Design of Analog and Digital Hybrid MAC Circuit for Artificial Neural Networks

Ki-Hyuk Park, Min-Hyung Cho, Young-Deuk Jeon, Joo-Hyun Lee

Processor Research Group
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
hyuk@etri.re.kr

***Abstract*—Demand for high-performance hardware acceleration for machine learning applications is increasing rapidly. This paper presents a low power analog and digital hybrid MAC (Multiply and Accumulation) circuit for artificial neural networks. The proposed MAC circuit consists of an analog synapse unit, digital preprocessing and postprocessing unit for support of the parallel analog synapse cores. As the hybrid MAC circuit supports relatively low power and fast multiple MAC operations, it provides a good advantage in developing hardware accelerator for artificial neural networks.**

Keywords—MAC; analog; digital; hybrid; neural network

I. INTRODUCTION

Deep Neural Network (DNN) technology is rapidly evolving on the basis of the Internet which enables acquiring large amounts of learning data and hardware technology capable of high-speed parallel processing. In recent years, various Deep Neural Network (DNN) frameworks have been proposed such as R-CNN, SSD and YOLO V2, etc. Since the convolution operation is composed of millions of multiplications and additions, a high-performance accelerator having a parallel structure is required for real-time image recognition. The purpose of this study is to design a new MAC circuit that can be efficiently used in DNN for an image recognition field which is one of the most popular application fields of artificial neural networks.

This paper is organized as follows. Section II presents the proposed MAC algorithm and the architecture of our hybrid MAC circuit. Synthesis results are given in Section III. Section IV presents the conclusions.

II. DESIGN OF ANALOG AND DIGITAL HYBRID MAC CIRCUIT

A. The proposed MAC algorithm

Each convolutional layer and fully connected layer of DNN carries out matrix operations that computes multiple multiplications and additions of the form

$$y = \sum_{i=0}^{31} F_i \cdot W_i \quad \textbf{(1)}$$

where F_i is a feature data, W_i is a weight for the feature data and the accumulation size is 32 in this case. As indicated in the equation (1), the MAC operation with 32 feature and weight inputs needs 32 multiplications and 31 additions.

The multiplication is mathematically the combination of right shifts and additions of each bit-by-bit multiplication result of two operands. The presented MAC algorithm generates all bit products implicated in overall multiplications and adds the bit products considering their bit positions. The multiplication is mathematically the combination of right shifts and additions of each bit-by-bit multiplication result of two operands.

In the presented MAC algorithm, we deal with operands with signed-magnitude format. The presented MAC algorithm generates all bit products implicated in overall multiplications and adds the bit products considering their bit positions of the form

$$F[m] \cdot W[n] = \sum_{i=0}^{31} (\mathrm{F}i[7] \oplus \mathrm{W}i[7]) \cdot (F_i[m] \cdot W_i[n]) \quad \textbf{(2)}$$

where F_i and W_i are 8-bit signed-magnitude numbers and $\oplus$ is the XOR operation. With equation (2), the result of our MAC operation is as follows.

$$y = \sum_{n=0}^{6} \sum_{m=0}^{6} \sum_{i=0}^{31} (F_i[7] \oplus W_i[7]) \cdot (F_i[m] \cdot W_i[n]) \cdot 2^{m+n} \quad \textbf{(3)}$$

B. Architecture of Hybrid MAC Circuit

Fig. 1 shows the block diagram of the proposed hybrid MAC architecture. We have designed an analog and digital mixed architecture composed of digital preprocessor for preparing data which are inputs to the analog synapse, the analog synapse unit for adding bitwise bit products and postprocessor for composing MAC output from analog synapse outputs.

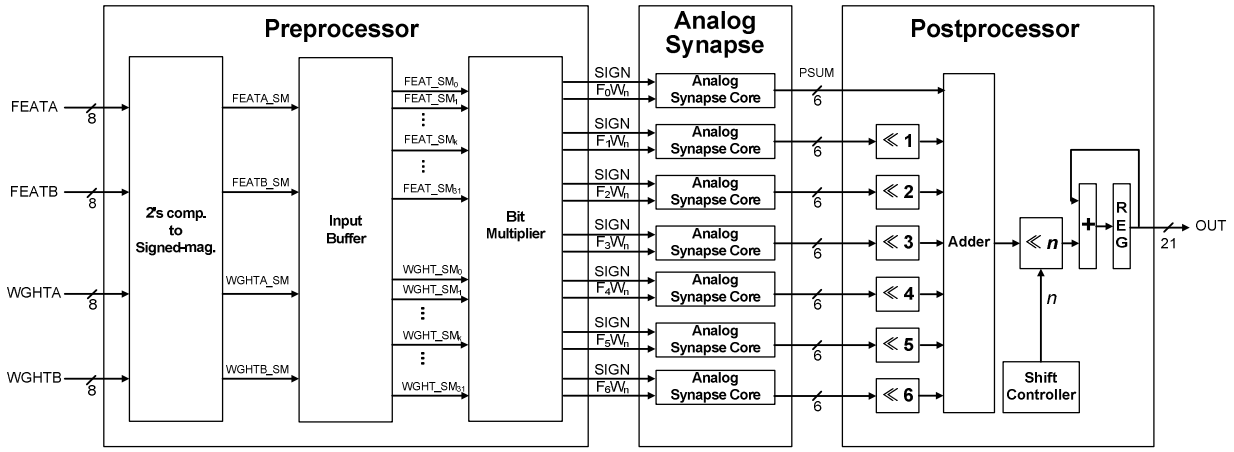


Fig. 1. The proposed hybrid MAC circuit

C. Preprocessing Unit

8-bit 2's complement inputs are converted to signed magnitude numbers and stored in the input buffer. When the buffer is full of thirty-two 8-bit feature and weight data, the bit multiplier generates 32-bit bitwise bit-products and their signs for each analog synapse core. For two 8-bit signed magnitude operands, 7 x 7 = 49 bit-products and 1-bit sign can be generated. In our design, each bit-product and sign is 32 bits wide, because there are 32 pairs of inputs. As shown in Fig. 2, 49 bit-products are divided into 7 groups of seven bit-products, so each group is fed to 7 analog synapse cores.

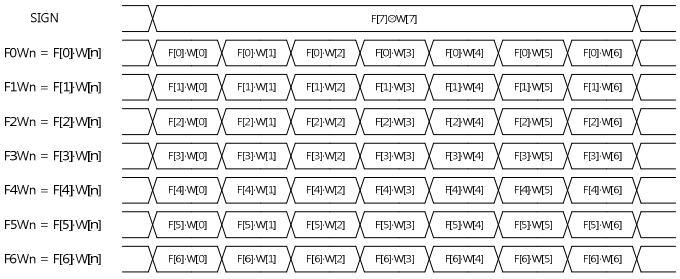
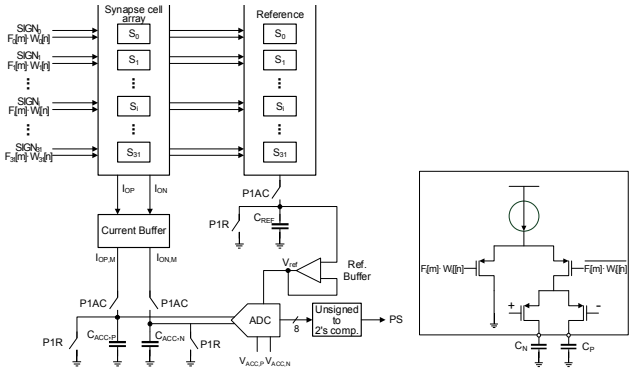


Fig. 2. Output sequence of Bit Multiplier

D. Analog Synapse Unit

The analog synapse unit performs the addition operation of 32 2-bit signed-magnitude numbers (a sign bit and a bit product) from the pre-processing unit. This unit consists of 7 parallel analog synapse cores. As shown in Fig. 3, each synapse core consists of a synapse cell array based on 1x32 identical current source synapse cells, a current buffer for weighting according to the bit position of processing two input bits, capacitors accumulating output currents from each synapse cell circuit, a reference voltage generator for an ADC reference voltage input, a ADC and unsigned-to-2's complement converter.



(a) Analog synapse core architecure (b) Synapse cell circuit

Fig. 3. Analog synapse core

E. Postprocessing Unit

Outputs from the analog synapse are shifted to the left individually by the appropriate bit position m in equation (3) before they are summed. The shifted outputs of the analog synapse are summed and the result is shifted to the left by the bit position n in equation (3). Finally, the shifted result is added to the stored accumulated value.

III. Experimental Results

We have implemented the digital parts of our hybrid MAC unit with SystemVerilog HDL. The analog parts have been designed and verified separately in Cadence analog design environment. Instead of analog/digital mixed simulation, we have implemented synthesizable high-level simulation model equivalent to the analog synapse core for the fast integrated simulation. Our MAC unit was synthesized by using Synopsys Design Compiler 2015-06 with TSMC 40nm technology library. Gate-level simulation with SDF annotation has been done for timing check. The main clock

frequency in our design is 400 MHz, and secondary clock for the analog synapse unit is 100 MHz.

The area and power report of the synthesis result are presented in TABLE I. As the table indicates, the analog synapse unit accounts for 90% of the total area. We also estimated power dissipation using Synopsys Design Compiler. Annotating switching activity on our design from gate-level simulation, we performed the power analysis. While the analog synapse consumes 33.2 % of the total power, the power dissipation in the preprocessor is 65.62%. It is estimated that frequent reading and writing data on the input buffer storing 32 pairs of input data in the preprocessor cause high power consumption.

TABLE I. AREA AND POWER ANALYSIS RESULT

Cell Name	Cell Area		Cell Power	
	Area [μm^2]	Percent [%]	Power [µW]	Percent [%]
Preprocessor	4,868.64	7.95	1,121.0	65.62
Synapse Core 0	4,967.09	12.84	81.0	4.74
Synapse Core 1	4,967.09	12.84	81.0	4.74
Synapse Core 2	4,967.09	12.84	81.0	4.74
Synapse Core 3	4,967.09	12.84	81.0	4.74
Synapse Core 4	4,967.09	12.84	81.0	4.74
Synapse Core 5	4,967.09	12.84	81.0	4.74
Synapse Core 6	4,967.09	12.84	81.0	4.74
Postprocessor	474.16	1.82	20.4	1.19
Top	40,112.46	100.00	1,708.4	100.00

* The area of NAND2 cell (ND2D1BWP) is 0.7 μm^2.

TABLE II compares the main results of the proposed MAC unit in this paper with those of a fully digital MAC unit which has a parallel architecture consisting of 32 8x8 bit multipliers and a multi-operand adder in terms of area, power and throughput. On one hand, the hybrid MAC structure has the advantage of low power consumption, but will increase the area and simultaneously decrease the throughput.

TABLE II. PERFORMANCE COMPARISON

	This work	Digital MAC
Operating Frequency (MHz)	400	400
Operating Voltage (V)	1.0 V	1.0 V
Technology	TSMC 40nm LP	TSMC 40nm LP
Area (μm^2)	40,112.5	16,133.5
Power (mW)	1.71	1.94
Max. Throughput (MHz)	7.55	25

IV. CONCLUSIONS

In conclusion, we have presented new analog and digital hybrid MAC unit suitable for the matrix operations performed in artificial neural networks. Adopting seven parallel current source based analog synapse cores instead of digital multipliers and a multi-operand adder makes the analog and digital hybrid MAC unit more energy-efficient than the fully digital one. The presented design was synthesized and simulated with TSMC 40nm technology library. Deploying our hybrid unit in a neuromorphic processor can yield power reduction in applications that compute many matrix operations, such as CNNs.

ACKNOWLEDGMENT

This work was supported by the ICT R&D program of MSIT/IITP [2016-0-00098, Brain-Inspired Neuromorphic Perception and Learning Processor].

REFERENCES

[1] D. Miyashita, S. Kousai, and T. Suzuki, "A Neuromorphic Chip Optimized for Deep Learning and CMOS Technology With Time-Domain Analog and Digital Mixed-Signal Processing," IEEE Journal of Solid-State Circuits, vol. 52, pp. 2679- 2689, 2017.

[2] J. Garland, and D. Gregg, "Low Complexity Multiply Accumulate Unit for Weight-Sharing Convolutional Neural Networks," IEEE Computer Architecture Letters, vol. 16. no. 2, issue. 2, pp. 132-135, 2017.

[3] H.O. Ahmed, M. Ghoneima, and M. Dessouky, "Concurrent MAC unit design using VHDL for deep learning networks on FPGA," IEEE Symposium on Comput. Applications & Industrial Electronics(ISCAIE), pp. 31-36, 2018.

[4] T.T. Hoang, M. Sjalander, and P. Larsson-Edefors, "A High-Speed, Energy-Efficient Two-Cycle Multiply-Accumulate (MAC) Architecture and Its Application to a Double-Throughput MAC Unit," IEEE Trans. on Circuits and Systems, vol 57, issue. 12, pp. 3073-3081, 2010.