q^2 > \frac{2V_{off}(2T_i - T_i^2)_{max}}{V_{DD}} \tag{8}$$

where C_{row} is the row capacitance of the non-synaptic and non-dummy capacitors. Once these constraints have been satisfied, the classifier may be designed to operate within safety margins. In the implemented circuit, the capacitor values used are 20 times larger than the minimum values required to have the classifier work safely.

B. Reference Voltage

As can be seen from the circuit schematic in Figure 1, the reference voltage V_{ref} is the initial value to which each of the row voltages is charged. Since it is the same for every row, it is a common mode term in (3) and (from a mathematical point of view) does not affect the discrimination process. From the point of view of the physical circuit, however, the choice of V_{ref} is very important for successful discrimination among the row voltages. If V_{ref} is too low (less than a threshold voltage V_T above the negative supply V_S), then for some input voltages it is possible that none of the discriminator transistors are conducting. This would be more of a problem for the smaller input voltages. On the other hand, if V_{ref} is too large, then discharging the row voltages would take longer. Also, simulation results have shown that when V_{ref} is too large, the discrimination process is not carried out correctly and other rows are falsely selected as the winner.

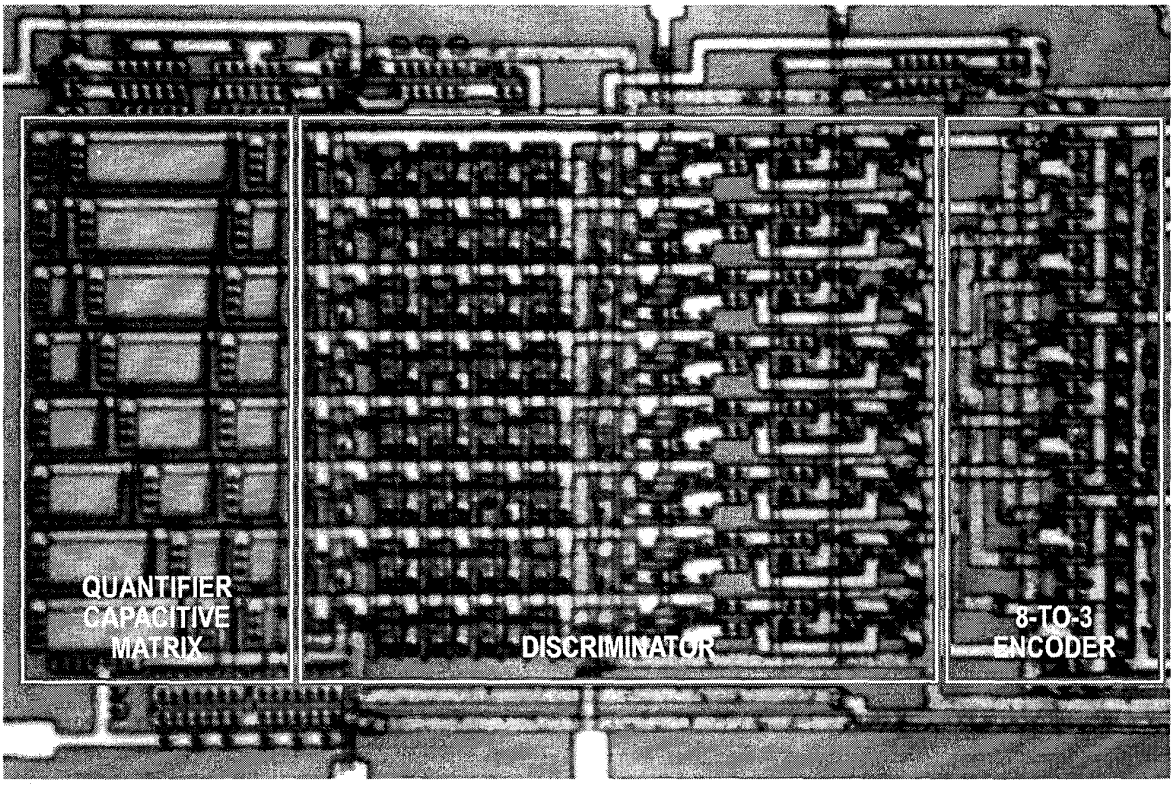

Figure 3: Die photo of circuit implemented in 2μm n-well CMOS process.

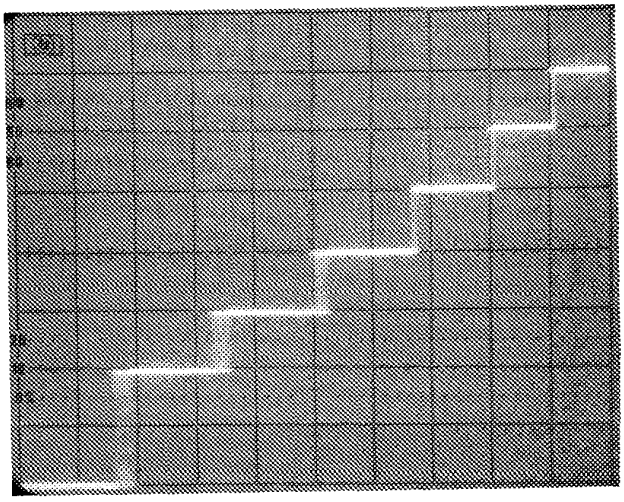

Figure 4: Input-output characteristic of neural classifier circuit as A/D converter.

In this implementation, the value of V_{ref} was chosen by iteration from simulation results. Unfortunately, no provision was included in the fabricated chip for adjusting V_{ref} to check if the simulation design choice was indeed optimal. As will be seen in the Experimental Results section, a problem was seen in the device characteristic that can probably be attributed to a too-low choice of V_{ref}.

IV. Implementation

The circuit of Figure 1 was fabricated in a 2μm n-well CMOS process through the MOSIS service. Figure 3 shows a die photo of the circuit, with the capacitive quantifer subnet and discriminator network identified. Also implemented was an 8-to-3 encoder which provided the digital output in binary format, rather than the 1-of-8 format at the discriminator output. The total area of the circuit is $500\mu\text{m}\times250\mu\text{m}$.

V. Experimental Results

A. Classification

To test the operation of the neural classifier A/D, the input-output characteristic was displayed on an oscilloscope. The A/D input was a triangle wave over the input range of 0 to +5V. The digital output was converted back into analog form with a D/A converter, which was displayed on the scope in X-Y mode.

Figure 4 shows the measured input-output characteristic at a conversion rate of 2.5MHz. The X axis is the input voltage at 0.5V/division; the Y axis is the analog representation of the digital output code at one LSB/division. Ideally the output would be a series of uniform steps (quantization intervals). Figure 4 shows good uniformity for the quantization intervals corresponding to codes 2 (010) through 7 (111); the peak-to-peak variation of step size from the average is $\pm24\%$ of a quantization interval.

A significant problem shown in Figure 4 is a missing code at code 1 (001). Two possible causes of this effect are being investigated. The first possible cause is the capacitance associated with a minimum geometry overlap between the row and orthogonal column lines of the quantifier subnet. In the analysis of Section II, this capacitance was assumed to be negligible. If its value were to be included, it would be subtracted from C_{ij} in equation (4). This would explain why larger quanta values are not affected by the overlapping capacitance, since this capacitance can be neglected for the larger C_{ij} values. The second possible cause for a missing code is the effect described in Section III.B, in which a too-low value of V_{ref} causes the discriminator to become inactive for low input voltages. Future work will address the influence of V_{ref} on the input range of this classification approach.

B. Power Consumption

At the maximum A/D conversion rate of 5MHz, the measured power consumption of the entire chip was 21 mW. Most of this power was required by the on-chip voltage references which was required for producing V_{ref}. The clock-rate-dependent component of power consumption was 450 μW, indicating the low power requirement of the classifier itself.

VI. Conclusions

This paper has presented the theory of a Neural Euclidean Classifier, examining its density, power, and speed advantages as well as its limitations due to offset and reference voltage design issues. A 5MHz analog to digital converter using the neural network has been designed and implemented. The performance of this converter confirms this paper's theoretical background for the Neural Euclidean Classifier.

References

[1] U. Çilingiroğlu, "A Charge-Based Neural Hamming Classifier" *IEEE Journal of Solid State Circuits,* January, 1993.

[2] Bora M. Onat "Backing up a Truck Using an Artificial Neural Network." *Istanbul Technical University BSEE thesis,* 1992.

[3] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design and Analysis.* Saunders College Publishing, 1987.

[4] L. A. Glasser and D. W. Dobberpuhl, *The Design and Analysis of VLSI circuits.* Reading MA. Addison Wesley 1985.