A Fine-Grain, Current Mode Scheme for VLSI Proximity Search Engine

Seiji Takeuchi and Takayasu Sakurai
Institute of Industrial Science, University of Tokyo, Tokyo, Japan, 106

Introduction

A Proximity Search Engine (PSE) which finds neighboring vectors of a specific input vector is used in various important applications such as neural network-based recognition systems, vector quantizers (VQ) and content addressable memories (CAM). It has been known, however, that the PSE is area and power hungry, since many arithmetic operations are required. In this paper, 53µW, 10 transistors per synapse architecture is presented which is based on current mode operations. The measurement results with 0.5µm CMOS process are described.

Compared with previously reported neural network LSI's [1], [2] which need several tens of transistors per synapse including storage elements, the present scheme uses 10 MOSFETs per synapse. There has been a proposal [3] to make use of charge-based operations in building a neural network to reduce the number of MOSFET's but it requires a double poly-Si process for capacitors of good precision. The present scheme is manufacturable with the conventional single poly CMOS process and hence is easily integratable w ith other CMOS processors and logic blocks. The present architecture is extensible to VQ's and CAM's and can be a promising macro in system LSI's.

Architecture and Basic Operation

Figure 1 shows the present PSE architecture. It consists of synapse cells, threshold cells, and sense-amplifier . The synapse circuit is shown in Fig.2. Each synapse cell has an SRAM cell which stores a weighting factor in the following neural function calculation.

$$Y_i = f\left(\sum_j W_{ij} \bullet X_j - \theta_i\right)$$

,where X_j is an input, W_{ij} is an weighting factor, θ_i is a threshold, f is a step function and Y_i is an output.

In most cases, learning is carried out by software before recognition operations start and good weighting factors and threshold values are provided to the PS E It is assumed that synapse weight is 1-bit accuracy but the bit depth is extended by assigning multiple synapses per input. Figures 3 and 4 show one column of the present PSE and operation waveforms.

Weighting factors are written in the synapse cells through write bit lines. Each threshold cell has a weighting factor of either of 1, 2, 4, 8... by having multiple pull-down and pull-up paths in a cell as shown in Fig.3. The threshold value is programmed by writing in the SRAM's in the threshold cells. Pull-up and pull-down paths make the read-out process differential.

First, the BL and BL bar are precharged to the voltage which is generated by one active pull-up path and one active pull-down path. By doing so, the bit line delay can be minimized. The simulated worst bit line delay time to make the BL voltage difference of 20mV is 1.3ns.

After the precharge, j-th word line is opened selectively depending on the input Xj. That is, if Xj is "1", the word line is opened. The pull-down path and pull-up path in the synapse cell is activated only when Xj is "1" and the stored Wij is "1". This corresponds to the 'AND' function in the basic neural function. Since synapse cells are connecte d to a bit line, the current is added, which realizes the 'Σ' function. The threshold value is subtr acted by adding current to the opposite bit line to the normal synapse cell. This is achieved by twisting the BL and BL bar as shown in Fig. 3. Once Yi's are obtained, they are sent to word drivers again to implement multi-stage neural network.

Measurement Results

Figure 5 shows a microphotograph of the chip fabricated with 0.5µm double metal and single poly CMOS process. 32 synapses are connected to one bit line. The area of one synapse is 19.6µm x 14.2µm. The maximum power consumption of 1.7mW per column is observed when there are 31 pull-up cells and 32 pull-down cells as seen in Fig.6. This corresponds to 53µW per synapse at 3.3V V_{DD}. The minimum bit line voltage difference is observed in the same condition as above and is 25mV (see Fig.7).

Further Extensions

The present architecture can be extended to realize VQ's and CAM's. For this objective, th e synapse cell needs to evaluate the following function.

$$Y = WTA_i\left(\sum_j W_{ij} \oplus X_j - \theta\right)$$

,where θ is set as 0 in a VQ and the number of synapses per bit line in a CAM. W_{ij} and X_j are assumed to be 1-bit. Figure 8 is an extended synapse circuit which can conduct XOR function as well as AND function. The WTA stands for Winner-Takes-All circuit (Fig.9). The WTA includes 31 comparators for 32 columns and carries out a tournament among all column output. A binary encoded location of the winner column is easily obtained by the tournament result A chip of this extended version is designed as shown in Fig. 10 and is under fabrication.

Acknowledgments

The chip is manufactured through the VLSI Design Education Center at Univ. of Tokyo collaborating with NTT Electronics Corp. and Dai-Nippon Printing Corp. Research support and useful discussions with T.Kuroda and T.Furuyama from Toshiba are to be acknowledged.

References

[1] T.Shima et al., "Neuro Chips with On-Chip BackProp and/or Hebbian Learning," ISSCC, pp.138-139, Feb.1992

[2] B.J.Sheu et al., "An analog Neural Network Processor for Self-Organizing Mapping," ISSCC, pp.136-137, Feb.1992.

[3] T.Shibata et al., "Advances in Neuron-MOS Application," ISSCC, pp.304-305, Feb.1996.